creative commons
COMMONS DEED

저작자표시 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.
- 이차적 저작물을 작성할 수 있습니다.
- 이 저작물을 영리 목적으로 이용할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

The evolution and diversification of *Scutellaria* (Lamiaceae) based on phylogenomic analyses

Yoonkyung Lee

Department of Biology

The Graduate School of Sungshin University

The evolution and diversification of *Scutellaria* (Lamiaceae) based on phylogenomic analyses

A Dissertation

Submitted to the

Graduate School of Sungshin University

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Yoonkyung Lee

October, 2020

This is to certify that we have examined the

Doctoral Dissertation of

Yoonkyung Lee

Submitted to Department of Biology

Approved as to style and content:

Thesis Advisor .....

Committee Chairman .....

Committee Member .....

Committee Member .....

Committee Member .....

Committee Member .....

The Graduate School of Sungshin University

ABSTRACT

The evolution and diversification of *Scutellaria* (Lamiaceae) based on phylogenomic analyses

Yoonkyung Lee

Department of Biology

Graduate School of

Sungshin University

Scutellaria is the second largest genus in Lamiaceae including approximately 470 taxa in the world. It is economically important as the medicinal and aromatic value. Plants are annual or perennial herbs that can be distinguished from other genera of Lamiaceae by their characteristics such as persistent synsepal, presence of gynophore, exalbuminous seeds, and so on.

Morphological classification systems of *Scutellaria* have been suggested by many researchers. However, most of them were regional studies. A global classification system of *Scutellaria* based on the morphological study was conducted by Paton (1990) and it is widely used to date. In this classification system, two subgenera,

seven sections, two subsections, and 38 species groups (unofficial rank under the subsections) were included. Regional molecular phylogenetic studies of *Scutellaria* have been performed by Zhao et al. (2017) and Safikhani et al. (2018). However, these studies failed to show the detailed phylogenetic structure of *Scutellaria* because of regional sampling.

In this study, global phylogenomic studies were performed to understand the evolution and diversification of *Scutellaria.* The aims of this study were: 1) a comparison between molecular data and the current classification system (Paton, 1990), 2) the morphological re-examination of each subgroup recognized from the molecular studies, and 3) providing evidence for the future new classification system based on phylogenomic data.

As the first step of this study, rapidly evolved chloroplast genes (*ccsA*, *matK*, and *ndhF*) were selected based on the analysis of previously reported chloroplast genomes in *Scutellaria* (*S. baicalensis* and *S. insignis*). As a result of combined phylogenetic analysis of these three genes for 200 representatives in *Scutellaria*, six major clades were recognized with subgenus *Apeltantus* being a monophyletic group and subgenus *Scutellaria* being a paraphyletic group. Both sections *Apeltanthus* and *Lupulinaria* included in subgenus *Apeltanthus* were paraphyletic groups. Most sections and species groups were paraphyletic groups, and only section *Perilomia* included in subgenus *Scutellaria* formed a monophyletic group.

The second step to address phylogenetic relationships in *Scutellaria* is the sequencing of chloroplast whole genome.

Forty-seven taxa representing each section and species group of *Scutellaria* including outgroups were selected for chloroplast whole-genome sequencing. Total lengths of chloroplast genomes sequenced in *Scutellaria* ranged from 151,030 bp to 152,594 bp, containing a pair of IR regions (25,190-25,314 bp), LSC region (83,008-84,622 bp), and SSC region (17,292-17,600 bp). Chloroplast genomes of *Scutellaria* included in this study contain 80 protein-coding genes, 30 tRNA genes, and 4 rRNA genes without gene-content variations and showed a highly conserved structure without rearrangements. Sequence analysis showed fast-evolving chloroplast regions in *Scutellaria*. Top ten fast-evolving chloroplast DNA regions were identified as *trnK-rps16, ycf1, ndhF, petA-psbJ*, *petN-psbM*, *ccsA-ndhD*, *trnT*-UGU-*trnL*-UAA, *rpoB-trnC*, *rpl32-ccsA*, and *rbcL-accD* (in order from the best regions). The phylogenetic analysis based on super-matrix phylogeny was performed for 213 taxa by combining sequences of three chloroplast genes (*ccsA*, *matK,* and *ndhF*) and chloroplast genome sequences of 50 taxa including genomes from previous studies. Except for the position of section *Salviifoliae,* the topology of the super-matrix phylogenetic tree was congruent with the phylogenetic tree based on three chloroplast genes, and bootstrap values were increased in most clades.

As the third step, phylogenetic analyses based on the nuclear genome were conducted using 272 nuclear loci with 95 *Scutellaria* representatives. The length of aligned concatenation sequences was 965,469 bp. A maximum-likelihood tree based on the concatenated

matrix produced a highly supported and well-resolved tree. The topology was similar to previous chloroplast trees and the resolution of the concatenated ML tree was improved with high supporting values compared to the previous chloroplast trees. Again, the monophyly of subgenus *Apeltanthus* and the paraphyly of subgenus *Scutellaria* was clearly confirmed in this analysis. Different from the previous super-matrix chloroplast tree, two species (*S. heterophylla* and *S. pontica*) of section *Salviifoliae* formed a paraphyletic group indicating the close maternal relationship of these species. Similar to the previous chloroplast trees, section *Salazaria* was a sister to a clade containing some species of section *Scutellaria.* Section *Perilomia* was a monophyletic group and section *Anaspis* was located at the base of the tree. Most of the taxa in which the phylogenetic positions had been ambiguous were resolved in the chloroplast trees in the concatenated ML tree. To find the best nuclear DNA region among 272 loci that were useful in phylogenetic analysis with additional samples, the number of shared quartets between individual gene trees with the concatenated ML tree was calculated. As a result, the subset containing only one DN17712_c1_g5_i1 (*Arabidopsis* AT3G54460 homolog: the protein containing SNF2 domain, the helicase domain, and F-box family protein) locus showed the highest shared quartet proportion (87.27 %).

As an independent study, nuclear ribosomal DNA sequences were determined for 49 representatives, and the structural variation of the secondary structure of the internal transcribed spacer 2 (ITS2)

region in nuclear ribosomal DNA was identified. The repeating unit of nuclear ribosomal DNA in *S. orientalis* was identified and the total length of the unit was 8,047 bp. The phylogenetic tree based on nuclear ribosomal DNA region was similar to that from nuclear 272 single-copy loci. It is reported that four structural characteristics in the secondary structure of ITS2 are shared in eukaryotes (Schultz et al., 2005). However, the structure was not found in *Scutellaria*. In addition, secondary structures did not provide a key to identify the subgroups of *Scutellaria* recognized from the molecular phylogenetic tree, indicating lineage-free independent structural evolution happened in *Scutellaria*.

This study showed that the first molecular phylogenetic relationships of *Scutellaria* based on global sampling with proper representatives and genome-scale data: including 52 chloroplast genomes (150 kbp) and nuclear 272 loci (960 kbp). Based on both chloroplast and nuclear phylogenetic trees, four major groups were recognized in *Scutellaria.* These groups were completely different from major groups in the current classification system (Paton, 1990), and these are potential subgenera in the new classification system. In addition, the result confirms potential hybridization origins of many taxa in two major groups, and highly supported common clades in both chloroplast and nuclear trees could be named by the concept of the phylogenetic classification. The result will provide the basis for a new classification system in *Scutellaria.*

CONTENTS

ABSTRACT ····· i
CONTENTS ····· vi
LIST OF FIGURES ····· xi
LIST OF TABLES ····· xv

CHAPTER 1. Introduction ····· 1
I. Lamiaceae ····· 2
1. Background of the family ····· 2
2. Classification history in the family ····· 5
3. Recent molecular phylogenetic studies ····· 9
II. *Scutellaria* ····· 14
1. Background of the genus ····· 14
2. Classification history in *Scutellaria* ····· 17
3. Recent molecular phylogenetic studies and the purposes of this study ····· 22

CHAPTER 2. Phylogenetic relationships in *Scutellaria* inferred from three chloroplast genes (*ccsA*, *matK*, and *ndhF*) ····· 24
I. Introduction ····· 25
II. Materials and Methods ····· 28
1. Taxon sampling ····· 28

2. DNA extraction, polymerase chain reaction, and sequencing 28
3. Phylogenetic analyses 34
III. Results 35
1. Sequence analyses 35
2. Phylogenetic relationships based on three fast-evolving chloroplast genes tree 36
IV. Discussion 42
1. Comparison between the phylogenetic tree and the current classification system 42
2. Comparison between trees from this study and previous molecular phylogenetic studies 46

CHAPTER 3. Comparative analyses of chloroplast genomes in *Scutellaria* and phylogeny based on the super-matrix approach 49
I. Introduction 50
II. Materials and Methods 53
1. Taxon sampling 53
2. DNA extraction, amplification, and sequencing 53
3. *De-novo* assembly and annotation 54
4. Sequence analyses 55
5. Phylogenetic analyses based on super-matrix approach 56
6. Calculation of quartet distance among phylogenetic trees 57
III. Results 58

1. Completion of chloroplast genome sequences ····· 58
2. Structural analyses of chloroplast genomes ····· 63
3. Repeat sequences ····· 71
4. Nucleotide diversity in chloroplast genomes ····· 77
5. Comparison between super-matrix tree and three chloroplast genes tree ····· 79
6. Quartet distance among phylogenetic trees ····· 85
IV. Discussion ····· 87
1. Characteristics of chloroplast genomes in *Scutellaria* ····· 87
2. Phylogenetic relationships based on super-matrix tree ····· 89
3. Screening of the best region for the phylogenetic study ····· 91

CHAPTER 4. Phylogeny of *Scutellaria* based on 272 nuclear single-copy genes using Hyb-Seq technique ····· 94
I. Introduction ····· 95
II. Materials and Methods ····· 100
1. Probe design ····· 100
2. Taxon sampling, DNA extraction, and sequencing ····· 102
3. Raw data analyses ····· 104
4. Phylogenetic analyses ····· 105
III. Results ····· 107
1. Sequence analyses ····· 107
2. Phylogenetic analyses based on concatenation approach ····· 111
3. Phylogenetic analyses based on coalescent approach ····· 117

IV. Discussion ····· 122
1. Discordance between the concatenated ML tree and the coalescent ASTRAL tree ····· 122
2. Reexamination of phylogenetic relationships based on the concatenated ML tree ····· 123

CHAPTER 5. Phylogenetic analyses based on nuclear ribosomal DNA sequences and the secondary structure of ITS2 in *Scutellaria* ····· 126
I. Introduction ····· 127
II. Materials and Methods ····· 130
1. Taxon sampling and *De-novo* assembly ····· 130
2. Completion of nrDNA sequences by PCR and Sanger sequencing ····· 131
3. Phylogenetic analyses ····· 133
4. Secondary structure analyses of ITS2 ····· 133
III. Results ····· 134
1. Completion of nrDNA sequences and phylogeny based on nrDNA ····· 134
2. Analyses of ITS2 secondary structures ····· 140
IV. Discussion ····· 146

CHAPTER 6. Conclusion ····· 148
I. Comparison between chloroplast and nuclear phylogenetic trees ····· 149

II. Comprehensive review of phylogenetic relationships in *Scutellaria* ····· 152
III. Futher studies ····· 156

REFERENCES ····· 158
ABSTRACT (in Korean) ····· 177
APPENDICES FOR CHAPTER 2 ····· 181
APPENDICES FOR CHAPTER 3 ····· 191
APPENDICES FOR CHAPTER 4 ····· 241
APPENDICES FOR CHAPTER 5 ····· 276

LIST OF FIGURES

Fig. 1. Phylogenetic position of Lamiaceae in the angiosperm phylogenetic tree. ····· 4

Fig. 2. A recent phylogenetic study of Lamiaceae (Li et al., 2016) based on five chloroplast DNA regions (*matK*, *ndhF*, *rbcL*, *rps16*, and *trnL−F*). ····· 12

Fig. 3. Phylogenetic position of *Scutellaria* (grey box) in the subfamily Scutellarioideae (Li et al., 2016) based on recent molecular phylogenetic study (Zhao et al., 2020). ····· 16

Fig. 4. A flowchart for finding fast−evolving chloroplast DNA regions for phylogenetic study. ····· 33

Fig. 5. A maximum likelihood tree based on chloroplast *ccsA, matK,* and *ndhF* genes. ····· 37

Fig. 6. A summarized ML tree based on taxonomic groups recognized in the current classification system (Paton, 1990). ····· 40

Fig. 7. Structure of chloroplast genomes in *Scutellaria.* ····· 59

Fig. 8. The result of progressive Mauve alignment based on the complete chloroplast genomes from 50 *Scutellaria* taxa. ····· 64

Fig. 9. Comparison of the IR borders in chloroplast genomes of 50 *Scutellaria* taxa. ····· 69

Fig. 10. The types and number of detected SSRs in 50 chloroplast genomes in *Scutellaria*. ····· 72

Fig. 11. Distribution of 20 types of SSR repeat units detected in 50 chloroplast genomes in *Scutellaria.* ····· 73

Fig. 12. Frequency of forward, palindromic, reverse, and complement repeat in each taxon. ····· 75

Fig. 13. Frequency of repeat length in each chloroplast genome. ····· ····· 76

Fig. 14. Nucleotide diversity of *Scutellaria* chloroplast genomes based on sliding window analysis. ····· 78

Fig. 15. A super-matrix maximum-likelihood tree based on three chloroplast genes (213 taxa) and chloroplast whole genome data. ··· ····· 80

Fig. 16. A summarized super-matrix ML tree based on taxonomic groups recognized in the current classification system (Paton, 1990). ····· 84

Fig. 17. A graph showing the percentage of quartet distance calculated from each phylogenetic tree. ····· 86

Fig. 18. Workflow of the Hyb-Seq technique. ····· 98

Fig. 19. A Venn diagram indicating the number of shared single-copy genes among *S. baicalensis*, *Arabidopsis thaliana, Populus trichocarpa, Oryza sativa,* and *Vitis vinifera*. ····· 101

Fig. 20. A heatmap visualizing the percentage recovered for 272 nuclear single-copy loci using HybPiper. ····· 110

Fig. 21. A concatenated ML tree obtained using IQ-TREE based on 272 nuclear single-copy loci. ····· 113

Fig. 22. A graph plotted by sorting in descending order of the shared quartet percentage of individual 272 loci with the concatenated ML tree. ····· 116

Fig. 23. A graph of recalculating the shared quartet percentage by combining the subset of top eight loci. ····· 117

Fig. 24. A coalescent ASTRAL tree obtained using ASTRAL based on 272 nuclear single-copy loci trees. ····· 119

Fig. 25. A tanglegram for comparing the concatenated ML tree (left) and the coalescent ASTRAL tree (right). ····· 120

Fig. 26. Piechart on nodes showing conflict and congruent between individual gene trees and the coalescent ASTRAL tree based on PhyParts analysis. ····· 121

Fig. 27. General structure of nrDNA repeating unit in plants. ···· 127

Fig. 28. A typical graphical visualization of the ITS2 secondary structure of eukaryotes (Schultz et al., 2005). ····· 129

Fig. 29. Phylogenetic tree of *Scutellaria* based on nrDNA using RAxML. ··· ····· 139

Fig. 30. Distribution of types of ITS2 secondary structure in the phylogenetic tree. ····· 141

Fig. 31. Twelve ITS2 structural types of *Scutellaria*. ····· 142

Fig. 32. A tanglegram showing phylogenetic trees based on nuclear (left) and chloroplast (right) datasets. ····· 151

LIST OF TABLES

Table 1. The classification history of Lamiaceae ····· 7

Table 2. The classification history of *Scutellaria* ····· 18

Table 3. The detailed infrageneric classification system of *Scutellaria* ····· 21

Table 4. Number of taxa in each subgroup based on the current classification system (Paton, 1990) and that included in this study ····· 30

Table 5. List of fast-evolving chloroplast DNA regions that are useful for the phylogenetic study of *Scutellaria* ····· 31

Table 6. The information of primers used in this chapter ····· 32

Table 7. Statistical summary of sequence for three fast-evolving genes in *Scutellaria* ····· 35

Table 8. Statistical summary for 52 chloroplast genomes including outgroup taxa ····· 60

Table 9. The functions and gene groups of genes included in the chloroplast genomes in *Scutellaria* ····· 62

Table 10. Top ten regions with high nucleotide diversity and their sequence variability in *Scutellaria* chloroplast genomes ····· 78

Table 11. Statistical summary of sequencing and assembly results using HybPiper ····· 108

Table 12. Statistical summary of concatenated matrix ····· 111

Table 13. Statiscical summary of nrDNA sequences of 49 *Scutellaria* taxa and outgroup taxa ····· 136

Table 14. Statistical summary of nrDNA sequences in *Scutellaria* excluding heterogeneous sequences (35 taxa) ····· 138

Table 15. Comparison between groups in the current classification system (Paton, 1990) and the characteristics of the ITS2 secondary structure of *Scutellaria* ····· 145

CHAPTER 1.

Introduction

I. Lamiaceae

1. Background of the family

Lamiaceae is well known as a mint family of flowering plants, with 236 genera and approximately 7,280 species (Stevens, 2001; Angiosperm Phylogeny Website). According to the APG IV classification system (Fig. 1; The Angiosperm Phylogeny Group, 2016), this family is included in Lamiales, euasterids, asterids, core-eudicots, eudicots, and angiosperms. Lamiaceae is the sixth-largest family in angiosperms, and most plants are cosmopolitan. Plants are economically important as medicinal and aromatic plants. For example, many plants such as mint, lavender, basil, rosemary, and oregano are widely cultivated for horticultural, familiar garden herbs. In Korea, 26 genera and 65 taxa in Lamiaceae are distributed (Suh et al., 2007).

Most plants of Lamiaceae often produce the iridoids and phenolic glycosides as a defense against herbivores or against infection by microorganisms (Mint Evolutionary Genomics Consortium, 2018). Typically, plants are herbaceous, shrub, or tree, and the cross-sections of the stems are mostly rectangular (Judd et al., 2016). The hair is a capitate-glandular hair, where ethereal oils (terpenoids) secrete. The non-glandular hair is commonly multicellular, or a mixture of multicellular hair and unicellular hair. Leaves are usually opposite, occasionally whorled, simple, entire, or toothed. Stipules are lacking, and the main axis of the inflorescence

is indeterminate. Flowers are bisexual, and the corolla is five-lobed, two-labiate, and the corolla lobe is imbricate. The lateral flowers are determinate and commonly zygomorphic. The sepal is five-lobed and connate, and the calyx is actinomorphic or zygomorphic, tubular- campanulate, heel-shaped, persistent, sometimes enlarged in fruit. The anterior stamens are longer than the posterior, sometimes equal, and occasionally reduced into two. The filaments adnate to the corolla, and the pollen is tricolpate or hexacolpate. The ovary is superior, syncarpous, axile placentation, and includes two-locules, which looks like four-locules because of the septum. The style commonly branched at the terminal and apical or gynobasic with the ovary. The stigma is two-fid, and it is located inconspicuously at the end of a small and branched style. Two ovules are laterally attached per carpel, and it has a thin-walled megasporangium within one integument and leptosporangium. The characteristic of the ovule has been considered to be an important characteristic that distinguishes Lamiaceae as a monophyletic group. The nectar disk is often present. The fruit is a drupe with one to four seeds, indehiscent, or a legume with four seed, or four-nutlets or four-schizocarp.

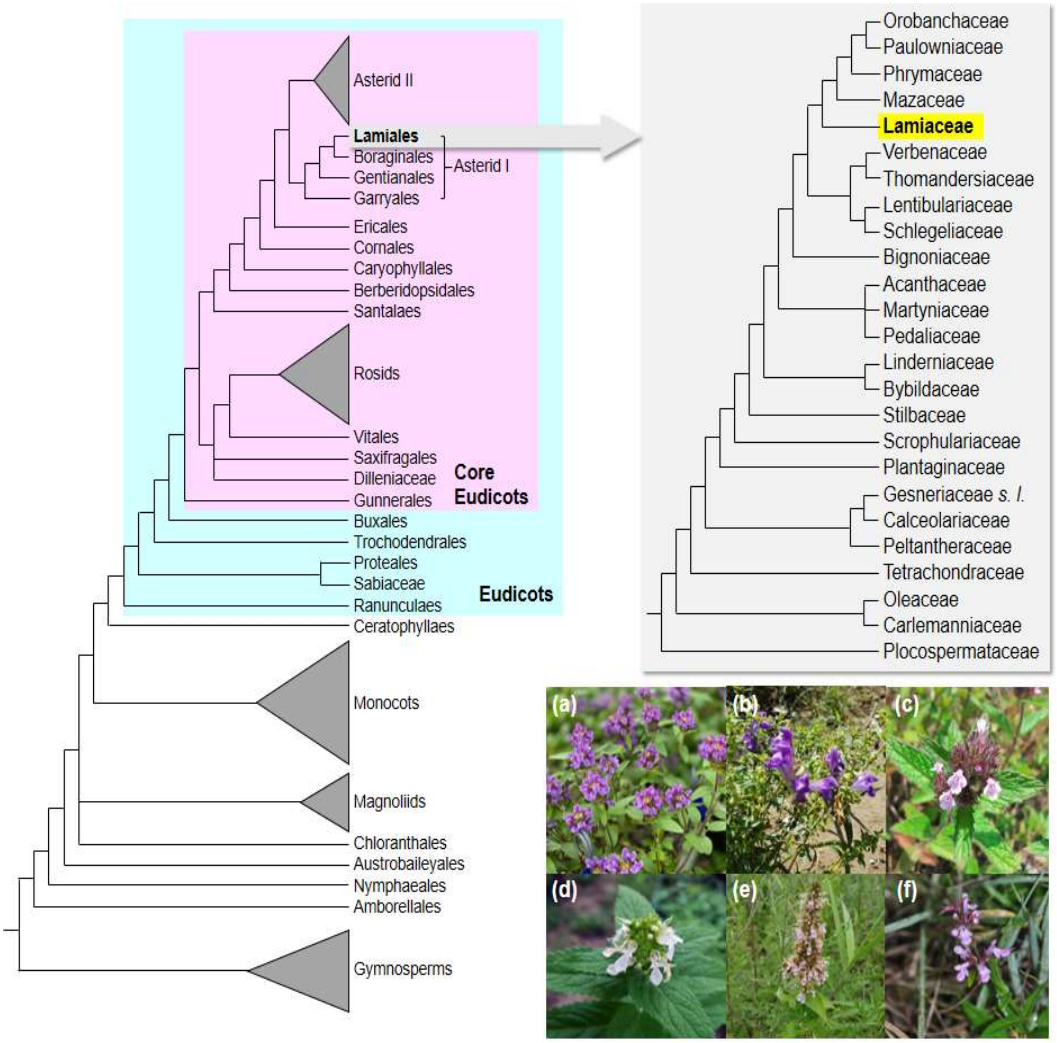


Fig. 1. Phylogenetic position of Lamiaceae in the angiosperm phylogenetic tree (summarized from Angiosperm Phylogeny Website: http://www.mobot.org/MOBOT/research/APweb/). The pictures (a–f) are some plants of Lamiaceae that distribute in Korea. (a) *Prunella vulgaris* var. *lilacina* Nakai; (b) *Scutellaria baicalensis* Georgi; (c) *Clinopodium chinense* var. *parviflorum* (Kudo) H. Hara; (d) *Nepeta cataria* L.; (e) *Teucrium veronicoides* Maxim.; (f) *Stachys japonica* Miq.

2. Classification history in the family

The classification system of Lamiaceae based on morphological characteristics has been well studied (Table 1). Bentham (1832-1836) suggested the first classification system in Lamiaceae. Later, Bentham (1876) modified this classification with detailed studies based on floral structure, which was divided into eight tribes; Ocimoideae, Satureineae, Moanrdeae, Nepeteae, Stachydeae, Prasieae, Prostanthereae, and Ajugoideae. Briquet (1895-1897) modified Bentham's classification system into eight subfamilies; Ajugoideae, Prostantheroideae, Prasioideae, Scutellarioideae, Lamioideae, Lavanduloideae, Ocimoideae, and Catoferioideae. Erdtman (1945) classified Lamiaceae into two subfamilies based on pollen traits, it was recognized that Lamioideae shed tricolpate pollen at a two-celled stage and Nepetoideae shed hexacolpate pollen at a three-celled stage. Wunderlich (1967) suggested six subfamilies based on the pollen, germination, and seed characters; Ajugoideae, Prostantheroideae, Scutellarioideae, Lamioideae, Nepetoideae, and Catoferioideae.

Lamiaceae and Verbenaceae have been studied of their phylogenetic relationship. Cronquist (1981) suggested that the characteristic of gynobasic style is a key that distinguishes the two families. According to Cronquist (1981), Lamiaceae have a deeply four-lobed ovary, in contrast, Verbenaceae have an unlobed ovary and a terminal style. Cantino (1982) performed a morphological study for two families and showed that Verbenaceae was more

closely related to Boraginaceae and Scrophulariaceae than Lamiaceae. However, Cronquist (1981; 1988) suggested that the relationship between Verbenaceae and Lamiaceae was unresolved and noted that shared characteristics were found in both subfamilies. This unresolved relationship was indicated similarly in other studies (Junell, 1934; Cantino, 1992a; Cantino, 1992b), and Lamiaceae has been recognized as a polyphyletic group. Junell (1934) suggested that the placentation of cymose was similar between Verbenaceae and Lamiaceae. Cantino (1992a; 1992b) conducted detailed morphological studies to investigate the relationship between Verbenaceae and Lamiaceae and suggested that Lamiaceae is a polyphyletic group and that several clades represent an independent clade of Verbenaceae. Therefore, morphological studies had not fully resolved clear phylogenetic delimitation between Verbenaceae and Lamiaceae.

Table 1. The classification history of Lamiaceae. Each row of the table includes the same taxonomic whatever their ranks are and arrows indicate combined, divided, or transpositioned groups.

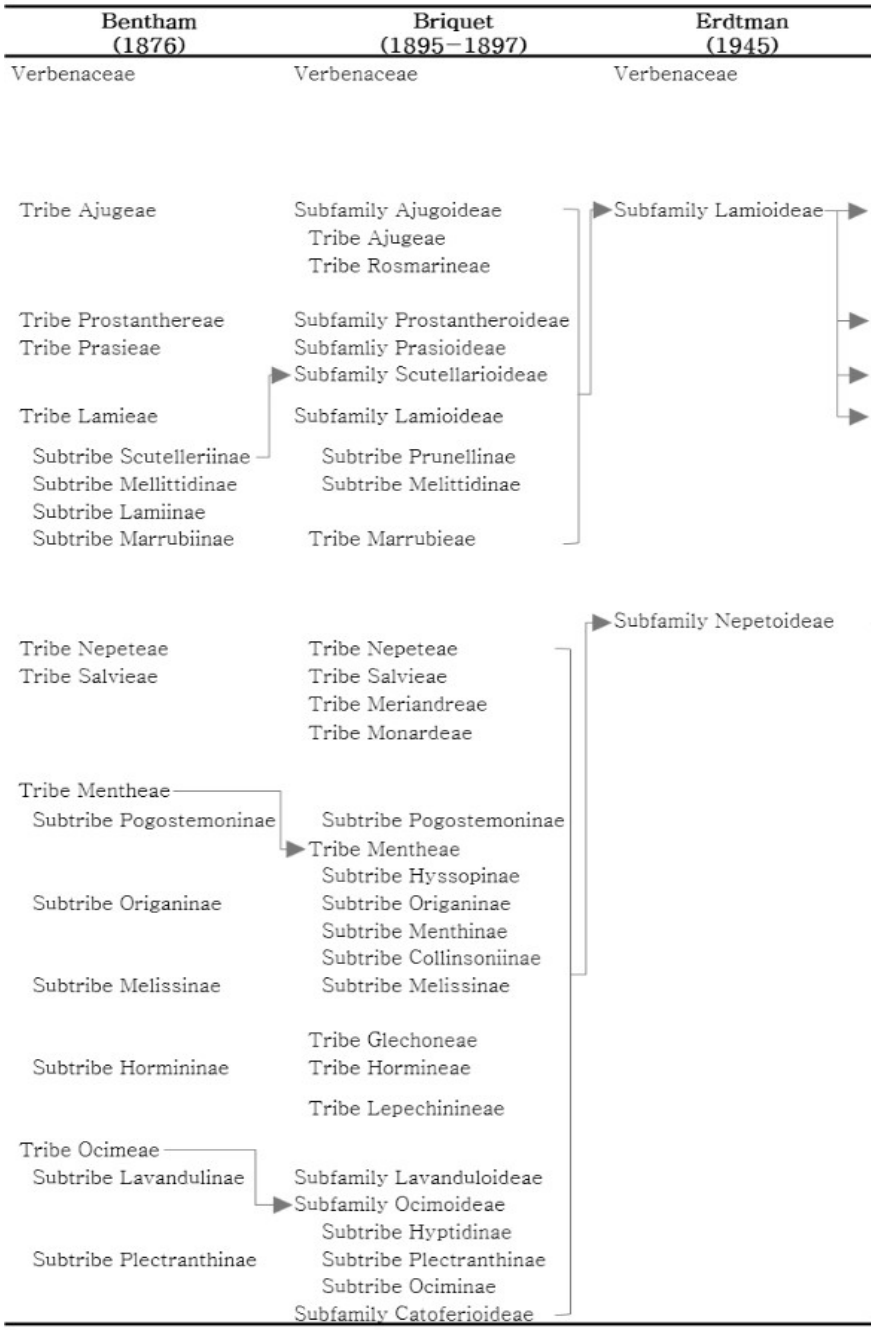


Table 1. Continued

Wunderlich (1967)	Harley et al (2004)	Li et al (2016–2017)
Verbenaceae	Subfamily Symphorematoideae	Subfamily Symphorematoideae
	Subfamily Viticoideae	Subfamily Viticoideae
		Subfamily Callicarpoideae
		Subfamily Premnoideae
		Subfamily Tectonoideae
Subfamily Ajugoideae	Subfamily Ajugoideae	Subfamily Ajugoideae
		Subfamily Peronematoideae
Subfamily Prostantheroideae	Subfamily Prostantheroideae	Subfamily Prostantheroideae
Subfamily Scutellarioideae	Subfamily Scutellarioideae	Subfamily Scutellarioideae
Subfamily Lamioideae	Subfamily Lamioideae	Subfamily Lamioideae
Tribe Lamieae		
Subtribe Melittidinae		
Subtribe Lamiinae		
Tribe Marrubieae		
Tribe Prasieae		
		Subfamily Cymarioideae
Subfamily Nepetoideae	Subfamily Nepetoideae	Subfamily Nepetoideae
Tribe Nepeteae		
Tribe Salvieae		
Tribe Meriandreae		
Tribe Monardeae		
Subtribe Elsholtzieae		
Tribe Mentheae		
Subtribe Hyssopinae		
Subtribe Origaninae		
Subtribe Menthinae		
Subtribe Collinsoniinae		
Subtribe Melissinae		
Tribe Prunelleae		
Tribe Glechoneae		
Tribe Hormineae		
Tribe Lepechinineae		
Tribe Ocimeae		
Subtribe Ociminae		
Subtribe Hyptidinae		
Subtribe Plectranthinae		
Subfamily Catoferioideae		

3. Recent molecular phylogenetic studies

With the development of sequencing technology, molecular phylogenetic studies based on chloroplast DNA sequences have been performed by many researchers to establish the classification system in Lamiaceae clearly. Wagstaff and Olmstead (1997) performed the comprehensive molecular phylogenetic study based on *rbcL* region in chloroplast and compared it with the classification suggested by Cantino (1992b). As a result of this study, the parsimony analysis supported the monophyly of Lamiaceae, which includes four subfamilies (Carypoteridoideae, Chloanthoideae, Viticoideae, and Symhorematacеae). However, some genera (*Hymenopyramis, Petraeovitex, Callicarpa,* and *Tectona*) were weakly supported and unresolved phylogenetic positions. Wagstaff et al. (1998) were also performed the molecular phylogenetic study inferred from *rbcL* and *ndhF* regions in chloroplast, and Lamiaceae and four subfamilies (Nepetoideae, Lamioideae, Pogostemonoideae, and Scutellarioideae) were showed as monophyletic groups with well-supported values. But the others were remained unresolved in phylogenetic tree.

Harley et al. (2004) defined seven subfamilies in Lamiaceae by integrated study combining previous morphological studies (Cantino and Sanders, 1986; Cantino et al., 1992) and molecular phylogenetic studies (Wagstaff et al., 1995; Wagstaff and Olmstead, 1997; Wagstaff et al., 1998): Ajugoideae, Prostantheroideae, Scutellarioideae, Lamioideae, Nepetoideae, Symphorematoideae, and

Viticoideae. Especially, Symphorematoideae has been previously considered belonging to Verbenaceae, but was first suggested as a subfamily of Lamiaceae in this classification based on molecular phylogenetic studies (Wagstaff et al., 1998; Bendiksby et al., 2011). Viticoideae remained unresolved position in previous studies, but Throne (1992) and Cantino (1992) first recognized Viticoideae as a subfamily of Lamiaceae, and Harley et al. (2004) adopted it. However, Viticoideae has not been showed as a clear monophyletic group. The classification system in Lamiaceae was relatively well-established by Harley et al, (2004), and this classification system has been widely used. However, the positions of some genera remained unresolved and needed further study. Therefore, nine genera (*Callicarpa*, *Hymenopyramis*, *Petraeovitex*, *Peronema*, *Garrettia*, *Cymaria*, *Acrymia*, *Holocheila*, and *Ombrocharis*) were still unresolved their phylogenetic position.

Recently, Li et al. (Fig. 2; 2016) performed the phylogenetic analysis based on five chloroplast DNA regions (*matK, ndhF, rbcL, rps16,* and *trnL-F*) including 288 species of 191 genera to re-establish the classification system of Lamiaceae. According to Li et al. (2016), clear phylogenetic relationships were showed than before, and a total of twelve major clades were recognized with highly supported values. In the classification system suggested by Li et al. (2016), six subfamilies (Ajugoideae, Lamioideae, Nepetoideae, Prostantheroideae, Scutellarioideae, and Symphorematoideae) suggested in the previous studies have

remained. In Viticoideae, only one clade was consistent with the previous studies (Bendiksby et al., 2011; Bramley et al., 2009) and the other clades were suggested as other subfamilies. As a result, five new clades were comprised as phylogenetically defined clades: Calliprostantherina, Cymalamiina, Perolamiina, Viticisymphorina, and Scutelamiina. Especially, Calliprostantherina formed a sister clade with the rest of the groups in the phylogenetic tree. Besides, three subfamilies were described: Cymarioideae, Peronematoideae, and Premnoideae. Cymarioideae included two genera *Acrymia* and *Cymaria* which phylogenetic position remained unclear. Premnoideae also included *Premna, Gmelina,* and *Cornutia.* Peronematoideae includes *Garrettia, Hymenopyramis, Peronema,* and *Petraeovites.* However, *Callicarpa* and *Tectona* did not belong to any subfamily. Later, Li and Olmstead (2017) gave them new subfamilies: Callicarpoideae and Tectonoideae. Callicalloideae was a sister to Prostantheroideae with strong support value, and Tectonoideae was a sister to a clade, including Lamioideae, Cymarioideae, Scutellarioideae, Peronematoideae, Ajugoideae, and Premnoideae in the phylogenetic tree (Fig. 2). Therefore, Lamiaceae currently includes 12 subfamilies (Table 1, Fig. 2): Ajugoideae, Callicarpoideae, Cymarioideae, Lamioideae, Nepetoideae, Peronematoideae, Premnoideae, Prostantheroideae, Scutellarioideae, Symphorematoideae, Tectonoideae, and Viticoideae.

According to previous studies by Li et al. (2016) and Li and Olmstead (2017), each subfamily shows synapomorphic

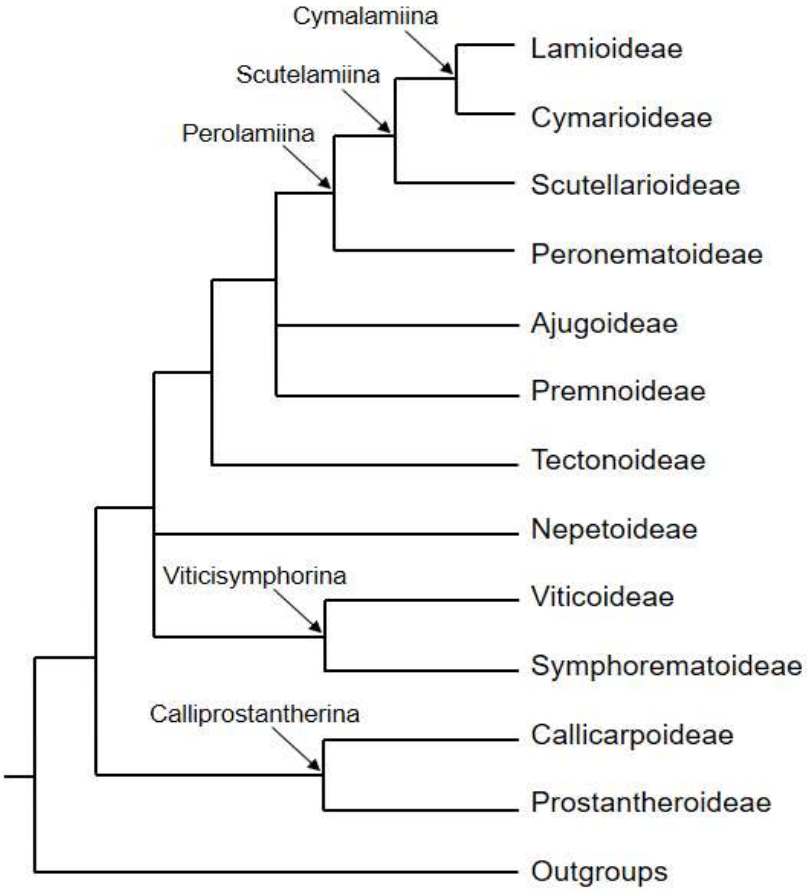


Fig. 2. A recent phylogenetic study of Lamiaceae (Li et al., 2016) based on five chloroplast DNA regions (*matK*, *ndhF*, *rbcL*, *rps16*, and *trnL−F*). Arrows indicate phylogenetically defined clades by Li et al. (2016).

characteristics. The synapomorphies of Ajugoideae are the pollen exine with spines and the branched exine. Callicarpoideae, unlike other subfamilies, the shape of stigma is peltate or capitate and has fruit with four pyrenes. Cymarioideae indicated synapomorphic characteristics in the structure of inflorescences that are cyme, axillary, loose, and have long peduncles. Lamioideae is distinguished from other subfamilies that all species have a gynobasic style. In Nepetoideae, pollen is mainly hexacolpate and three-celled. Peronematoideae have a woody stem, unlobed ovary, nectar disk not developed, or absent. Premnoideae has drupaceous fruit with one four-seeded pyrene, and Prostantheroideae has a dry schizocarp with albuminous seeds divided into four one-seeded mericarps. Scutellarioideae have tuberculate or elongate pericarps, and xylem fibers exist on the calyx. All species belonging to Symphorematoideae are climbers, and inflorescences are 3-7-flowered capitate cymes. In Tectonoideae, the flowers and calyx are actinomorphic, and have 5-7-lobed corolla. Viticoideae did not suggest morphological characteristics distinguished from other subfamilies, but the phytochemical was described as synapomorphies (Pedersen, 2000). Pedersen (2000) suggested that further studies are needed to find more obvious morphological synapomorphies of Viticoideae.

II. *Scutellaria*

1. Background of the genus

Scutellaria is included in the subfamily Scutellarioideae, along with four genera *Holmskioldia, Renschia*, *Tinnea,* and *Wenchengia* (Fig. 3). Except for *Scutellaria* and *Tinnea,* three genera (*Holmskioldia, Renschia,* and *Wenchengia*) are monotypic which consisting of only one species (*Holmskioldia sanguinea*, *Renschia heterotypica*, and *Wenchengia alternifolia*). Scutellarioideae is classified by the morphological synapomorphies which are pericarps with tuberculate or elongate processes (Li et al., 2012), high densities of xylem fibers in the calyces (Ryding 2017), and racemose or cymose inflorescences. The monophyly of Scutellarioideae has been recognized by molecular phylogenetic studies (Li et al., 2016; Zhao et al., 2020).

Scutellaria is the second largest genus in Lamiaceae containing approximately 470 taxa (The Plant List; http://www.theplantlist.org/). Most taxa are traditionally well known medicinal plants used for sleep disorders, anxiety, and strokes. Most plants distribute cosmopolitan, mainly in the temperate regions. It is known that approximately 15 taxa are distributed in Korea (Kim, 2018; Paton, 1990). Plants are annual or perennial herbs, or rarely subshrub and aquatic herbs. Leaves are arranged opposite, petiolate, entire, toothed, or rarely pinnatifid. Inflorescences are terminal or axillary racemes or spikes. Flowers

are arranged opposite, or secund in a one-sided. Their common name 'Skullcap' refers to the feature of the flower. Corolla is zygomorphic and usually 2-lipped, and 4-lobed. Corolla tube is straight or curved upwards at the base, and the interior of the corolla tube is glabrous. Anterior stamens are longer than the posterior. Nutlets are usually four, ovoid with a small scar. Besides these general characteristics, *Scutellaria* is distinguished from other genera in Lamiaceae by several morphological characteristics (Kim and Lee, 1995b; Paton, 1990). Typically, their calyx is bilabiate, the upper calyx lip has a shield-like structure called 'scutellum' derived from the Latin word 'scutella' meaning a dish. Lower calyx lips are larger than upper calyx lips, and the feature of upper calyx lips is galeate enclosing the anthers. Gynoecium has a peg-like gynophore. Anterior stamen is usually dimidiate, and these are unilocular thecae because of the abortion of the upper locule. Fruits are nutlets, and the surface of nutlets are smooth or papillate. Papillae are rarely forming a band or wing on the nutlet surface. An embryo in seeds is commonly incumbent or rarely bent. The arrangement of the inflorescence has been used for infrageneric delimitation. In addition, the patterns and types of the epidermal cell on nutlets were also considered important characteristics of the infrageneric classification system (Kim and Lee, 1995a; Paton, 1990).

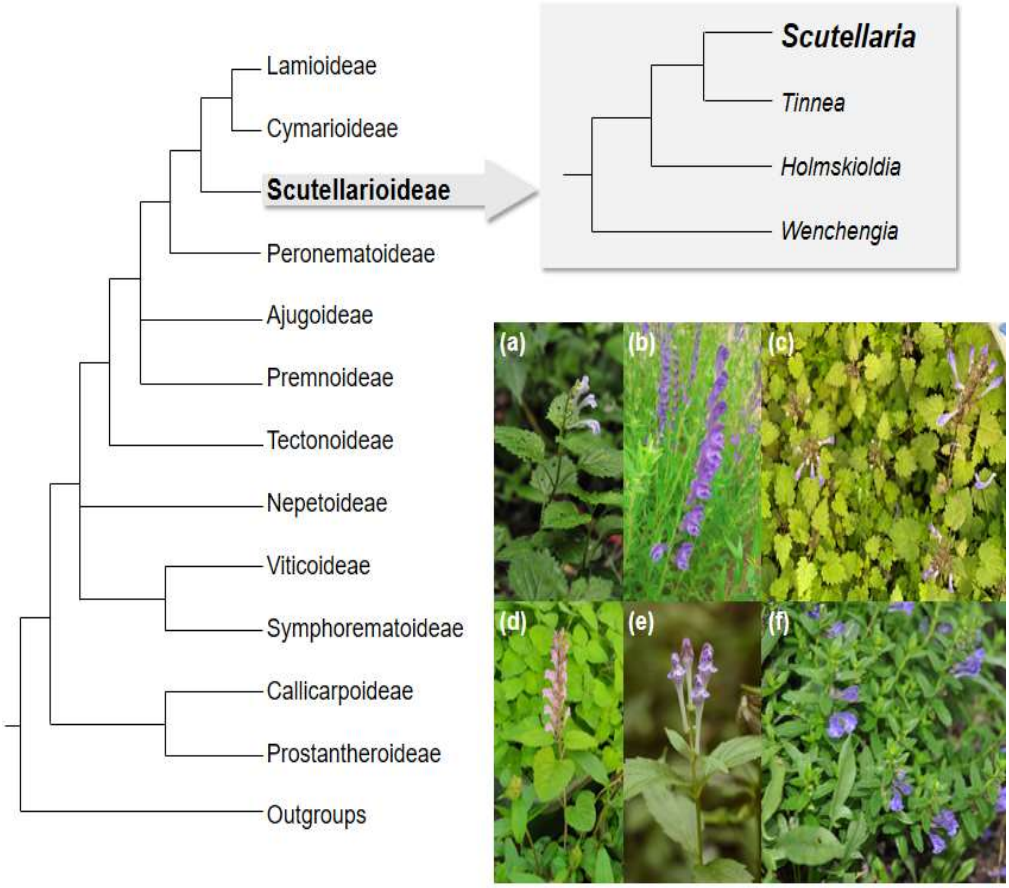
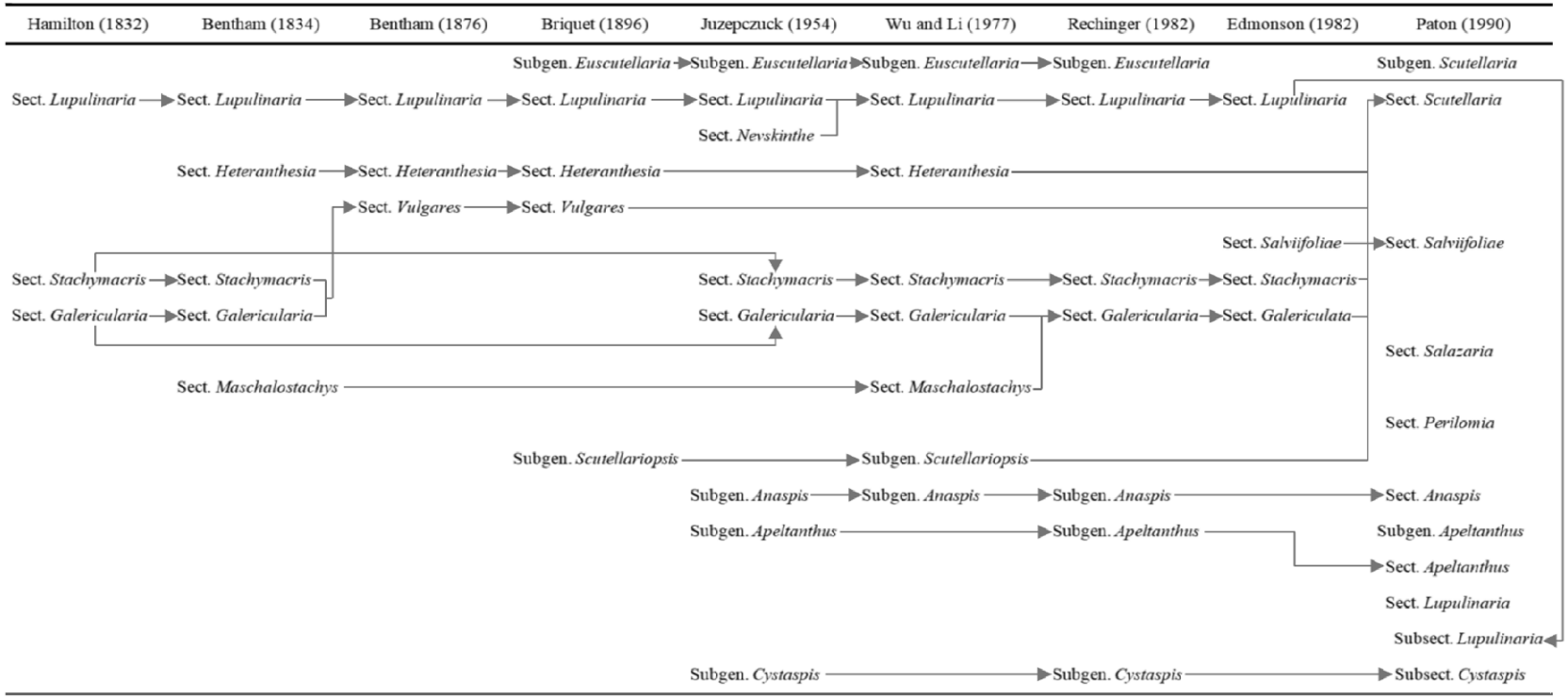


Fig. 3. Phylogenetic position of *Scutellaria* (grey box) in the subfamily Scutellarioideae (Li et al., 2016) based on recent molecular phylogenetic study (Zhao et al., 2020). The pictures (a–f) are some plants of *Scutellaria* that distribute in Korea. (a) *S. indica* L.; (b) *S. baicalensis* Georgi; (c) *S. pekinensis* var. *alpina* (Makino) H. Hara; (d) *S. indica* var. *coccinea* S. T. Kim & S. Lee; (e) *S. insignis* Nakai; (f) *S. strigillosa* Hemsl.

2. Classification history in *Scutellaria*

The morphological studies for the infrageneric classification system of *Scutellaria* have been well studied (Table 2). Hamilton (1832) suggested the first classification system of *Scutellaria,* and three sections were divided by characteristics of flower and inflorescence: section (sect.) *Galericularia*, *Lupulinaria*, and *Stachymacris*. Bentham (1834) accepted most of Hamilton's classification system and included two additional sections: sect. *Maschalostachys* with axillary inflorescences and sect. *Heteranthesia* with spirally inflorescences. Later, Bentham (1876) combined sects. *Galericularia* and *Stachymacris* into sect. *Vulgares*, and remained the previous three sections (sects. *Lupulinaria, Heteranthesia,* and *Vulgares*) according to plant distribution. Briquet (1896) first recognized subgenus in *Scutellaria,* and divided into two subgenera which were identified by the characteristics of nutlets: subgenus (subgen.) *Euscutellaria* and subgen. *Scutellariopsis*. Subgen. *Euscutellaria* included three sections of Bentham's classification system (sects. *Lupulinaria, Heteranthesia,* and *Vulgares).* Briquet's classification system (Briquet, 1896) had been widely used as the global treatment of *Scutellaria,* and most of the subsequent studies were based on this classification system. Juzepczuk (1954) recognized four subgenera according to characteristics of scutellum and calyx: subgen. *Euscutellaria, Anaspis, Apeltanthus*, and *Cystapis.* Subgen. *Euscutellaria* included sects. *Lupulinaria, Stachymacris,* and *Galericularia* of Hamilton's classification system

Table 2. The classification history of *Scutellaria*. Each row of the table includes the same taxonomic whatever their ranks are and arrows indicate combined, divided, or transpositioned groups.



(Hamilton, 1832) and added sect. *Nevskinthe* which is similar features with sect. *Lupulinaria.* Wu and Li (1977) remained subgen. *Anaspis* of Juzepczuk's classification system (Juzepzuk, 1954), and followed Briquet's subgenera (subgen. *Euscutellaria* and *Scutellariopsis*; Briquet, 1896). Five sections of subgen. *Euscutellaria* were followed Bentham's classification system (Bentham, 1834). Rechinger (1982) adopted Juzepczuk's classification system (Juzepzuk, 1954), but sect. *Nevskinthe* was excluded. Edmonson (1982) established a classification system based on the characteristics of flowers and inflorescences, and remained three sections (sects. *Lupulinaria, Stachymacris,* and *Galericularia)* following with the previous studies (Hamilton, 1832; Bentham, 1834), and added sect. *Salviifoliae*.

Previous classification systems were established based on geographically limited sampling, and a globally consistent classification system had not been clearly suggested. The most recent classification system of *Scutellaria* based on the morphological study was suggested by Paton (1990), which includes 110 global species (Table 3). This classification system is divided into two subgenera (subgen. *Apeltanthus* and *Scutellaria)* according to the characteristics of inflorescence and bracts. Subgen. *Apeltanthus* is four-sided inflorescence and flowers subtended by cucullate bracts. In contrast, subgen. *Scutellaria* is one-sided or rarely spiral inflorescence and flowers are subtended by leaves or leaf-like bracts. Subgen. *Apeltanthus* is divided into two sections

based on their habitats: sects. *Apeltanthus* and *Lupulinaria*. Sect. *Apeltanthus* is chasmophytes and lower leaves have not petiole, and sect. *Lupulinaria* is not chasmophytes and lower leaves have petiole. Sect. *Lupulinaria* includes two subsections (subsects. *Cystaspis* and *Lupulinaria),* these sections are distinguished by characteristics of scutellum. The scutellum of subsect. *Cystaspis* is membranous-inflated, and subsect. *Lupulinaria* is thick-textured. In subsect. *Lupulinaria,* four species groups (sp-gr.) are included. These species groups are classified by the characteristics of bracts, rhizome, and leaves. Subgen. *Scutellaria,* approximately 275 species are included, and it is divided into five sections: sects. *Scutellaria*, *Salviifoliae*, *Salazaria*, *Perilomia*, and *Anaspis.* Sect. *Scutellaria* is the largest section including 240 species and 34 species groups. This section is erect or prostrate herbs, and suffruticose or not. Thirty-four species groups in sect. *Scutellaria* are distinguished by various characteristics. Sect. *Salviifoliae* is suffruticose mat- or cushion-forming herbs, and prostrate or sacending. Sect. *Salazaria* is a densely branched shrub to one-meter tall, and stems are round or round-quadrangular. Sect. *Perilomia* is a shrub to four-meter tall, and it has corolla lips with equal size and shape, and the upper lip is not galeate. Sect. *Anaspis* is suffruticose chasmophyte. Since the classification system was established by Paton (1990), this classification system has been most widely used, which has been the basis for understanding infrageneric relationships of *Scutellaria.*

Table 3. The detailed infrageneric classification system of *Scutellaria* (Paton, 1990) including "species group", which is an unofficial rank under subsection

Subgenus	Section	Subsection	Species group
Apeltanthus			
	Apeltanthus		
	Lupulinaria		
		Lupulinaria	
			COLPODEA
			GLECHOMOIDES
			LINEARIS
			ORIENTALIS
		Cystapis	
Scutellaria			
	Anaspis		
	Salazaria		
	Salviifoliae		
	Perilomia		
	Scutellaria		
			ALBIDA
			ANGUSTIFOLIA
			BALEARICA
			CAERULEA
			COSTARICANA
			DISCOLOR
			FORRESTII
			GALERICULATA
			GUILIELMII
			HASTIFOLIA
			HAVANENSIS
			HINTONIANA
			HUMILIS
			INCANA
			LATERIFLORA
			LONGITUBA
			LUTEA
			NUMMULARIIFOLIA
			OVATA
			PARVULA
			RACEMOSA
			REPENS
			RESINOSA
			SELERIANA
			SPECIOSA
			STRIGILLOSA
			SUFFRUTESCENS
			TUBEROSA
			TUBIFERA
			ULIGINOSA
			UTRICULATA
			VIOLACEA
			VIOLASCENS
			YUNNANENSIS

3. Recent molecular phylogenetic studies and the purposes of this study

The recent molecular phylogenetic study *Scutellaria* has been performed using internal transcribed spacer (ITS) and external transcribed spacer (ETS) sequences in nuclear ribosomal DNA (nrDNA) based on 41 Chinese taxa (Zhao et al., 2017). This study was conducted on Chinese taxa and analyzed based on Li and Hedge's classification system suggested in Flora of China (Li and Hedge, 1994). As an updated edition of Lamiaceae in Flora of China, Li and Hedge divided *Scutellaria* (98 spp.) into seven groups on the basis of morphological characters of inflorescence, bract, calyx, and nutlet. As a result of comparing the obtained phylogenetic tree with Li and Hedge's classification system (Li and Hedge, 1994), it was described that all sections and groups appeared the paraphyletic groups. In that study, it was described that further studies including more species and different regions are needed. Besides, taxa included in Zhao et al. (2017) were overlapped only 24 taxa compared with the current classification system (Paton, 1990). It was confirmed that only the small subclade including subgen. *Apeltanthus* was a monophyletic group and subgen. *Scutellaria* was a paraphyletic group.

Safikhani et al. (2018) was performed a molecular phylogenetic study focusing on 42 Iranian taxa using nuclear ITS and chloroplast *trnL−F* region. Only nine taxa included in Safikhani et al. (2018) were overlapped with the current classification system (Paton, 1990), and it was also difficult to compare with the phylogenetic

relationship. Both studies (Safikhani et al., 2018; Zhao et al., 2017) had limited sampling, and it was difficult to compare with the current classification system (Paton, 1990). Also, both studies did not suggest a clear phylogenetic classification. Recently, a phylogenomic study of subfamily Scutellarioideae was performed (Zhao et al., 2020), and eleven taxa of *Scutellaria* were included. In that study, chloroplast genomes of 12 taxa newly reported. It was also confirmed that subgen. *Apeltanthus* was a monophyletic group and subgen. *Scutellaria* was a paraphyletic group same as in the previous study (Zhao et al., 2017).

The previous molecular phylogenetic studies have been difficult to understand the phylogenetic relationship of *Scutellaria* due to regional sampling, and re-evaluation of previous studies has been needed based on more extensive studies. To understand the phylogenetic relationships of *Scutellaria* in the world, 1) chloroplast *ccsA, matK,* and *ndhF* sequences for proper representatives were determined, 2) chloroplast genomes were completed and super-matrix phylogeny was performed, 3) the data for a large amount of single-copy nuclear genes was generated, and evolutionary events such as ancient hybridization was investigated, and 4) the structural variation of nrDNA and secondary structure of ITS2 were identified as an independent study. All phylogenetic trees were compared and re-examined with the current classification system based on morphological study (Paton, 1990). The final aim is to the suggestions for the new classification system based on these phylogenomic studies.

CHAPTER 2.

Phylogenetic relationships in *Scutellaria* inferred from three chloroplast genes (*ccsA*, *matK*, and *ndhF*)

I. Introduction

Genes in chloroplast DNA have been widely used to understanding the phylogeny of *Scutellaria* and Lamiaceae (Kaufmann and Wink, 1994; Wink and Kaufmann, 1996; Wagstaff and Olmstead, 1997; Wagstaff et al. 1998; Harley et al., 2004; Bendiksby et al., 2013; Li et al., 2016; Zhao et al., 2020). Especially, chloroplast genes such as *rbcL*, *rps16*, *trnL, trnL-F, matK*, and *ndhF* have been widely used for phylogenetic analyses. These regions have contributed to resolving the evolutionary relationship of plants as useful molecular markers.

The phylogenetic studies of *Scutellaria* were mainly conducted based on morphological and cytological studies (Bentham, 1834; Briquet, 1896; Edmonson, 1982; Hamilton, 1832; Juzepzuk, 1954; Lane, 1983; Paton, 1990; Rechinger, 1982; Wu and Li, 1977), and the molecular phylogenetic studies of *Scutellaria* in the world has not been performed yet. Recently, Zhao et al. (2017) performed a molecular phylogenetic study used internal and external transcribed spacer (ITS and ETS) in nuclear ribosomal DNA (nrDNA) sequences to confirm the phylogenetic position of *S. wuana*. A total of 51 Chinese taxa were included in Zhao et al. (2017), and 24 taxa were overlapped with the current classification system (Paton, 1990). As a result of phylogenetic analyses, the monophyly of *Scutellaria* was confirmed. However, subgenus (subgen.) *Scutellaira* was a paraphyletic group. Subgen. *Apeltanthus* was a monophyletic group, but only five taxa

included. Therefore, the phylogenetic relationships among subgroups of *Scutellaria* could not clearly confirmed in that study (Zhao et al., 2017). Later, Safikhani et al. (2018) conducted a molecular phylogenetic study based on nuclear ITS and chloroplast *trnL−F* regions. According to the current classification system (Paton, 1990), 29 taxa were overlapped. Unlike Zhao et al. (2017), Safikhani et al. (2018) showed that subgen. *Apeltanthus* and *Scutellaria* were a monophyletic group in phylogenetic tree based on ITS region, but subgen. *Scutellaria* was a paraphyletic group in *trnL−F* tree. However, the taxa of subgen. *Scutellaria* included only 11 taxa, so it was difficult to confirm the clear phylogenetic relationships of *Scutellaria.* Therefore, both studies (Safikhani et al., 2018; Zhao et al., 2017) could not clearly showed phylogenetic relationships of *Scutellaira* because of narrow geographical sampling and a lack of variability within data. Zhao et al. (2020) recently completed chloroplast genomes of 12 taxa in subfamily Scutellarioideae, including 11 taxa of *Scutellaria.* According to Zhao et al. (2020), the sequence divergence and nucleotide diversity of *Scutellaria* were analyzed in chloroplast genomes of 11 taxa of *Scutellaria,* and ten variable regions were suggested: *ndhF, ycf1* genes, and eight intergenic spacer regions (*trnH−psbA, trnK−rps16 intron, petN−psbM, rbcL−accD, petA−psbJ, petB−petD, intron, rpl32−trnL, and rps15−ycf1*). These regions were identified as “hotspot” for further DNA barcoding research and phylogenetic study. However, the analysis of this chapter had conducted before

published Zhao et al. (2020)'s study and did not reflect the region suggested by Zhao et al. (2020).

Since the global phylogenetic study has not been performed on *Scutellaria,* this chapter will clarify the phylogenetic relationship through expanded analysis compared to the previous studies and compare it extensively with the morphological classification system suggested by Paton (1990). A total of 200 taxa were included representing each subgroup of *Scutellaria* and also containing questionable unplaced taxa, which account for approximately 43% of the recognized taxa in *Scutellaria.* To address the phylogeny of the *Scutellaria* in the world, three fast-evolving genes in chloroplast (*ccsA*, *matK*, and *ndhF*) were selected based on the analysis of the two previously reported chloroplast genomes in the genus (*S. baicalensis*, NC_027262.1; *S. insignis,* NC_028533.1). This study will provide the first global molecular phylogenetic relationships among taxa of *Scutellaria,* and it will also be helpful for establishing the new classification system in the genus.

II. Materials and Methods

1. Taxon sampling

Comparing with the current classification system (Paton, 1990), a total of 200 taxa representing all subgenera and sections (sect.) are included, and thirty-four out of 38 (89%) species-groups (sp-gr.) were included (Table 4). Some samples were obtained from the field (Korea and China) and others are from major herbaria (BM, K, KH, KUN, KYO, SWU, MO, PE, and WTU). *Holmskioldia sanguinea* was also included as an outgroup which is a sister taxon in the previous molecular study of Lamiaceae (Li et al., 2016). The taxa list, voucher information, and GenBank accession number are present in Appendix 1.

2. DNA extraction, polymerase chain reaction, and sequencing

All samples were extracted using the GeneAll Plant SV Mini Kit (GeneAll Biotechnology Co. Ltd, Seoul, Korea) following the manufacturer's protocol. The concentration of extracted DNA was estimated using a NanoDrop 2000 (Thermo Scientific, Wilmington, USA) and diluted to a concentration of 10 ng/㎕.

To find fast-evolving genes for the phylogenetic study of *Scutellaria,* two previously reported chloroplast genomes (*S.*

baicalensis and *S. insignis*) in GenBank were analyzed. In each gene or intergenic spacer, the length in alignment and number of variable sites were obtained. After that, filtered-out regions of less than 500 bp in length, having over 5 % of the gap, and less than 5 % of sequence variations. Finally, sorted by a portion of variable sites in the entire sequence, and selected top-ranked three genes (*ccsA*, *matK*, and *ndhF*; Table 5).

Table 6 represents the information of primers using polymerase chain reaction (PCR). Fig. 4 shows the flowchart for finding fast-evolving chloroplast DNA regions in this chapter. Each pair of primers was newly designed to obtain a PCR product, and the PCR product length was approximately 0.8-1.0 kbp. The PCR was conducted with a total volume of 20 ㎕ containing 1 ㎕ template DNA, 10 ㎕ Master Taq (2X PCR Master mix Solution i-Taq, intron, Seoul, Korea), 1 ㎕ of each primer, and 7 ㎕ of distilled water. The PCR reaction cycle was performed using S1000 Thermal cycler (BioRad, Hercules, USA) with the following program: 3 min pre-denaturation at 94 ℃ followed by 35 cycles of 30 s denaturation at 95 ℃, 30 s annealing at 55 ℃, and 45 s extension at 72 ℃, and finished by a final extension of 7 min at 72 ℃.

Table 4. Number of taxa in each subgroup based on the current classification system (Paton, 1990) and that included in this study

Subgenus	Section	Subsection	Species group	No, of taxa included in this study	No. of taxa included in Paton (1990)
Apeltanthus				20	142
	Apeltanthus			4	12
	Lupulinaria			16	131
		Lupulinaria		15	120
			colpodea	0	3
			glechomoides	0	1
			linearis	1	1
			orientalis	14	104
		Cystapis		1	11
Scutellaria				98	275
	Anaspis			4	15
	Salazaria			1	1
	Salviifoliae			2	5
	Perilomia			2	11
	Scutellaria			89	240
			albida	8	11
			angustifolia	8	8
			balearica	1	1
			caerulea	1	5
			costaricana	1	2
			discolor	1	1
			forrestii	2	2
			galericulata	4	7
			guilielmii	1	1
			hastifolia	1	1
			havanensis	2	3
			hintoniana	1	1
			humilis	4	4
			incana	12	15
			lateriflora	1	2
			longituba	1	1
			lutea	1	1
			nummulariifolia	1	1
			ovata	3	5
			parvula	3	3
			racemosa	1	2
			repens	4	4
			resinosa	6	19
			seleriana	1	1
			speciosa	4	6
			strigillosa	5	5
			suffrutescens	1	1
			tuberosa	1	1
			tubifera	0	1
			uliginosa	4	9
			utriculata	0	1
			violacea	2	3
			violascens	2	4
			yunnanensis	1	1

Table 5. List of fast-evolving chloroplast DNA regions that are useful for the phylogenetic study of *Scutellaria*

Region name	Length in alignment	No. of variable sites	Portion of variable sites in complete deletion (%)	Pairwise distance
ndhF (3' half)	1105	56	5.07	0.053
ndhF (total)	2155	70	3.25	0.034
matK (including trnK intron)	2589	55	2.12	0.022
ccsA	975	20	2.05	0.021
matK (3' half)	1545	31	2.01	0.021
matK (5' half)	910	13	1.43	0.015
accD	1479	22	1.49	0.015
ndhA	2115	30	1.42	0.015
ndhA (3' half)	1057	21	1.99	0.020
ndhA (5' half)	1057	9	0.85	0.009
ndhF (5' half)	1050	14	1.33	0.013
cemA	690	9	1.30	0.013
intergenic_*ycf4-cemA*	1545	20	1.29	0.016
rbcL	1452	18	1.24	0.013
rps3	663	7	1.06	0.011
rpl16	1310	13	0.99	0.011
rps2	711	7	0.98	0.010
atpF	1237	11	0.89	0.009
rpoC2	4206	35	0.83	0.008
trnV-UAC	653	5	0.77	0.008
ndhG	531	4	0.75	0.008
intergenic_*petA-psbJ*	1097	7	0.64	0.006
intergenic_*trnH_psbA*	1144	7	0.61	0.006
intergenic_*psbE-petL*	359	1	0.28	0.003

Table 6. The information of primers used in this chapter

DNA region	Primer name	Primer sequence 5'→3'
ccsA	ccsA_9F	ACTCATATTTCTTTTTCGATCGTTT
	ccsA_10R	TAGACGCCACAATTGCTGAA
	ccsA_11F	CCTTTTTAGTCGATGAAATCGT
	ccsA_4R	TTAATGTAAATGAACCATAACTATG
matK	matK_909F	GAAYAATWGGAACAARRGTMTC
	matK_1998R	AGTGCGATACAGTMAAAACAA
	matK_2FF	CTADAATTTGACTCCGTACYAC
	matK_2RR	ATAGRAARTCGTGTTGGTWAGATC
	matK_3FF	ATGCCGCAATCAGAGGAATA
	matK_3RR	TTCGAGTAATTAACCGTTTCACAA
	matK_4FF	GGAAAATCCGGGTTATGACA
	matK_4RR	CYATTCCTTTGAATTTWTGGG
	matK_5FF	GGCGTGCAACAAGAATTTG
	matK_5RR	CCTTCTTGGATAAAACCACAGC
ndhF	ndhF_1F	GTATTCCACCGCTTGCTTGT
	ndhF_2R	CATGACACCAACTCCATTCG
	ndhF_3F	CTTCGCCGATTTTCGCTAT
	ndhF_4R	CCCGCYGCTAAAGGATAATG

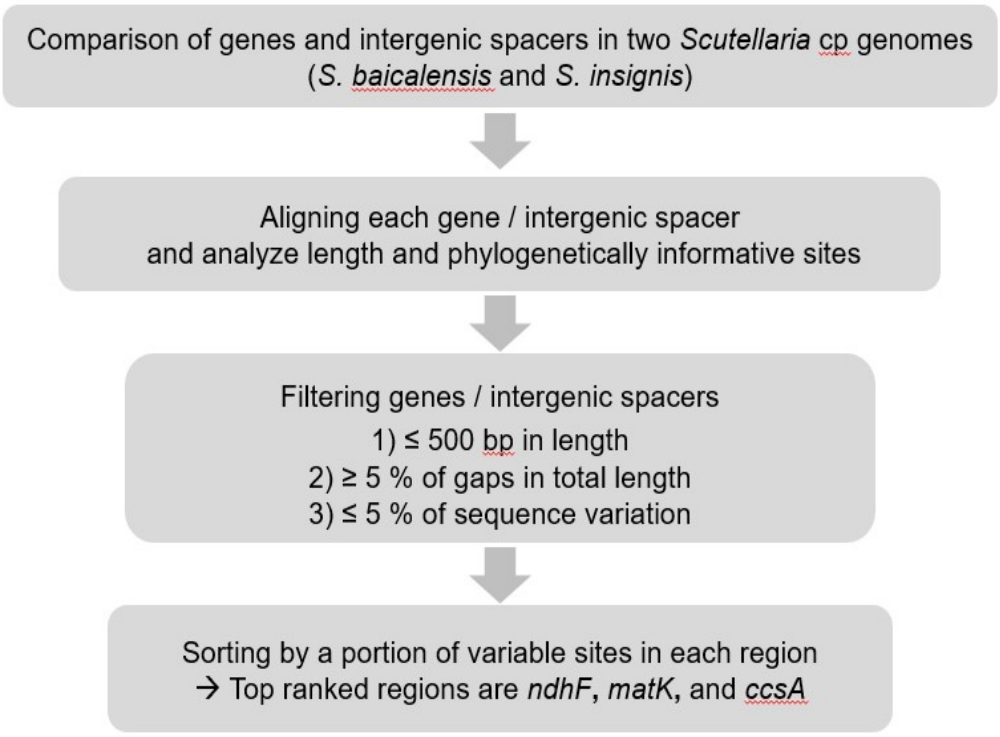


Fig. 4. A flowchart for finding fast-evolving chloroplast DNA regions for phylogenetic study.

3. Phylogenetic analyses

Sequences were assembled and edited using Sequencher 4.9 (Gene Code Corporation, Ann Arbor, USA). Sequence alignments for each region were performed using CLUSTALW (Thompson et al., 2003) with default options in MEGA 7 (Kumar et al., 2016) and manually adjusted by eye. After alignments for each region, the analysis of each sequence was performed using MEGA 7. Three different datasets were combined for phylogenetic analysis, and the suitable substitution model was determined by using MEGA 7 before Maximum-Likelihood (ML) analysis and Bayesian inference (BI) analysis. The GTR + I + Γ was selected as the best model, and ML analysis was conducted with RAxML GUI v.1.5 (Silvestro and Michalak, 2012) using the rapid bootstrap option with 500 bootstrap replications. Bayesian inference (BI) analysis was executed using MrBayes v3.2.7 (Ronquist et al., 2012). The Markov chain Monte Carlo (MCMC) algorithm was run for 10,000,000 generations with four chains, and trees were recorded to files every 100 generations. The first 25 % of sampled generations were discarded as burn-in, and a 50 % majority-rule consensus tree was obtained using the remaining trees. Finally, the obtained each phylogenetic tree was visualized in MEGA7 (Kumar et al., 2016) and Figtree v.1.4.4 (Rambaut, 2018).

III. Results

1. Sequence analyses

The result of sequence analysis for the *ccsA, matK,* and *ndhF* genes with 200 species in *Scutellaria* is summarized in Table 7. The final length of aligned sequences of the three genes was 2,475 bp. The length of aligned *ccsA* sequences was 696 bp, variable sites were 160 bp (23 %), and informative characters were 92 bp (13.2 %). The length of aligned *matK* sequences was the longest (1,023 bp), but variable sites had the lowest with 228 bp (22.3 %), and informative characters also were the lowest with 118 bp (11.5 %). The length of aligned *ndhF* sequences was 756 bp, variable sites were 241 bp (31.9 %), informative characters were 144 bp (19.8 %).

Table 7. Statistical summary of sequence for three fast-evolving genes in *Scutellaria*

	ccsA	*matK*	*ndhF*	Total
Aligned length (bp)	696	1,023	756	2,475
Variable sites (bp)	160	228	241	629
Variable sites (%)	23	22.3	31.9	25.4
Informative characters (bp)	92	118	144	354
Informative characters (%)	13.2	11.5	19	14.3

2. Phylogenetic relationships based on three fast-evolving chloroplast genes tree

The ML tree based on three fast-evolving chloroplast genes for the 201 taxa including one outgroup taxon is shown in Fig. 5, and this phylogenetic tree is summarized based on taxonomic groups recognized in the current classification system (Paton, 1990; Fig. 6). The BI tree is shown in Appendix 2. The topologies of ML tree and BI tree were congruent, and the relationship between the taxa was investigated focused on the ML tree, and bootstrap support (BS) values and posterior probability (PP) values were indicated on the node of the ML tree. Each subgenus, sections, species groups in this tree were indicated based on the current classification system (Paton, 1990; Fig. 5).

According to the phylogenetic tree (Fig. 5), it was different compare to that of current classification system (Paton, 1990), and six major clades were recognized (clades #1-6). The monophyly of subgen. *Apeltanthus* clade was strongly supported (BS = 100; Fig. 5, clade #1), however, subgen. *Scutellaria* was a paraphyletic group.

Subclades in the subgen. *Apeltanthus* were not clearly recognized: both sect. *Apeltanthus* and sect. *Lupulinaria* was not a monophyletic group (Figs. 5 and 6). Some taxa belonging to both sections were located in different clades. In the current classification (Paton, 1990), four species groups were included in subgen. *Apeltanthus,* and only two species groups (sp-grs.

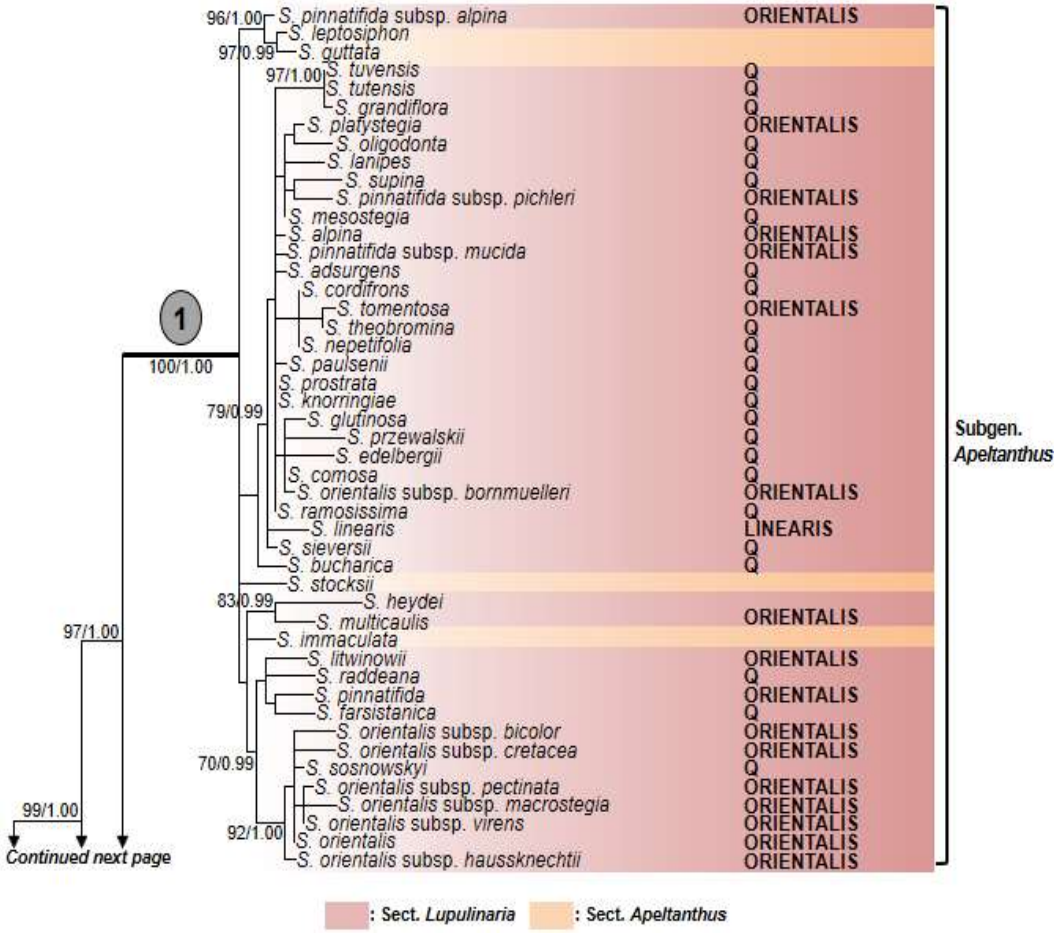


Fig. 5. A maximum likelihood tree based on chloroplast *ccsA, matK,* and *ndhF* genes. Numbers above the branches are support values and numbers only BS ≥ 50% or PP ≥ 0.90 are indicated (BS/PP). Circled numbers above the nodes (#1−6) correspond to major clades described in the text. Colored boxes are sections and capitalized names after the taxon names are "species group" recognized in Paton (1990). "Q" indicates a taxon of which taxonomic position (section and "species group") have not been reported.

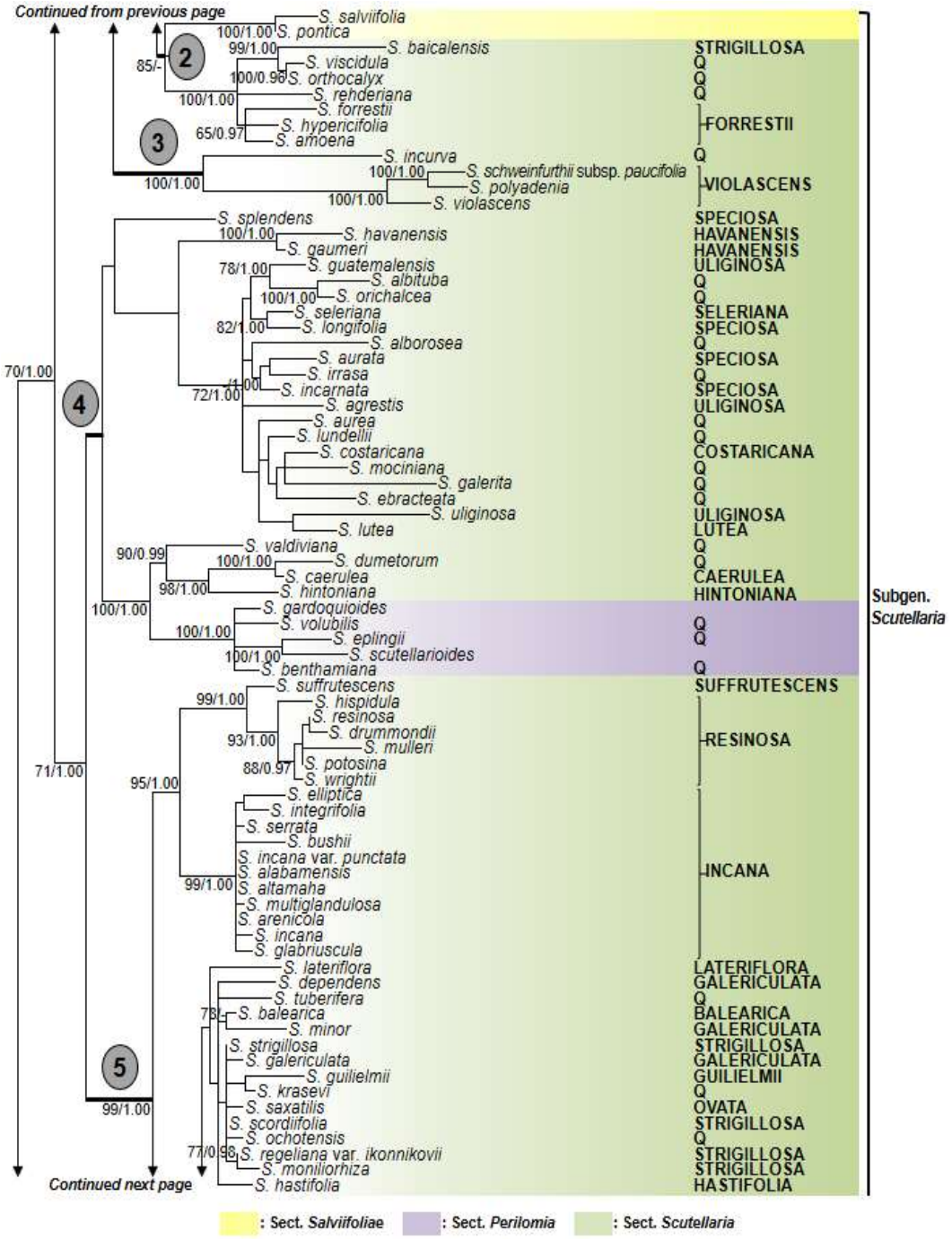


Fig. 5. Continued.

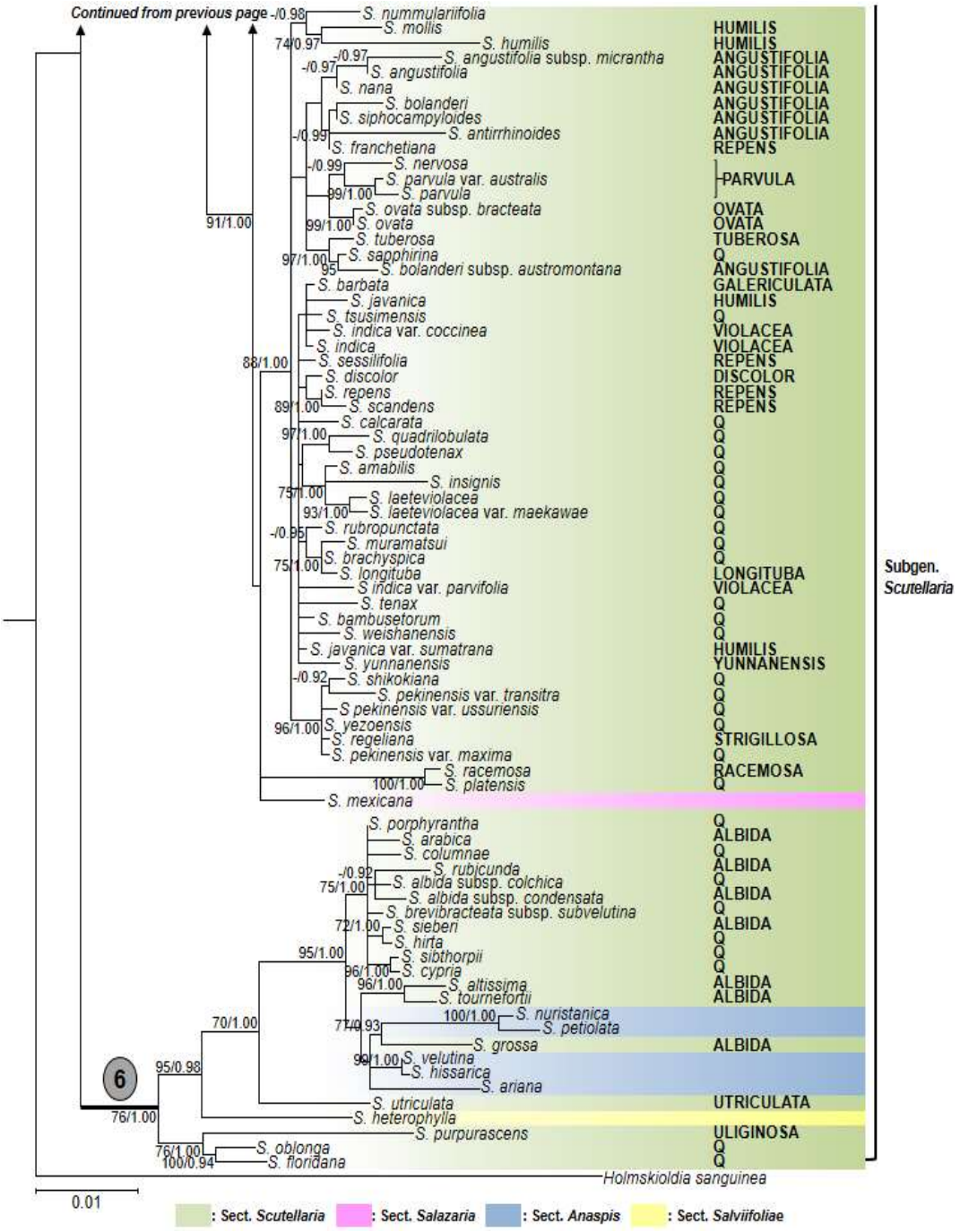


Fig. 5. Continued.

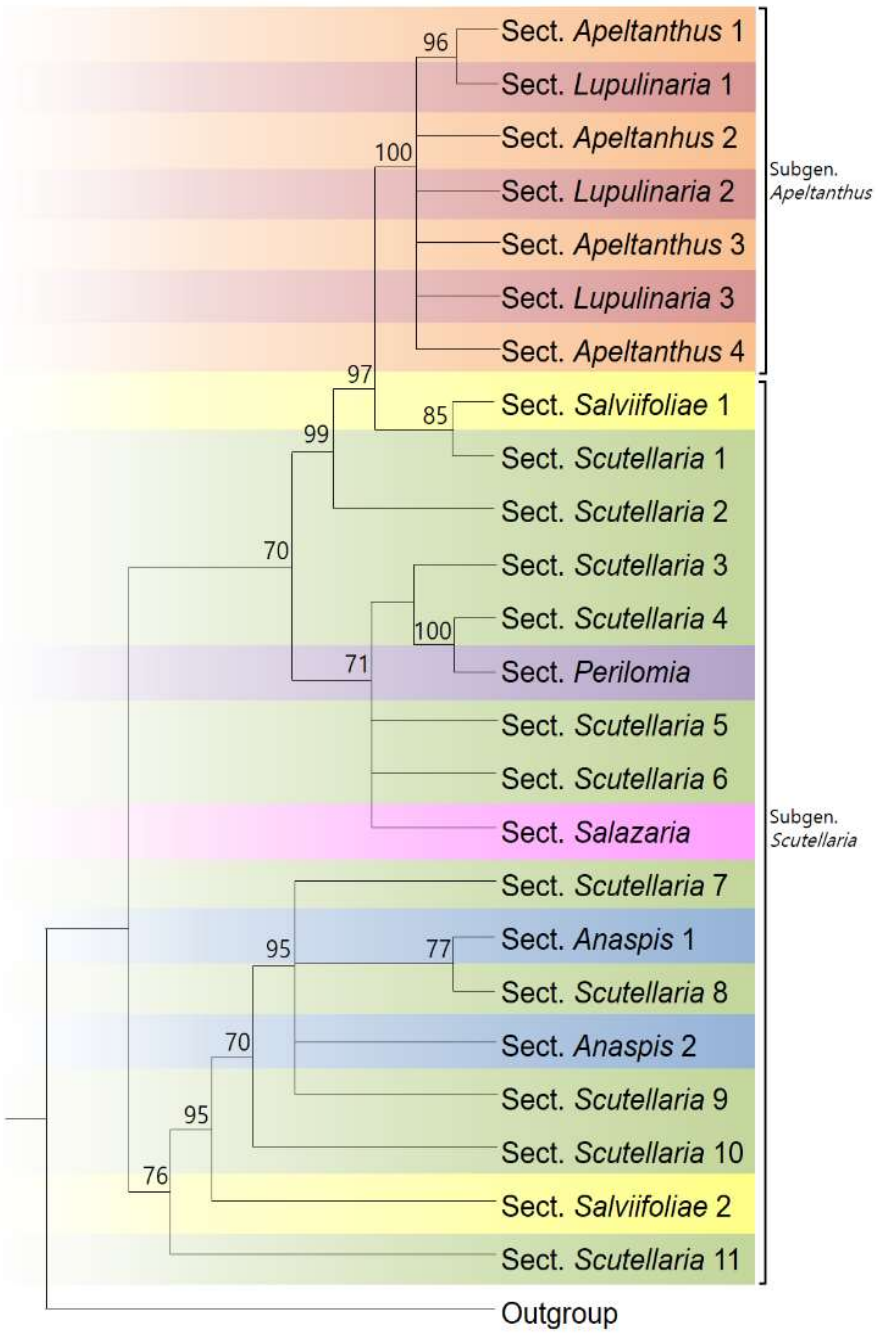


Fig. 6. A summarized ML tree based on taxonomic groups recognized in the current classification system (Paton, 1990). Only nodes over 70 % bootstrap value are recognized.

ORIENTALIS and LINEARIS) were included in the phylogenetic tree. Monotypic sp-gr. LINEARIS was included a part of sp-gr. ORIENTALIS, which means sp-gr. ORIENTALIS was clearly paraphyletic group.

Two groups of subgen. *Scutellaria* were form a grade as sisters to subgen. *Apeltanthus* (clade #2 and #3). The first group (clade #2) is a clade of sect. *Salviifoliae* and some members of sp-grs. FORESTII and STRIGILLOSA of sect. *Scutellaria,* which is a sister to subgen. *Apeltanthus* (BS = 85). Inside of this clade, sp-gr. FORRESTII of sect. *Scutellaria* was a monophyletic group but the supporting value was not high (BS = 65); sect. *Salviifoliae* was a highly supported monophyletic group (BS = 100). The second group (clade #3) contains members of sp-gr. VIOLASCENS and this was a sister to a clade of the first group plus subgen. *Apeltanthus* (BS = 100). Sect. *Salvifolliae* was placed two different places in the tree (yellow boxes in Fig. 5). In clades #4 and #5, various sections and species-groups remained unresolved in subgen. *Scutellaria*. Two major groups were recognized although one of them was highly supported (BS = 99; clade #4) but another was not (BS < 50; clade #5). Sect. *Perilomia* formed a clade as a monophyletic group inside of clade #4 (BS = 100), and their sister clade included sp-grs. HINTONIANA and CAERULEA. In clade #5, only three species-groups (RESINOSA, INCANA, PARVULA) were monophyletic groups. Sp-grs. RESINOSA and INCANA were shown the sister relationship and strongly supported (BS = 95). A

monotypic clade of sp-gr. PARVULA was weakly supported (BS < 50), and Sect. *Salazaria,* which was also a monotypic taxon, was weakly supported (BS < 50). At the base of the tree, clade #6 was a sister group of all other species (BS = 76). This clade includes sect. *Anaspis,* sp-gr. ALBIDA in sect. *Scutellaria,* and a species in sect. *Salviifoliae. S. utriculata* in sect. *Scutellaria* was a sister to sp-gr. ALBIDA and sect. *Anaspis* (BS = 70). *S. heterophylla* was strongly supported as sister to sp-gr. ALBIDA, *S. utriculata* of sect. *Scutellaria*, and sect. *Anaspis* (BS = 95). *S. purpurascens*, *S. oblonga, S. floridana* was sister to the remaining species of clade #6 (BS = 76).

IV. Discussion

1. Comparison between the phylogenetic tree and the current classification system

The phylogenetic tree obtained in this chapter is possible to compare extensively with the current classification system (Paton, 1990). Conspicuously, the phylogenetic tree (Figs. 5 and 6) clearly showed the monophyly of subgen. *Apeltanthus* and paraphyly of subgen. *Scutellaria* with highly supported values. Paton (1990) suggested various relationships between some species and species groups, but most of them were placed in totally different positions or showed the low supporting value in the phylogenetic tree (Fig.

5). These phylogenetic relationships show that the current classification system (Paton, 1990) is not supported in the phylogenetic tree.

Subgen. *Apeltanthis* divided into two sections (sects. *Apeltanthus* and *Lupulinaria*) according to whether sections are chasmophytes plant or not by Paton (1990). However, in clade #1 (Figs. 5 and 6), two sections were not monophyletic group, taxa of sect. *Apeltanthus* were scattered between sect. *Lupulinaria*. Sp-grs. LINEALIS and ORIENTALIS were described in Paton's study (Paton, 1990) as being classified by leaves shape, however, in sect. *Lupulinaria* of the tree, both species groups were not a monophyletic group. A clade #2 containing some species of sect. *Salviifoliae*, sp-grs. STRIGILLOSA, and FORRESTII were indicated as sister to subgen. *Apeltanthus*. Paton (1990) noted that sp-gr. FORRESTII is close to sp-grs. ALBIDA, VIOLACEA, and YUNNANENSIS. However, these groups were completely mixed in different clades in the tree. In clade #3, a monophyletic clade of sp-gr. VIOLASCENS was placed. Paton (1990) suggested a close relationship between *S. violascens* of sp-gr. VIOLASCENS and *S. alttisima* of sp-gr. ALBIDA, but these species were placed in different positions in the tree. In clade #4, sect. *Perilomia* was a monophyletic group and this sister group included sp-gr. HINTONIANA and CAERULEA. Paton (1990) recognized that Sect. *Perilomia* possibly evolved from Sect. *Scutellaria*, perhaps from the ancestors of the sp-grs. CAERULEA, LUTEA, HINTONIANA or

NUMMULARIFOLIA sometime in the last three million years, possibly in the Pleistocene in response to climatic changes. *S. nummulariifolia* also suggested that it has a similar embryo to those of members of Sect. *Perilomia.* This result showed that *S. nummulariifolia* was placed in a different clade with sect. *Parilomia,* and sect. *Perilomia* was closer to *S. caerulaea*, *S. lutea*, and *S. hintoniana* (BS = 100) than *S. nummlariifolia.* It showed that these relationships are mostly consistent with Paton's suggestion (Paton, 1990). In addition, Paton (1990) described the inflorescence of sp-gr. SPECIOSA is similar to sp-gr. UNLIGINOSA, and the corolla of sp-gr. SPECIOSA close to *S. costaricana*. The sister clade next to clade #4, although this clade was weakly supported, sp-grs. SPECIOSA, UNLIGINOSA, and COSTARICANA were grouped in the same clade. From the top of a clade #5, *S. suffrutescens* showed the closest relationship with sp-gr. RESINOSA, which was highly supported (BS = 99). This relationship is consistent with Paton's suggestion (Paton, 1990) that *S. suffrutescens* has a similar habit to sp-gr. RESINOSA. Sect. *Salazaria* in the base of the clade #5, which is monotypic taxon, was indicated as polytomy with some species of sect. *Scutellaria.* Therefore, further studies between sect. *Salazaria* and other species are needed. Paton (1990) noted that habit and corolla forms of sect. *Salazaria* are closely related to the sp-gr. RESINOSA, but sp-gr. RESINOSA was placed in a different clade in the tree. Sp-gr. PAVULA showed close to sp-gr. ANGUSTIFOLIA although the clade was weakly supported. Paton

(1990) was recognized that sp-gr. PARVULA related to the sp-gr. ANGUSTIFOLIA through the characteristics of *S. bolanderi.* Additionally, sp-grs. REPENS and VIOLACEA were shown paraphyletic groups, but it has respectively potential to be a monophyletic group because this clade of the tree is unresolved. In clade #6, sp-grs. ALBIDA, UTRICULATA, ULIGINOSA, sect. *Anaspis*, and sect. *Salviifoliae* were formed a clade. Paton (1990) suggested that the nutlet morphology of sp-gr. UTRICULATA relates to sp-gr. ALBIDA. In this clade, sp-gr. UTRICULATA was sister to a group including sp-gr. ALBIDA and sect. *Anaspis*. Sect. *Anaspis* was also suggested in Paton's study (Paton, 1990) that it was most likely evolved from the ancestors of sp-gr. ALBIDA and likely a monophyletic group indicating the apomorphy of a testudinate calyx. However, sp-gr. ALBIDA was not appeared an ancestor of sect. *Anaspis* in this clade, and it was considered that sp-gr. ALBIDA and sect. *Anaspis* were closely related.

According to this chapter, most of the phylogenetic relationships in *Scutellaria* are not clearly identified with the current system (Paton, 1990). It indicates that the key morphological features such as inflorescence and shape of bracts traditionally used to distinguish subgenera in the past do not provide the unique diagnostic character of clades. Besides, a re-examination of the relationship between sections and species groups is needed later. Therefore, further studies need to the fully-resolved tree with well-supported values.

2. Comparison between trees from this study and previous molecular phylogenetic studies

The previous molecular phylogenetic studies have been investigated the phylogenetic relationship of *Scutellaria,* but it was difficult to understand the infrageneric relationships because of regional sampling. This chapter shows new contributions to classifying *Scutellaria* through more extensive research.

The relationships of subgen. *Scutellaria* and subgen. *Apeltanthus* suggested in the phylogenetic tree of this chapter are the same as the previous molecular phylogenetic analyses (Zhao et al., 2017). The length of the dataset was 1,110 bp in Zhao et al. (2017), of which 332 characters were parsimony-informative characters. It included less length and informative character than this chapter, and they described the need to include more species and other regions. According to Zhao et al. (2017), subgen. *Apeltanthis* was a monophyletic group and subgen. *Scutellaria* was a paraphyletic group. These relationships were the same as the phylogenetic tree in this chapter. It means that the subgeneric classification system needs to be reexamined.

Recently, according to the study of Safikhani et al. (2018), the length of the aligned sequences included in the analysis was 1,464 bp in total, and the number of parsimony-informative characters was the lowest at 255. Safikhani et al. (2018) showed that the monophyly of subgen. *Apeltanthus* in nuclear ITS and chloroplast *trnL-F* trees. subgen. *Scutellaria* was indicated as monophyletic

group in nuclear ITS tree, but was paraphyletic group in *trnL−F* trees. However, the infrageneric relationships were unresolved due to the regional sampling, and it is difficult to compare with the current system (Paton, 1990). Conspicuously, according to Safikhani et al. (2018), *S. galericulata* was indicated as an independent clade in chloroplast *trnL−F* tree, and it was suggested that it could be presented as a new section. However, in this chapter, the members of the sp−gr. GALERICULATA (*S. barbata*, *S. dependens*, *S. minor,* and *S. galericulata*) did not form an independent clade and did not place at the base of the tree. Therefore, it is considered that sp−gr. GALERICULATA is not classified as a new section. The delimitation between the sect. *Scutellaria* including sp−gr. GALERICULATA and other sections is also considered still unresolved.

The previous molecular phylogenetic studies (Safikhani et al., 2018; Zhao et al., 2017) were not considered proper DNA regions and analyzed only two regions, and the dataset of sequence for phylogenetic analyses were insufficient. In addition, both studies confiremd that the section−, subsection−, and species group−level classification system established by Paton (1990) did not support the evolutionary relationship of *Scutellaria.* In this chapter, a molecular phylogenetic tree of *Scutellaria* was reconstructed using proper three chloroplast DNA genes for 200 representatives of *Scutellaria,* and parsimony−informative characters were the most included than previous studies. Although each section and species

group suggested by Paton (1990) were located in different clades, six major clades were recognized and the phylogenetic position of the groups in the genus was confirmed extensively. This chapter showed the first molecular phylogenetic relationship of *Scutellaria* in the world and provided the phylogenetic tree containing the most data and taxa. It will be the basis for a new classification system in *Scutellaria.*

CHAPTER 3.

Comparative analyses of chloroplast genomes in *Scutellaria* and phylogeny based on the super-matrix approach

I. Introduction

Many molecular phylogenetic analyses have been performed based on variable DNA sequence data, mainly chloroplast DNA regions (Bunsawat et al., 2004; Li et al., 2016; Steane et al., 1997; Wagstaff and Olmstead, 1997; Wagstaff et al., 1998). Sanger sequencing technique based on the PCR experiment has been traditionally used for phylogenetic studies. However, this method was costly and time-consuming to produce large amounts of data. Recently, many phylogenomic studies using the complete chloroplast genome sequences have been commonly conducted based on next-generation sequencing (NGS) technology (Bendiksby et al., 2011; Fragoso-Martinez et al., 2017; Zhao et al., 2020). NGS technique was developed to be possible to obtain large amounts of data with cost-effectively.

Chloroplast genome consists of circular double-stranded DNA molecules, mainly containing proteins related to photosynthesis and genes related plastid. Especially, chloroplast genome is mostly maternal inheritance in angiosperms and does not recombine when breeding. Therefore, individual variability is low and provides important information for species identification. Chloroplast genome is also known to have relatively well conserved in structures and contents of the gene, and the length of chloroplast genomes of angiosperm ranged from 120,000-160,000 bp. Generally, the structure of chloroplast genome is consist of a large single-copy (LSC) region, a small single-copy (SSC) region, and a pair of

inverted repeat (IRa and IRb) region.

The phylogenetic analysis using insufficient data reduces the resolution of the phylogenetic tree, which makes it difficult to understand the evolutionary relationship. It has been found that the resolution of phylogenetic trees can be increased depending on the number of taxa and the number of characteristics (Wortley et al., 2005; Philippe et al., 2011). To improve the phylogenetic tree with small data, the super-matrix phylogeny combining multiple gene sequences and whole-genome data has been performed (Diaz-Tapia et al., 2017; Roure et al., 2013; Williams et al., 2016; Uribe-Convers et al., 2017). Especially, the studies conducted by Williams et al. (2016) and Uribe-Convers et al. (2017) analyzed phylogenetic trees by integrating chloroplast DNA region sequences using Sanger sequencing and complete chloroplast genome sequences using high-throughput sequencing (HTS) technology. According to Williams et al. (2016), they were demonstrated that the phylogenetic tree combining chloroplast DNA region sequences and chloroplast genome sequences had higher resolution than the phylogenetic tree analyzed with only chloroplast DNA region sequences, although the missing data was included in a large proportion. In the study by Uribe-Convers et al. (2017), the phylogenetic tree obtained using only complete chloroplast genome data showed high resolution. However, the resolution of the phylogenetic tree more decreased as the taxa containing missing data was included. Therefore, it needs to confirm the resolution of

a tree by applying super-matrix phylogeny in *Scutellaria.*

Recently, Zhao et al. (2020) performed the phylogenomic study for 14 species of subfamily Scutellarioideae including 11 taxa of *Scutellaria.* Including these taxa, a total of 16 taxa in *Scutellaria* have been available at GenBank to date. Zhao et al. (2020) confirmed that subgenus (subgen.) *Apeltanthus* was a monophyletic group, and subgen. *Scutellaria* was a paraphyletic group. However, it was also difficult to compare with the current system because of the insufficient number of species.

In this chapter, chloroplast genomes of 49 taxa were completed including outgroup taxa using NGS technology. However, chloroplast genome data for 11 species published by Zhao et al. (2020) were not included in this chapter because these data were unavailable at the time this chapter was analyzed. The available chloroplast genomes of three species (*S. baicalensis*, MF521632; *S. insignis*, NC_028533; and *S. lateriflora*, NC_034693) were downloaded and included in the analyses. Based on the complete chloroplast genome data, 1) the structure and diversity between the complete chloroplast genomes were comprehensively investigated, 2) a super-matrix phylogeny was performed combining previously obtained three chloroplast DNA region sequences and complete chloroplast genome sequences to obtain a well-supported phylogenetic tree, 3) an obtained phylogenetic tree was compared with the current system (Paton, 1990) to understand the evolutionary relationship of *Scutellaria,* and 4) the phylogenetic

trees based on 52 complete chloroplast genomes, 10 regions representing high nucleotide diversity, and three genes (*ccsA, matK,* and *ndhF*) of Chapter 2 were compared.

II. Materials and Methods

1. Taxon sampling

A total of 49 taxa were sampled for chloroplast genome analyses including outgroup taxa. Comparing with the current system (Paton, 1990), all subgenera and sections (sect.) are included, and twenty-four out of 38 (63%) species group (sp-gr.) were included. Some samples were obtained from the field (Korea and China) and others are from major herbaria (KH, KYO, SWU, MO, NYBG, PE, and WTU). *Holmskioldia sanguinea* and *Tinnea aethiopica* were also included as an outgroup which is a sister taxon in the previous molecular study of Lamiaceae (Li et al., 2016). Taxa list and voucher information are shown in Appendix 3.

2. DNA extraction, amplification, and sequencing

Total genomic DNA (gDNA) was extracted using the GeneAll Plant SV Mini Kit (GeneAll Biotechnology Co. Ltd, Seoul, Korea) following the manufacturer's protocol. The obtained gDNA was stained using Quant-iT™ PicoGreen™ dsDNA Assay Kit

(Invitrogen, Oregon, USA) and analyzed for quality and quantity using SpectraMax M5 Microplate Reader (Molecular Device, California, USA). The low concentration samples (BGISEQ ≤ 12.5 ng/ul, Rapidgenomics ≤ 30 ng/ul) were was amplified using a whole-genome amplification (WGA) kit according to the manufacturer's instructions (Qiagen REPLI-g Mini Kit, Qiagen, Valencia, CA). The gDNA was run on 1.3% agarose gels to assess DNA quality. The gDNA of 31 taxa were conducted library construction by BGI (Shenzhen, China). The whole-genome re-sequencing was performed with 100 bp pair-end reads by the BGISEQ-500 platform (BGI, Shenzhen, China). The rest taxa were done library preparation and sequenced by RAPiD Genomics LLC (Florida, USA), and 100-bp paired-end reads were sequenced by Illumina Hiseq 3000 platform (Illumina Inc., San Diego, CA, USA).

3. *De-novo* assembly and annotation

Filtered reads were obtained from the BGI (Shenzhen, China), and the obtained raw reads from Rapid Genomics LLC were checked quality using FastQC v.0.11.5 (Andrews, 2010). The raw reads were removed adapters, barcodes, and low quality reads using Trimmomatic v.0.36 (Bolger et al., 2014). A trimming step was set to LEADING:20 TRAILING:20 SLIDINGWINDOW:5:20 MINLEN:36. To obtain chloroplast genome sequences, the paired-end reads of each taxon were mapped against a chloroplast

genome from *S. insignis* (GenBank accession: NC_028533) as a related taxon using Geneious (v9.0.5; Kearse et al., 2012) with the 'medium-low sensitivity option'. Subsequently, the quality of consensus sequences and their mapping condition were examined visually. Chloroplast genome sequences were annotated using GeSeq (Tillich et al., 2017). The annotated genome was compared with a chloroplast genome of *S. insignis* in the alignment generated by MAFFT v.7.308 (Katoh and Standley, 2013) module in the Geneious v.9.0.5 (Kearse et al., 2012).

4. Sequence analyses

The comparative analysis of 50 *Scutellaria* chloroplast genomes excluding outgroup taxa was performed using Mauve v.1.1.1 (Darling et al., 2004) plugin in Geneious v.9.0.6 (Kearse et al., 2012) to investigate the sequence rearrangement. The simple sequence repeats (SSRs) of *Scutellaria* chloroplast genomes were confirmed using the MISA website (Deier et al., 2018). SSR search parameters were set to ten or more for mono-nucleotide, five for di-nucleotide, four for tri-nucleotide, and three for tetra-nucleotide, penta-nucleotide, and hexa-nucleotide repeats following Zhao et al. (2020). The repetitive sequences including forward, reverse, complement, and palindromic repeats of *Scutellaria* chloroplast genomes were confirmed using the REPuter program online (Kurtz et al., 2001;

https://bibiserv.cebitec.uni-bielefeld.de/reputer).

Nucleotide diversity was calculated using DNASP v6.12.03 (Rozas et al., 2017), in this program, Pi value is defined as the average number of nucleotide differences per site. The degenerate codes in chloroplast genome sequences were converted to Ns using the option with'convert a fasta file with ambiguity codes to Ns'. The step size and sliding window were set to 200 bp and 600 bp. The number of variable site, parsimony informative site and p-distance was calculated using MEGA-X (Kumar et al., 2018).

5. Phylogenetic analyses based on super-matrix approach

The super-matrix phylogenetic analysis was performed by combining the data of Chapter 2 and this chapter. The obtained data which was the same taxa but from different vouchers in each chapter were included in the analysis as separate taxa. As a result, the combined dataset of 217 taxa including outgroup was used for phylogenetic analysis.

The missing data was filled with gaps (-), and GTR + I + Γ model was selected the best model using ModelFinder module (Kalyaanamoorthy et al., 2017) in IQ-tree v.1.6.3 (Nguyen et al., 2015). The ML analysis was conducted using Standard RAxML v.8.2.12 (Stamatakis, 2014) with 500 bootstrap replication. Bayesian inference (BI) analysis was performed using MrBayes v.3.2.7

(Ronquist et al., 2012) with the same options as Chapter 2. Each tree was visualized in MEGA7 (Kumar et al., 2016) and Figtree v.1.4.4 (Rambaut, 2018).

6. Calculation of quartet distance among phylogenetic trees

For ten regions representing high nucleotide diversity and the previous three genes (*ccsA*, *matK,* and *ndhF* in Chapter 2), each quartet distance was calculated to evaluate the phylogenetic information and confirm the conflict among branches in the phylogenetic trees. First, the phylogenetic analyses were conducted based on 52 taxa which chloroplast genome sequences were completed. The phylogenetic analyses were performed in the same method as super-matrix approach. The phylogenetic trees were obtained using complete chloroplast genome sequences, each of the high nucleotide diversity regions, the subsets of these ten regions, and the previous three genes. Subsequently, the quartet distance of each obtained phylogenetic tree was calculated using R package Quartet (Sand et al., 2014; Smith, 2018), and each quartet percentage was compared.

III. Results

1. Completion of chloroplast genome sequences

Sequencing using Illumina Hiseq produced the raw reads ranging from 2,563,825 to 6,536,592 reads, and raw reads using BGIseq were obtained ranging from 14,422,636 to 54,878,596 reads. A summary of the sequencing results of 49 taxa including outgroup taxa is shown in Appendix 4.

A total of 47 chloroplast genomes of *Scutellaria* were newly completed, and the general length of chloroplast genomes of *Scutellaria* ranged from 151,030 to 152,594 bp (Fig. 7, Table 8). Including the three previously published chloroplast genomes, the completed chloroplast genomes of *Scutellaria* divided the same as the general chloroplast genome structure of plants, and consist of a pair of IR regions (25,190–25,314 bp), LSC regions (83,008–84,622 bp), and SSC regions (17,292–17,600 bp). All chloroplast genomes of *Scutellaria* were encoded with 114 genes, 80 proteins, 30 tRNA genes, and 4 rRNA genes. The total GC content was 38.2% to 38.6% and indicated a similar percentage among taxa (Table 8). The newly completed chloroplast genomes were uploaded in GenBank, and the GenBank accession numbers are shown in Table 7. In addition, the functions and groups of all genes that consist of chloroplast genomes of *Scutellaria* are shown in Table 9. All genes that consist of chloroplast genomes of *Scutellaria* were identified in all taxa.

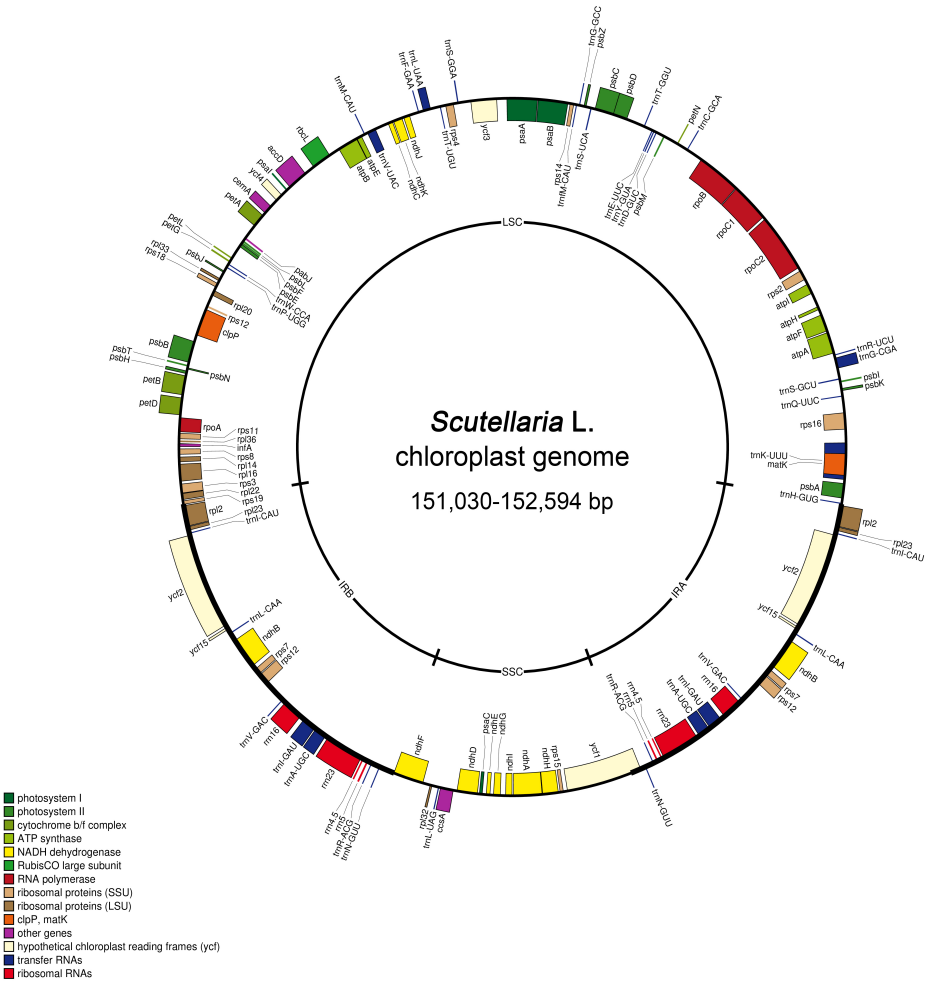


Fig. 7. Structure of chloroplast genomes in *Scutellaria.*

Table 8. Statistical summary for 52 chloroplast genomes including outgroup taxa

Taxa	Length (bp)				Gene	Protein coding	tRNA	rRNA	Total GC (%)	GenBank accession No.
	Total	LSC	SSC	IRs						
H. sanguinea (outgroup)	153,273	84,693	17,330	25,625	114	80	30	4	38.2	NC_047486
S. albida subsp. *condensata*	151,908	83,924	17,488	25,248	114	80	30	4	38.3	MW420937
S. altamaha	151,521	83,793	17,348	25,190	114	80	30	4	38.3	MW420957
S. angustifolia	152,539	84,524	17,527	25,244	114	80	30	4	38.3	MW420978
S. baicalensis	151,824	83,518	17,338	25,255	114	80	30	4	38.3	MF521632
S. bolanderi	152,310	84,438	17,406	25,233	114	80	30	4	38.3	MW420972
S. costaricana	151,823	83,946	17,415	25,231	114	80	30	4	38.4	MW420973
S. dependens	152,434	84,407	17,539	25,244	114	80	30	4	38.3	MW420952
S. discolor	151,762	83,971	17,309	25,241	114	80	30	4	38.4	MW420951
S. drummondii	152,266	84,383	17,493	25,195	114	80	30	4	38.3	MW420974
S. forrestii	151,750	83,908	17,338	25,252	114	80	30	4	38.3	MW420950
S. franchetiana	152,099	84,080	17,531	25,244	114	80	30	4	38.3	MW420968
S. gardoquioides	151,763	83,835	17,472	25,228	114	80	30	4	38.5	MW413436
S. gaumeri	152,339	84,366	17,463	25,255	114	80	30	4	38.4	MW420958
S. guatemalensis	152,330	84,371	17,463	25,248	114	80	30	4	38.4	MW420965
S. hastifolia	152,333	84,283	17,560	25,245	114	80	30	4	38.3	MW420934
S. heterophylla	152,365	84,419	17,538	25,204	114	80	30	4	38.3	MW420938
S. heydei	152,110	84,311	17,317	25,241	114	80	30	4	38.3	MW420941
S. hintoniana	151,030	83,008	17,550	25,236	114	80	30	4	38.4	MW420959
S. hissarica	152,042	83,962	17,568	25,256	114	80	30	4	38.3	MW420942
S. immaculata	152,008	84,280	17,292	25,218	114	80	30	4	38.3	MW420944
S. incana	151,634	83,770	17,454	25,205	114	80	30	4	38.3	MW420975
S. indica	151,915	83,922	17,525	25,234	114	80	30	4	38.4	MW420946
S. indica var. *coccinea*	151,956	83,951	17,539	25,233	114	80	30	4	38.3	MN047312

S. insignis	151,908	83,913	17,517	25,239	114	80	30	4	38.4	NC_028533
S. lateriflora	152,283	84,340	17,465	25,239	114	80	30	4	38.3	NC_034693
S. leptosiphon	152,146	84,346	17,320	25,240	114	80	30	4	38.3	MW420940
S. linearis	151,745	83,951	17,346	25,224	114	80	30	4	38.3	MW420945
S. longifolia	151,742	83,871	17,419	25,226	114	80	30	4	38.4	MW420966
S. mexicana	152,544	84,611	17,461	25,236	114	80	30	4	38.3	MW420976
S. mollis	152,558	84,622	17,468	25,234	114	80	30	4	38.3	MW420947
S. moniliorhiza	152,594	84,548	17,546	25,250	114	80	30	4	38.3	MW420954
S. multicaulis	152,234	84,407	17,343	25,242	114	80	30	4	38.3	MW420936
S. nana	152,529	84,502	17,541	25,243	114	80	30	4	38.3	MW420969
S. nummulariifolia	152,523	84,604	17,469	25,225	114	80	30	4	38.4	MW420960
S. orientalis	152,176	84,382	17,314	25,240	114	80	30	4	38.3	MW420933
S. orthocalyx	151,827	83,955	17,340	25,266	114	80	30	4	38.3	MW420955
S. ovata	151,637	83,777	17,454	25,203	114	80	30	4	38.3	MW420970
S. parvula	152,584	84,553	17,563	25,234	114	80	30	4	38.3	MW420964
S. pekinensis var. *maxima*	152,530	84,502	17,568	25,230	114	80	30	4	38.3	MW420953
S. pinnatifida subsp. *mucida*	151,813	84,011	17,324	25,239	114	80	30	4	38.3	MW420948
S. pinnatifida subsp. *pichleri*	151,551	83,796	17,305	25,225	114	80	30	4	38.3	MW420935
S. platensis	152,484	84,555	17,475	25,227	114	80	30	4	38.3	MW420949
S. rubicunda	152,016	83,958	17,568	25,245	114	80	30	4	38.3	MW420939
S. scutellarioides	151,544	83,494	17,422	25,314	114	80	30	4	38.6	MW420961
S. seleriana	151,744	83,898	17,384	25,231	114	80	30	4	38.4	MW420962
S. siphocampyloides	152,533	84,451	17,600	25,241	114	80	30	4	38.3	MW420977
S. splendens	151,861	83,977	17,430	25,227	114	80	30	4	38.4	MW420963
S. strigillosa	152,578	84,557	17,545	25,238	114	80	30	4	38.3	MW420956
S. suffrutescens	152,184	84,275	17,495	25,207	114	80	30	4	38.3	MW420967
S. tomentosa	151,777	83,985	17,326	25,233	114	80	30	4	38.3	MW420943
T. aethiopica (outgroup)	153,186	84,653	17,914	25,284	114	80	30	4	38.4	MW420971

Table 9. The functions and gene groups of genes included in the chloroplast genomes in *Scutellaria*

Genes category	Gene group	Gene name
Self replication	Ribosomal RNA genes	*rrn5*(2x), *rrn4.5* (2x), *rrn16* (2x), *rrn23* (2x)
	Transfer RNA genes	[a]*trnA-UGC*(2x), *trnC-GCA*, *trnD-GUC*, *trnE-UUC*, *trnF-GAA*, *trnfM-CAU*, [a]*trnG-GCC*, *trnG-UCC*, *trnH-GUG*, [a]*trnI-CAU* (2x), [a]*trnI-GAU* (2x), [a]*trnK-UUU*, *trnL-CAA* (2x), [a]*trnL-UAA*, *trnL-UAG*, *trnM-CAU*, *trnN-GUU* (2x), *trnP-UGG*, *trnQ-UUG*, *trnR-ACG* (2x), *trnR-UCU*, *trnS-GCU*, *trnS-GGA*, *trnS-UGA*, *trnT-GGU*, *trnT-UGU*, *trnV-GAC* (2x), [a]*trnV-UAC*, *trnW-CCA*, *trnY-GUA*
	Small subunits of ribosomal protein genes	*rps2*, *rps3*, *rps4*, *rps7* (2x), *rps8*, *rps11*, *rps12*, *rps14*, *rps15*, [a]*rps16*, *rps18*, *rps19*
	Large subunit of ribosomal protein genes	[a]*rpl2* (2x), *rpl14*, *rpl16*, *rpl20*, *rpl23* (2x), *rpl32*, *rpl33*, *rpl36*
	RNA polymerase subunits	*rpoA*, *rpoB*, [a]*rpoC1*, *rpoC2*
Genes for photosynthesis	Subunits of photosystem Ⅰ	*psaA*, *psaB*, *psaC*, *psaI*, *psaJ*,[b]*ycf3*
	Subunits of photosystem Ⅱ	*psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbH*, *psbI*, *psbJ*, *psbK*, *psbL*, *psbM*, *psbN*, *psbT*, *psbZ*
	Subunits of Cytochrome	*petA*, [a]*petB*, *petD*, *petG*, *petL*, *petN*
	Subunits of ATP-synthases	*atpA*, *atpB*, *atpE*, [a]*atpF*, *atpH*, *atpI*
	RuBisCo large subunit	*rbcL*
	Subunits of NADH dehydrogenase	[a]*ndhA*, [a]*ndhB* (2x), *ndhC*, *ndhD*, *ndhE*, *ndhF*, *ndhG*, *ndhH*, *ndhI*, *ndhJ*, *ndhK*
Other genes	Translational initiation factor	*infA*
	Maturase	*matK*
	Protease	[b]*clpP*
	Envelope membrane protein	*cemA*
	Subunit of Acetyl-CoA-carboxylase	*accD*
	C-type cytochrome synthesis gene	*ccsA*
	Component of TIC complex	*ycf1*
Unknown function	Conserved open reading frames	*ycf2* (2x), *ycf4*, *ycf15* (2x)

[a]genes contain one intron; [b]genes contain two introns; (2x) duplicate genes in IR.

2. Structural analyses of chloroplast genomes

The multiple alignment results using Mauve v1.1.1 showed that chloroplast genomes of *Scutellaria* are highly conserved (Fig. 8). The total chloroplast genomes were compared with *S. insignis* as reference taxon, the rearrangement such as translocation and inversions was not detected.

The expansion and contraction of IR regions (Fig. 9) were also investigated, chloroplast genome boundaries of all taxa were highly conserved. In all taxa, it was confirmed that five genes were located in common at the IR boundaries. The *rps19* gene located in LSC/IRa boundary and 41−57 bp was expanded to the IRa region. The IRa/SSC boundary was located in *ndhF* gene with 16−45 bp protrusion, and the SSC/IRb boundary was located *ycf1* gene with 729−784 bp protrusion. The *rpl2* gene (1,482−1,489 bp) was located on the left side of the IRb/LSC boundary, and the *trnH* gene (74 bp) was located on the right side. The distance between *rpl2* and the IRb/LSC boundary is 95−111 bp, and the *trnH* gene was 1−9 bp.

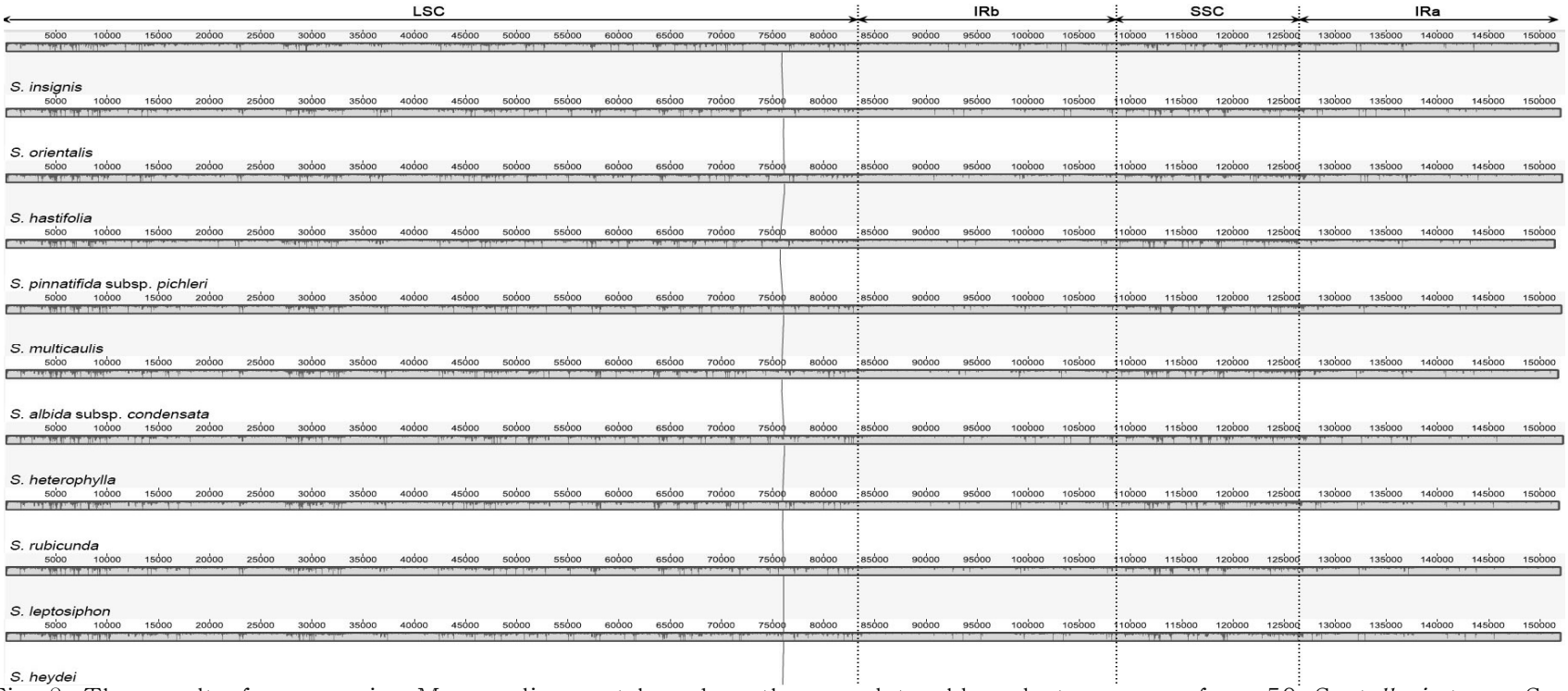


Fig. 8. The result of progressive Mauve alignment based on the complete chloroplast genomes from 50 *Scutellaria* taxa. *S. insignis* is a reference genome (top) for the analyses. Local collinear blocks (single block in this study) are connected by a vertical line. Dotted lines indicate borders of IR regions.
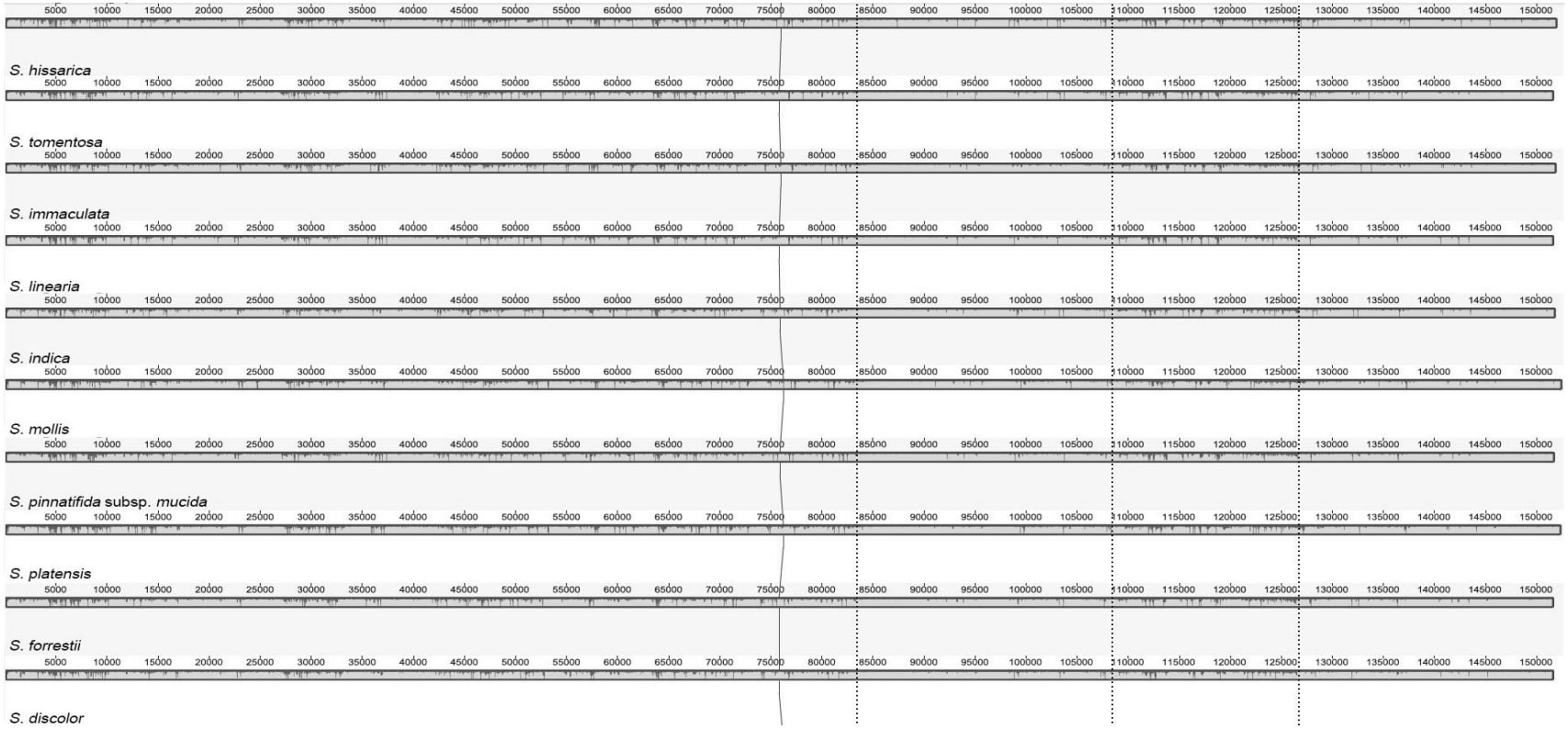


Fig. 8. Continued.
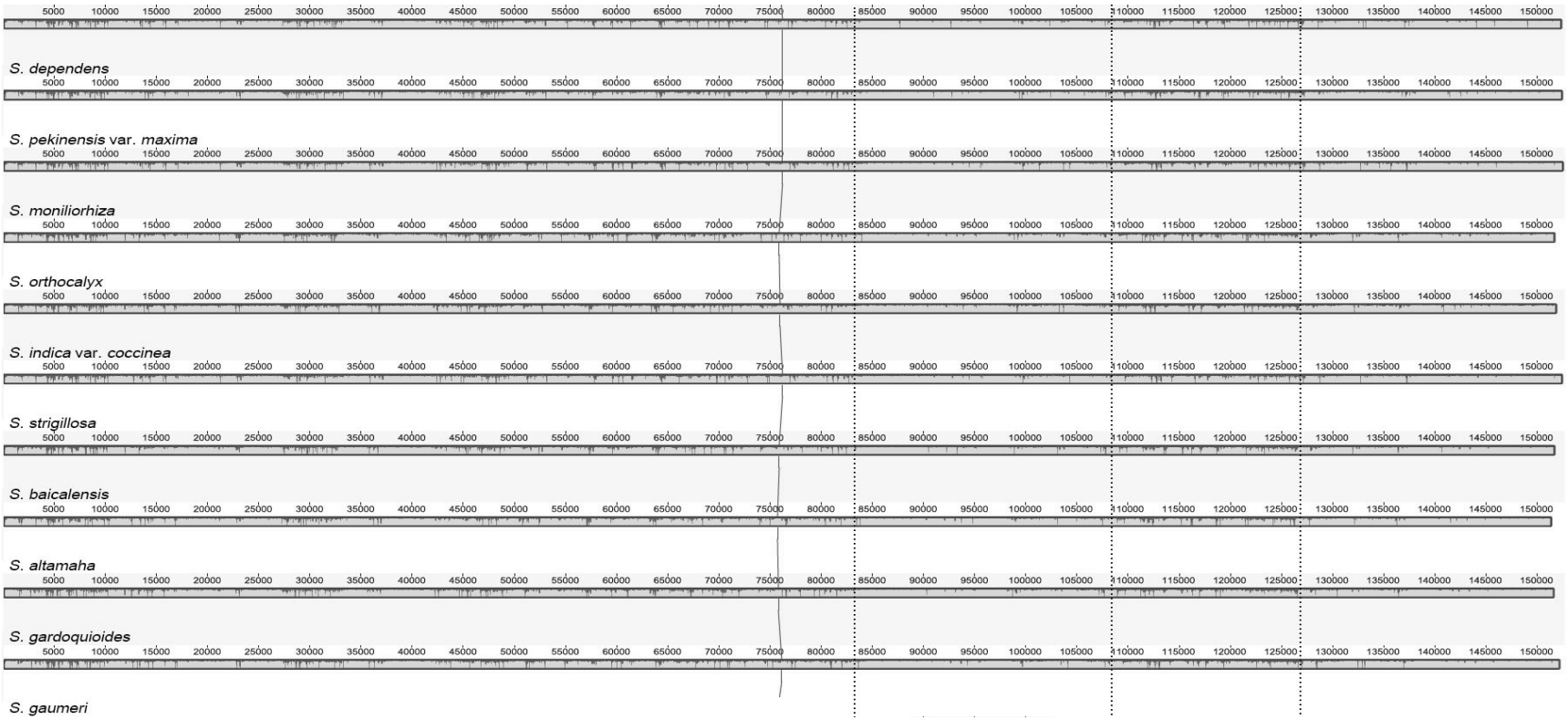


Fig. 8. Continued.



Fig. 8. Continued.



Fig. 8. Continued.

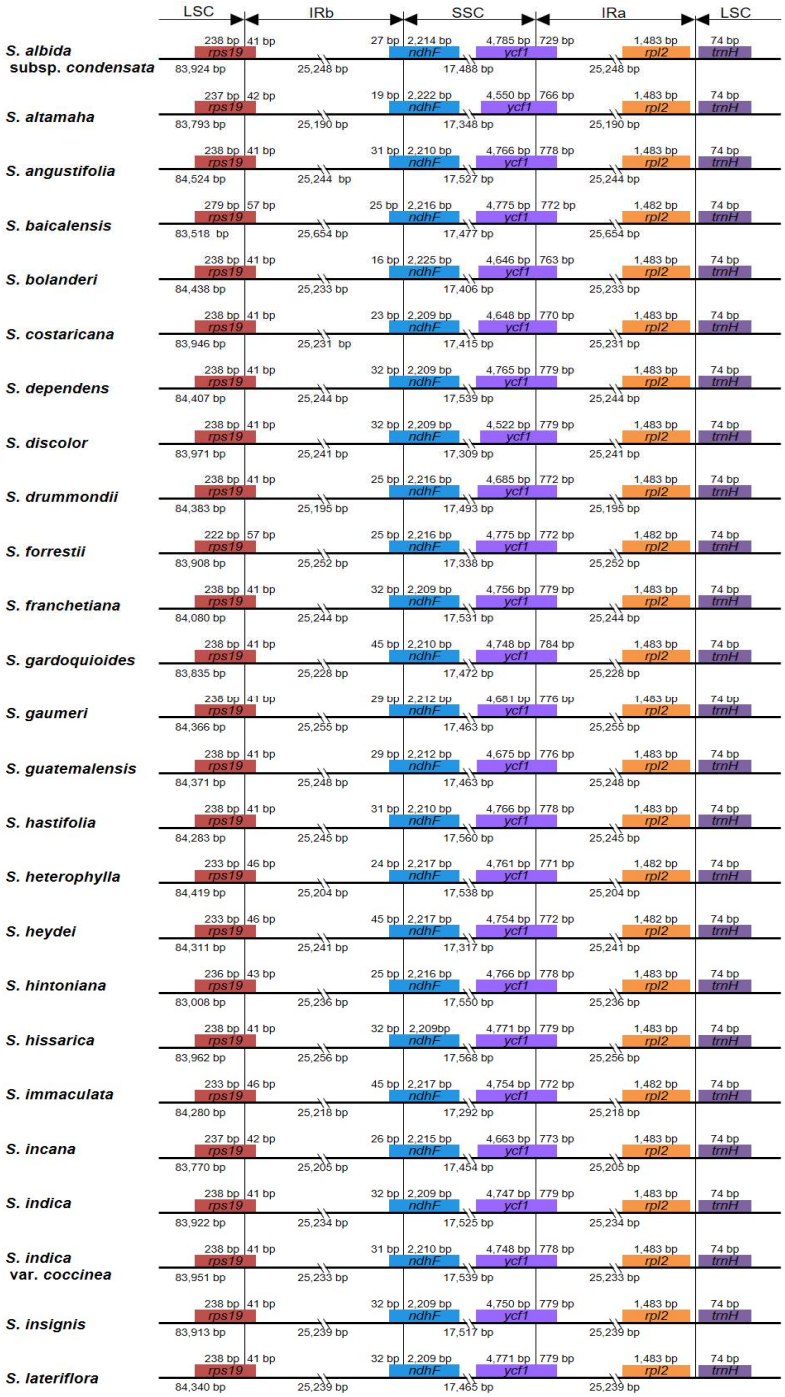


Fig. 9. Comparison of the IR borders in chloroplast genomes of 50 *Scutellaria* taxa.

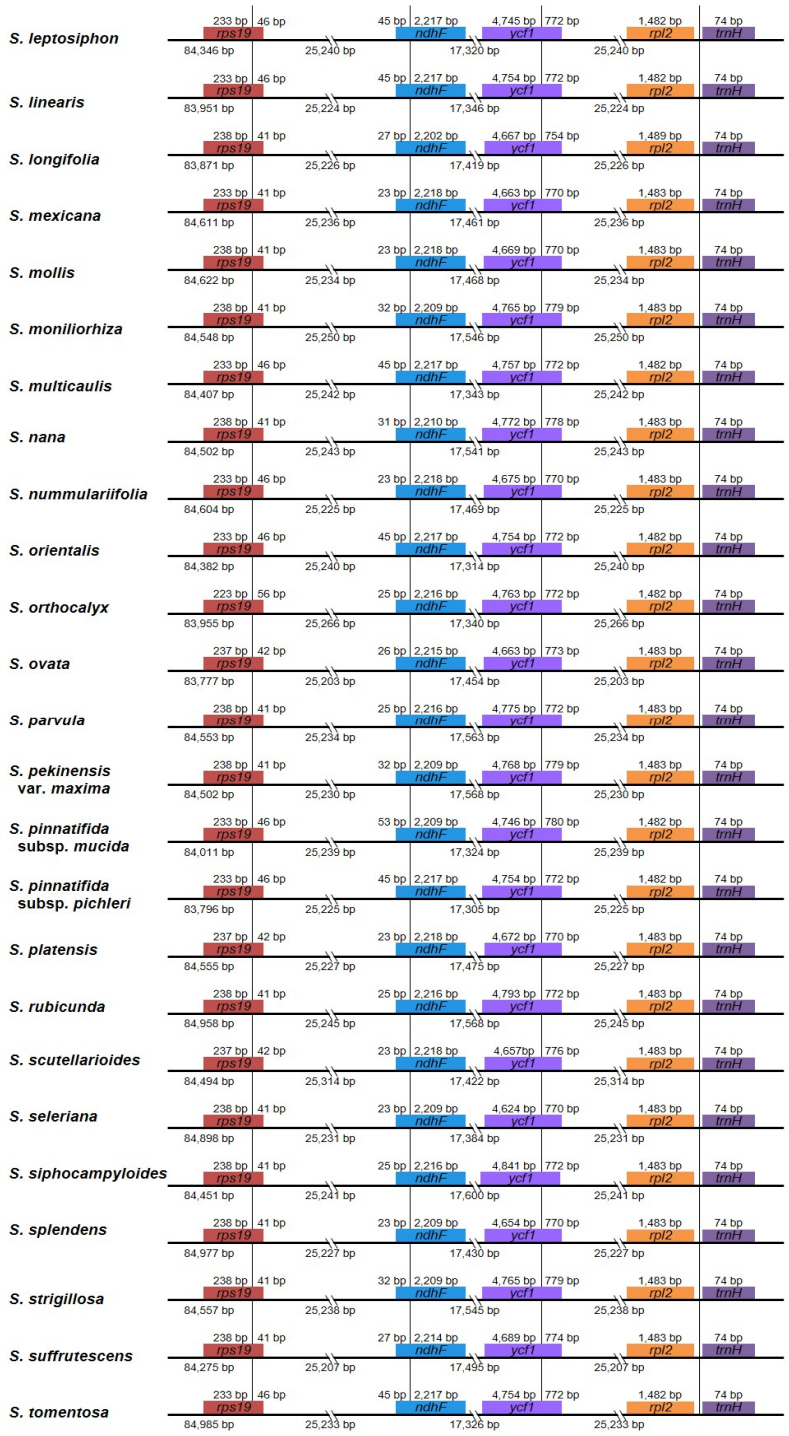


Fig. 9. Continued.

3. Repeat sequences

The total number of SSRs for each taxon in *Scutellaria* chloroplast genomes ranged from 32 to 56 (Fig. 10, Appendix 5). Mono, di, tri, and tetra nucleotides were commonly detected in all taxa, and pentanucleotides were found in only nine taxa (*S. costaricana*, *S. gaumeri*, *S. guatemalensis*, *S. longifolia, S. mollis*, *S. moniliorhiza, S. multicaulis*, *S. splendens,* and *S. strigillosa*). Hexanucleotides were found in only four taxa (*S. baicalensis, S. bolanderi*, *S. rubicunda*, and *S. siphocampyloides*).

A total of 1,730 SSRs were identified in the *Scutellaria* chloroplast genomes, 20 types of SSR repeat units were confirmed (Fig. 11, Appendix 6). The five types of repeat units were shared in all taxa: mononucleotide repeat unit (A/T), dinucleotide repeat unit (AT/AT), trinucleotide repeat unit (AAT/ATT), tetranucleotide (AAAG/CTTT and AAAT/ATTT). Mononucleotide repeat unit (C/G) was not detected in *S. seleriana,* and trinucleotide repeat unit (AAG/CTT) was absent in *S. orthocalyx*. Additionally, tetranucleotide repeat unit (ACAG/CTGT) was not present in *S. pekinensis* var. *maxima* and *S. splendens*. The rest of the tetranucleotide repeat units (AACC/GGTT, AGAT/ATCT, AATC/ATTG, and AATT/AATT) showed the most polymorphism. Pentanucleotide repeat units (AAAAG/CTTTT, AAGAT/ATCTT, and AAAGG/CCTTT) was each detected in five taxa (*S. costaricana*, *S. gaumeri, S. guatemalensis, S. longifolia,* and *S. splendens*),

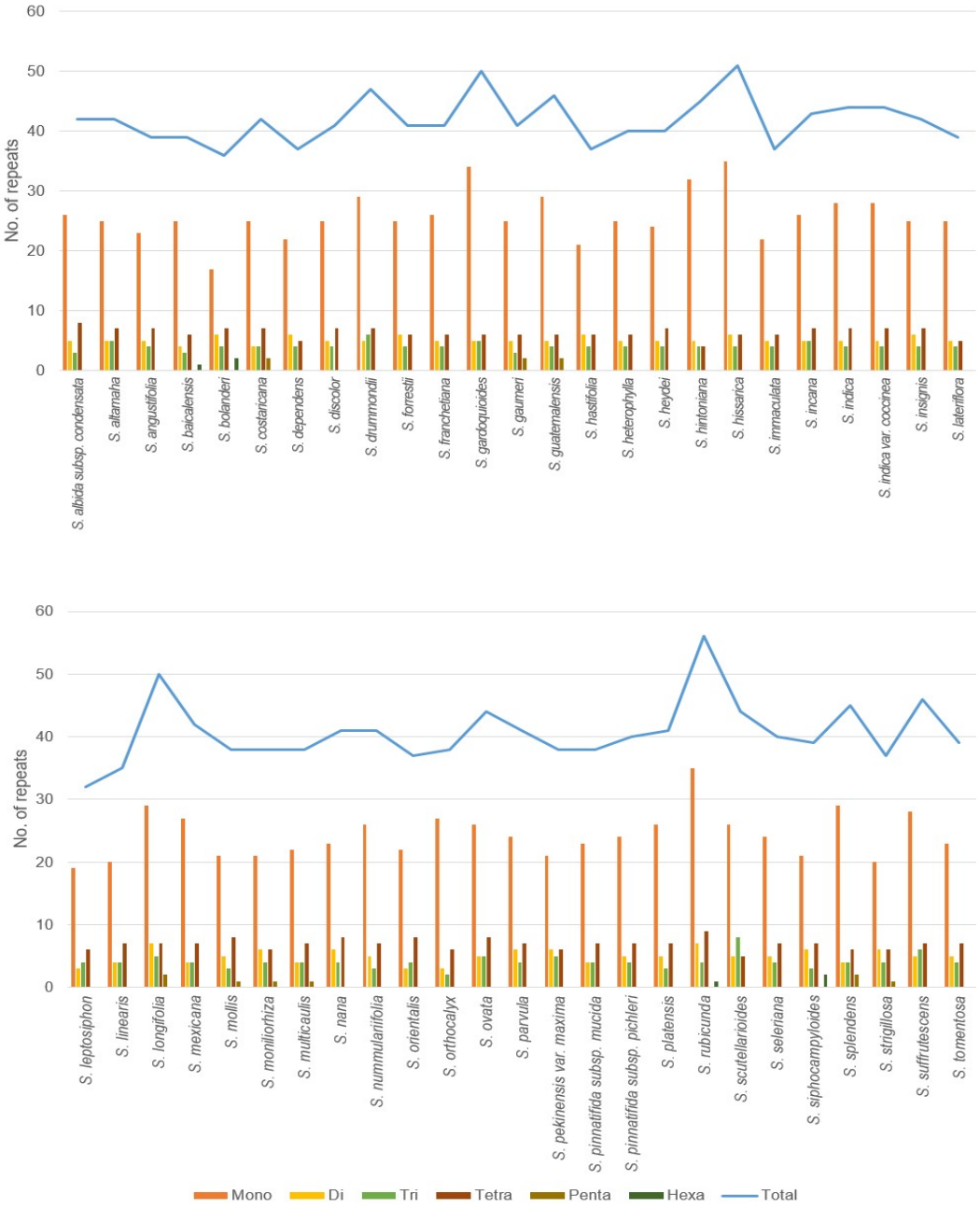


Fig. 10. The types and number of detected SSRs in 50 chloroplast genomes in *Scutellaria*. Mono, di, tri, tetra, penta, hexa indicate the types of SSR, and total means the sum of the number in each type.

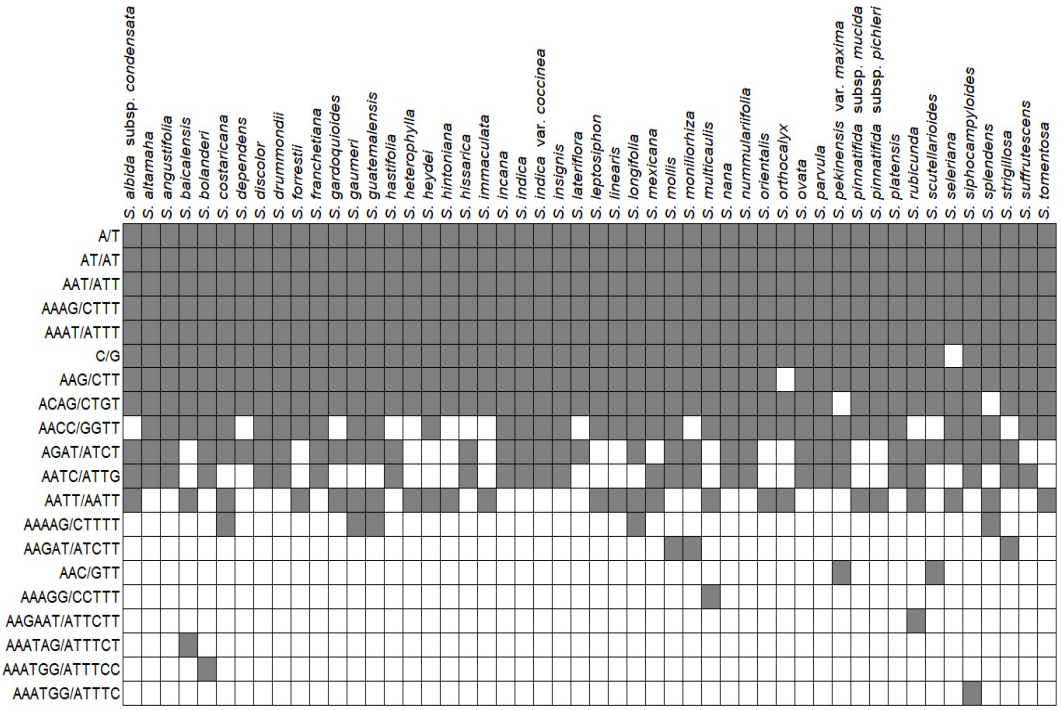


Fig. 11. Distribution of 20 types of SSR repeat units detected in 50 chloroplast genomes in *Scutellaria.* The x-axis represents each species and the y-axis represents types of SSR repeat unit. Filled squares indicate present, and empty squares indicate absent.

three taxa (*S. mollis, S. moniliorhiza,* and *S. strigillosa*), and one taxa (*S. multicaulis*). Trinucleotide repeat unit (AAC/GTT) was presented in only two taxa (*S. pekinensis* var. *maxima* and *S. scutellarioides*). Each of the four hexanucleotide units (AAGAAT/ATTCTT, AAATAG/ATTTCT, AAATGG/ATTTCC, and AAATGG/ATTTC) was detected in only four taxa *(S. baicalensis, S. bolanderi, S. rubicunda,* and *S. siphocampyloides*). Among a total of SSRs in *Scutellaria* chloroplast genomes, the number of A/T bases accounted for the largest percentage with 954 (55.1 %; Appendix 7), and G/C bases were rarely found with a total of 136 (7.9 %). Also, most of the A/T repeats were detected in non-coding regions.

For repeat sequences of length 30 bp or more, forward, palindromic, reverse, and complement repeats were variously detected in chloroplast genomes of *Scutellaria* (Fig. 12), and the total number of repeat ranged from 17 to 28. The forward repeat was the most common, followed by a palindromic repeat. The reverse repeat was confirmed in eight taxa (*S. albida* subsp. *condensata*, *S. pinnatifida* subsp. *pichleri*, *S. tomentosa*, *S. indic*a, *S. pinnatifida* subsp. *mucida*, *S. dependens*, *S. orthocalyx*, and *S. hintoniana*), and complement repeat was confirmed only in *S. heterophylla*. The length of the repeat ranged from 30 bp to 215 bp (Fig. 13), and the repeat of 30 bp was the most common. Information on SSR and repeat sequences for each taxon is shown in Appendix 5.

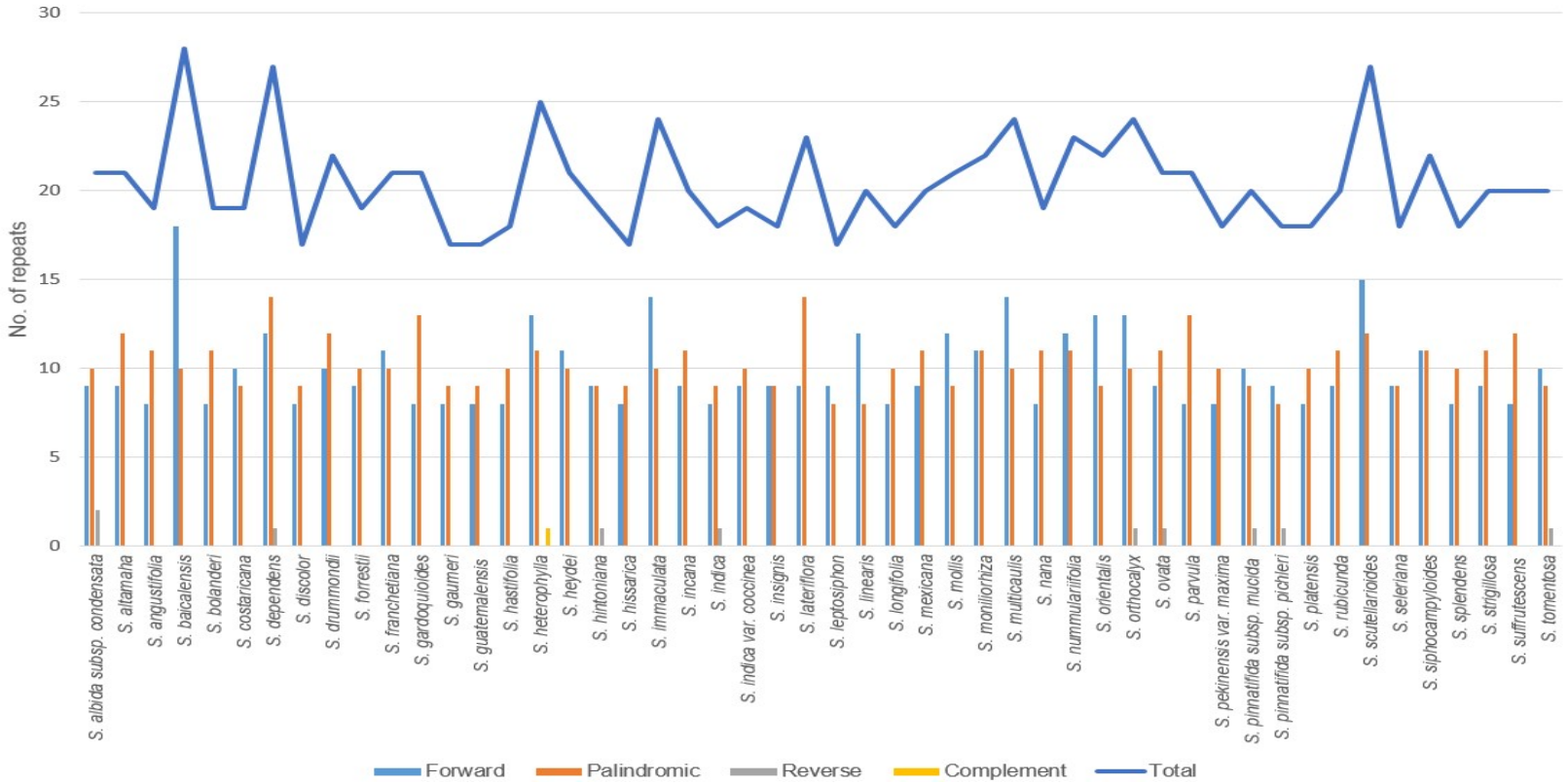


Fig. 12. Frequency of forward, palindromic, reverse, and complement repeat in each taxon.

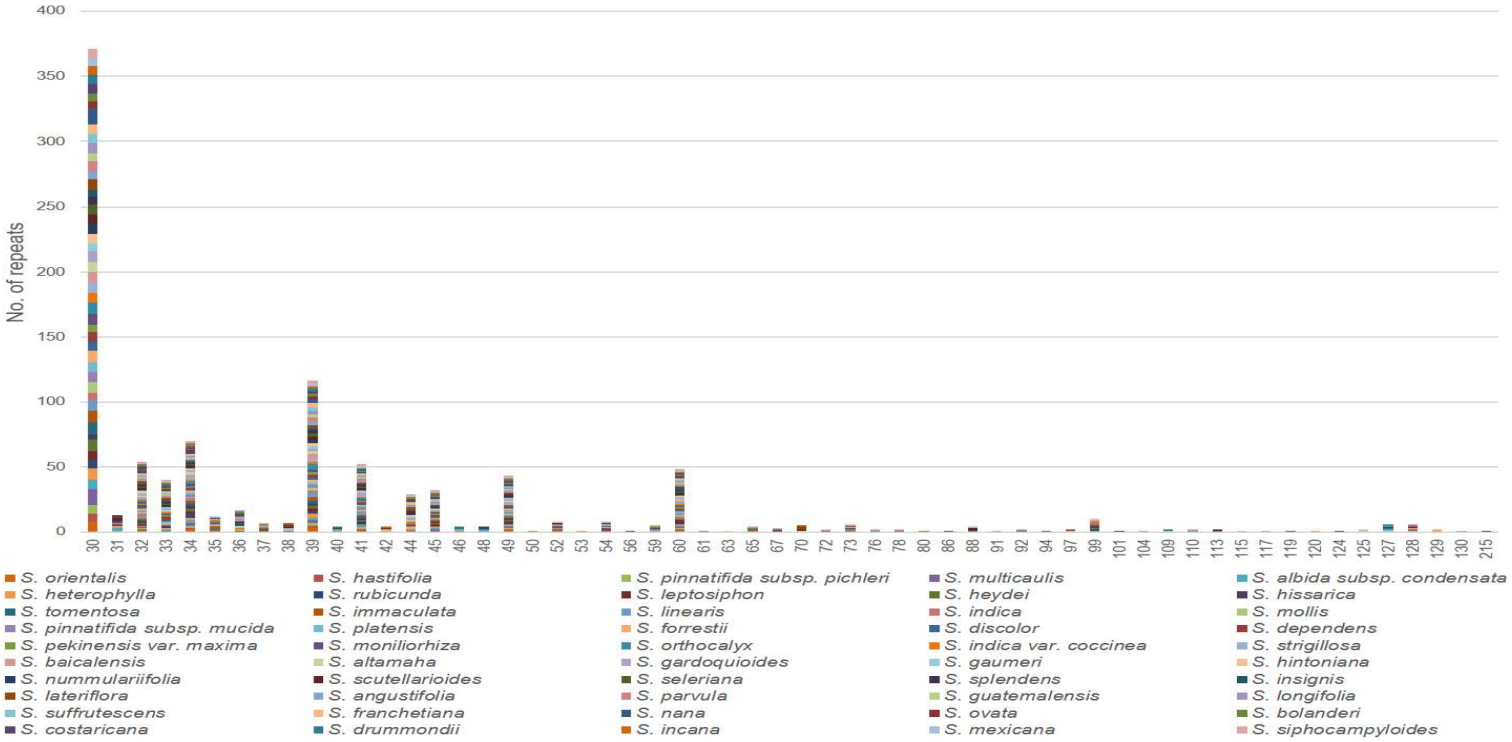


Fig. 13. Frequency of repeat length in each chloroplast genome. Numbers in X-axis represent the number of bases per repeating unit.

4. Nucleotide diversity in chloroplast genomes

The nucleotide diversity (Pi) values were analyzed including LSC, SSC, and IR regions (Fig. 14, Appendix 8) of chloroplast genomes of *Scutellaria,* and ten regions with high Pi values were investigated (Table 10). The average of Pi was 0.008 in the complete chloroplast genomes and ranged from 0 to 0.044. In the SSC region, the Pi value was the highest at 0.017, followed by the LSC region (Pi = 0.01) and the IR region (Pi = 0.002). The highest Pi value region was *trnK-rps16* (Pi = 0.04405) intergenic spacer region. The second highest value of Pi was *ycf1* gene, and the third was *ndhF* gene (Pi=0.04405). The rest seven regions were intergenic spacer regions, *petA-psbJ* (Pi = 0.02867), *petN-psbM* (Pi = 0.02861), *ccsA-ndhD* (Pi = 0.02472), *trnT*-UGU-*trnL*-UAA (Pi = 0.02408), *rpoB-trnC* (Pi = 0.02379), *rpl32-ccsA* (Pi = 0.02333) and *rbcL-accD* (Pi = 0.02288) indicated high Pi values sequentially. As a result of investigating the ratio of variable sites and informative sites for 10 regions, *trnK-rps16* region with the highest Pi value had the second-highest variable sites, and the highest variable region was *trnT-trnL* region. In addition, the region with the highest informative site was the *ycf1* region, and the second most informative site was the *trnK-rps16* region with the highest Pi value.

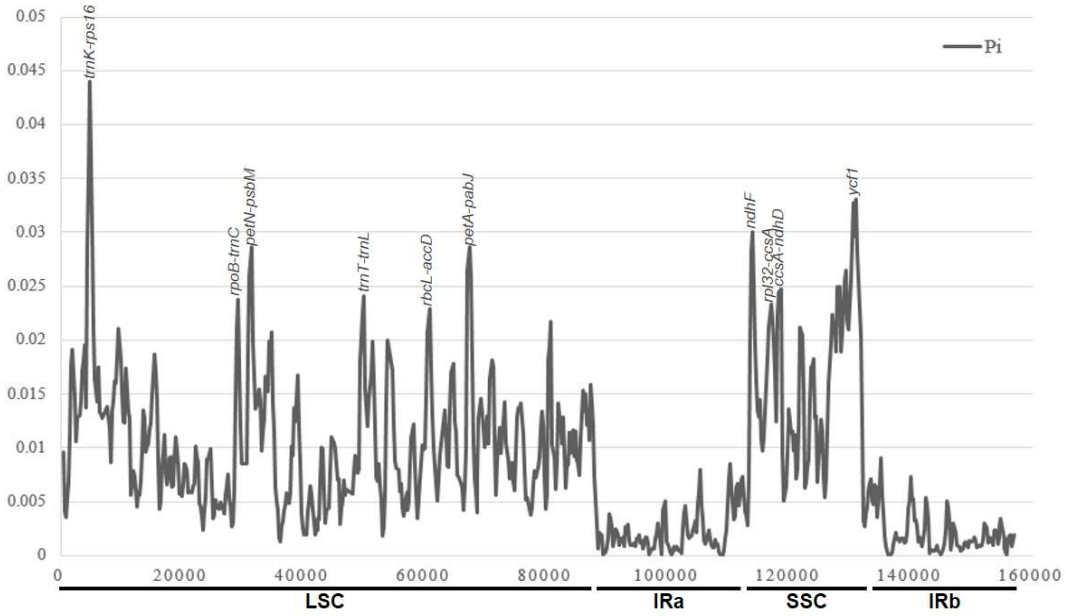


Fig. 14. Nucleotide diversity of *Scutellaria* chloroplast genomes based on sliding window analysis (window length = 600 bp, step size = 200 bp).

Table 10. Top ten regions with high nucleotide diversity and their sequence variability in *Scutellaria* chloroplast genomes

Position in the matrix (range)	Midpoint	Pi	Theta	Region Name	Length	No. of Variable sites [bp]	No. of Informative sites [bp]
4419-5280	4844	0.04405	0.06177	*trnK-rps16*	862	212 (24.6%)	110 (12.8%)
130538-131308	130999	0.03308	0.05135	*ycf1*	771	179 (23.2%)	118 (15.3%)
113559-114182	113882	0.03004	0.03981	*ndhF*	624	113 (18.1%)	78 (12.5%)
66834-67770	67316	0.02867	0.03721	*petA-psbJ*	937	162 (17.3%)	106 (11.3%)
30951-32011	31388	0.02861	0.03684	*petN-psbM*	1061	125 (11.8%)	72 (6.8%)
118198-118870	118570	0.02472	0.03907	*ccsA-ndhD*	673	116 (17.2%)	77 (11.4%)
49481-50278	49950	0.02408	0.05023	*trnT-trnL*	798	204 (25.6%)	73 (9.1%)
28838-29552	29195	0.02379	0.03349	*rpoB-trnC*	715	100 (14%)	69 (9.7%)
116516-117388	117024	0.02333	0.03498	*rpl32-ccsA*	873	159 (18.2%)	104 (11.9%)
60281-61056	60682	0.02288	0.03014	*rbcL-accD*	776	98 (12.6%)	69 (8.9%)

5. Comparison between super-matrix tree and three chloroplast genes tree

The total length of aligned sequences using phylogenetic analysis was 168,419 bp. The phylogenetic tree based on ML analysis of the 213 taxa including the outgroup taxa is shown in Fig. 15, and this phylogenetic tree is summarized based on taxonomic groups recognized in the current classification system (Paton, 1990; Fig. 16). The tree based on BI analysis is shown in Appendix 9. The topology of the super-matrix tree that combining three chloroplast genes and the complete chloroplast genome sequences was similar to that of the analysis with only three chloroplast DNA region data (previous tree), and the positions of major clades (clades #1-6) were the same as in the previous tree (Fig. 5). In the super-matrix tree, the resolution of some nodes that showed the consensus of bootstrap support (BS) value below 50 % was higher than the previous tree (Fig. 5). Among clades #1-6, clades #4 and #6 were indicated slightly lower BS value, and the rest of the clades were highly supported. Clades #1 and #3 showed 100 % BS value, the same as the previous tree, and were strongly supported. Clade #2 was indicated 88% BS value, and 3 % higher than the previous tree. Clade #4 showed a value of less than 50 % in the previous tree, but 55 % BS value in the super-matrix tree. In addition, clade #5 was 1 % higher than the previous tree and was strongly supported with 100 % BS value. However, clade #6 showed a lower BS value (53 %) than the previous tree (76 %).

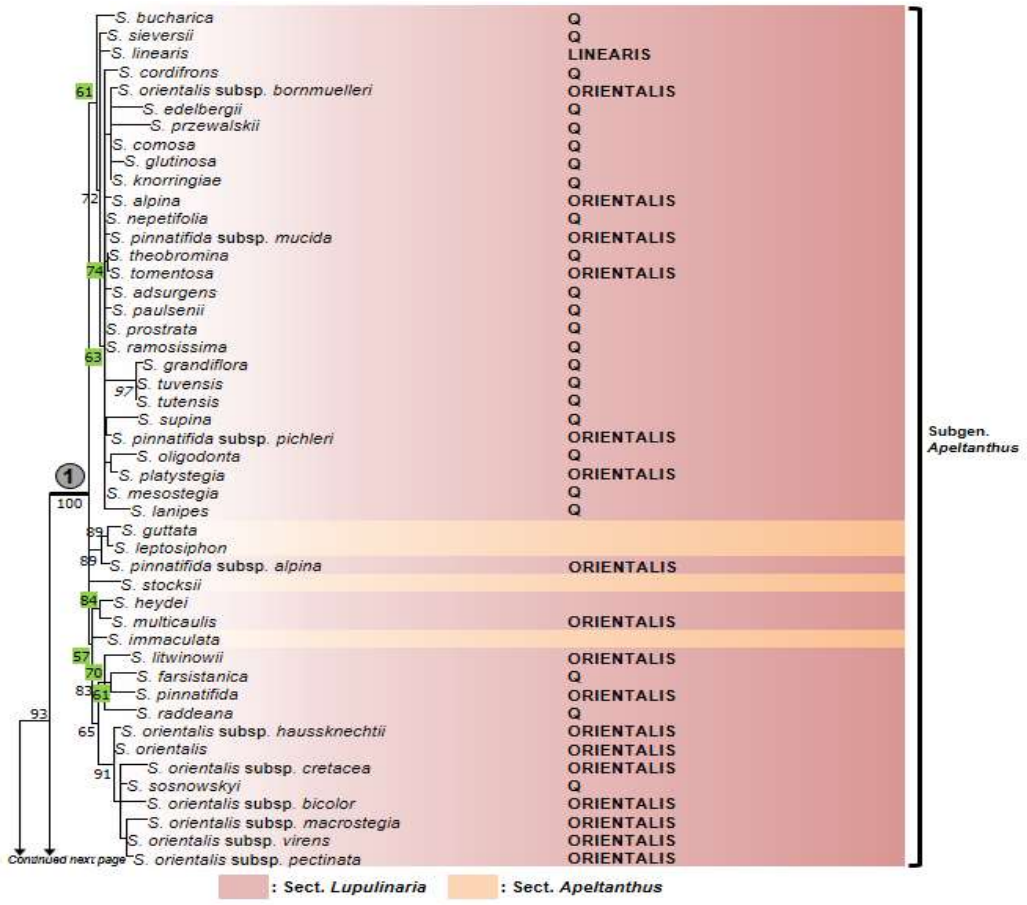


Fig. 15. A super-matrix maximum-likelihood tree based on three chloroplast genes (213 taxa) and chloroplast whole genome data. Numbers above the branches are bootstrap values, and numbers only BS ≥ 50% are indicated. The colored values in green indicate improved values compare to the three chloroplast genes tree. Colored boxes are sections and capitalized names after the taxon names are "species groups" recognized in Paton (1990). Circled numbers indicate six major clades recognized in this tree and these are the same as in the three chloroplast genes tree except the position of *S. heterophylla*. "Q" indicates a taxon of which taxonomic position (sections and "species group") have not been reported.

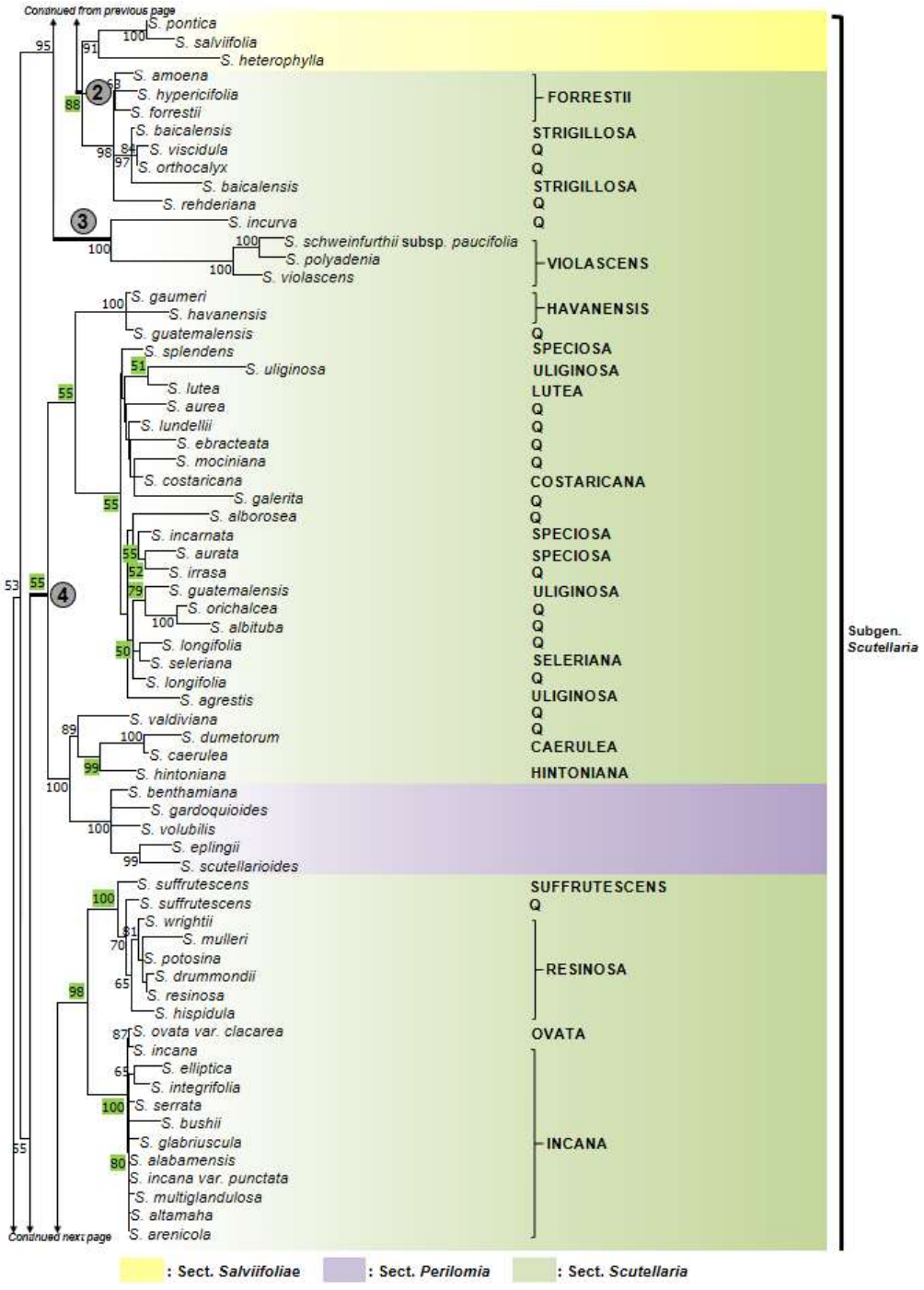


Fig. 15. Continued.

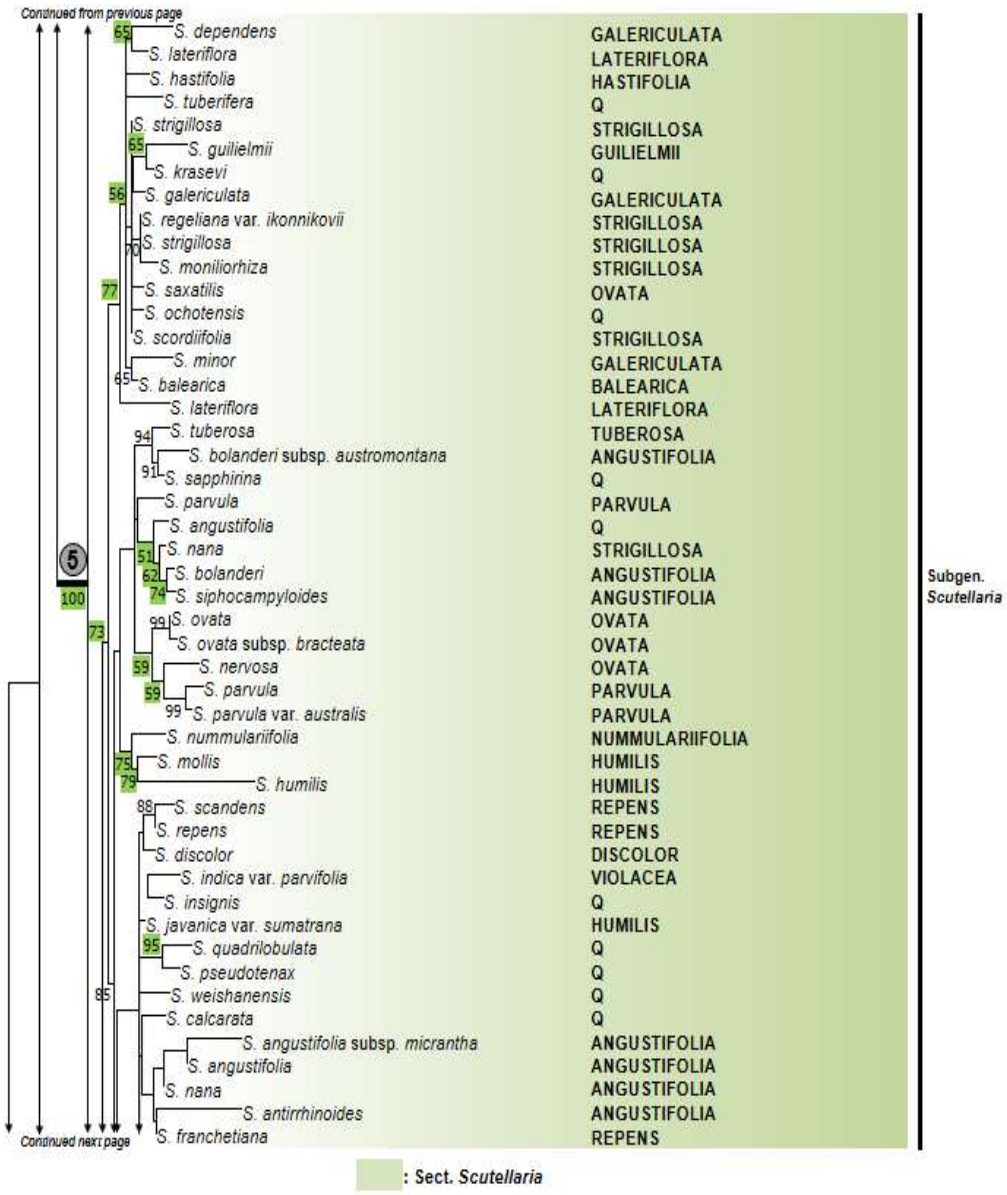


Fig. 15. Continued.

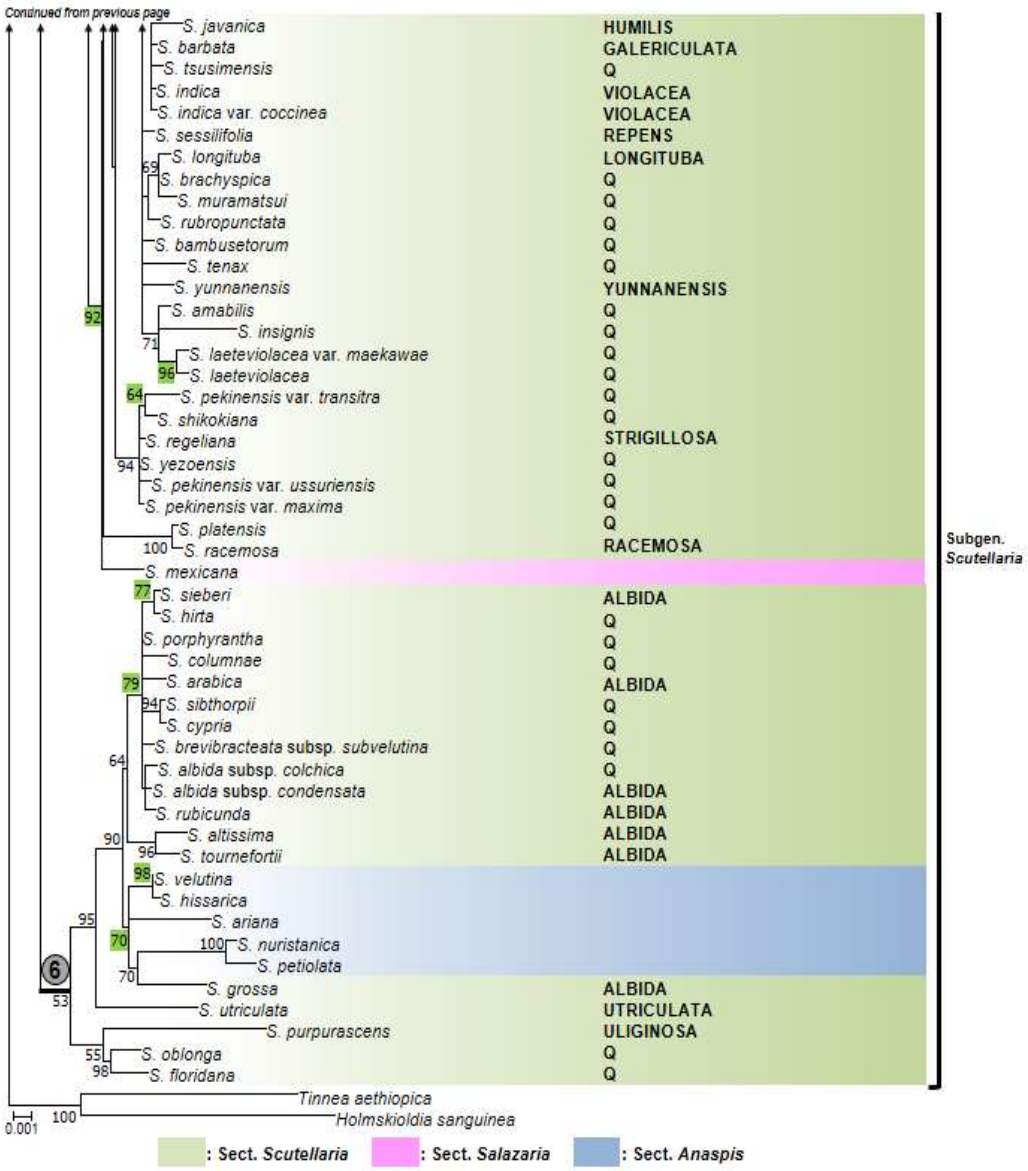


Fig. 15. Continued.

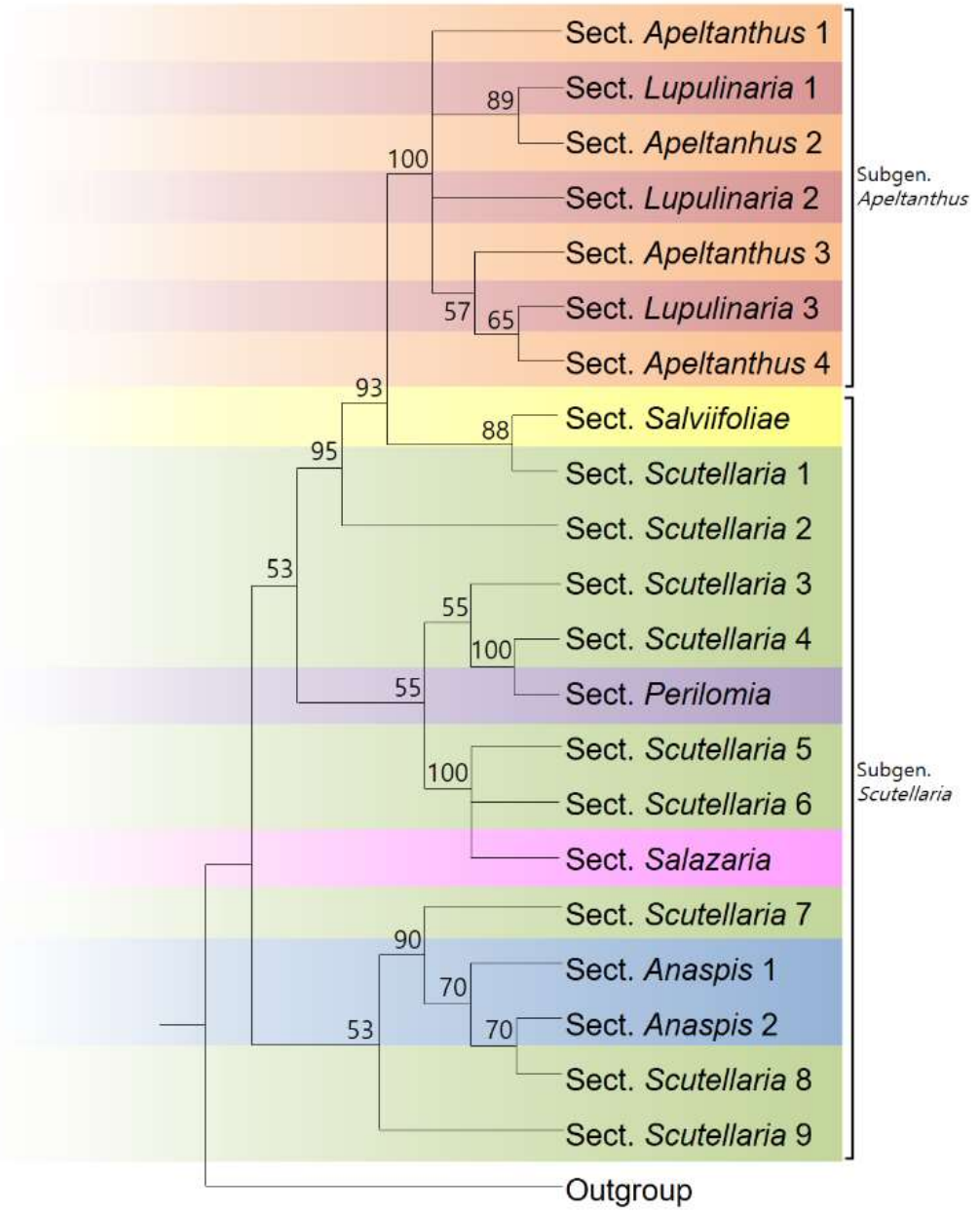


Fig. 16. A summarized super-matrix ML tree based on taxonomic groups recognized in the current classification system (Paton, 1990). Only nodes over 70% of bootstrap value are recognized.

6. Quartet distance among phylogenetic trees

To compare the topologies and resolutions among the phylogenetic trees based on 10 regions with high nucleotide diversity and the previous phylogenetic tree based on three genes, a total of 20 phylogenetic trees was obtained (Appendix 10). Each phylogenetic tree was calculated quartet distance comparing with the phylogenetic tree of the complete chloroplast genome (Fig. 17). The percentage of quartet distances for phylogenetic trees ranged from 47.62 to 77.46 %. All percentages of quartet distance analyzed by each phylogenetic tree using each region were lower (47.62–74.02 %) than the combined subset percentage. Among the phylogenetic trees of subsets, the subset that combined the two regions *trnK–rps16* and *ycf1* showed the lowest percentage (74.64 %). The subsets with the highest percentage (77.46 %) were two phylogenetic trees using a combined nine regions and a combined 10 regions. The percentage of quartet distance for a phylogenetic tree based on the previous three genes (*ccsA*, *matK*, and *ndhF*) was 76.19 %, showing higher than the subsets of the two regions (*trnK–rps16* and *ycf1*) and three regions (*trnK–rps16*, *ycf1*, and *ndhF*) with high nucleotide diversity.

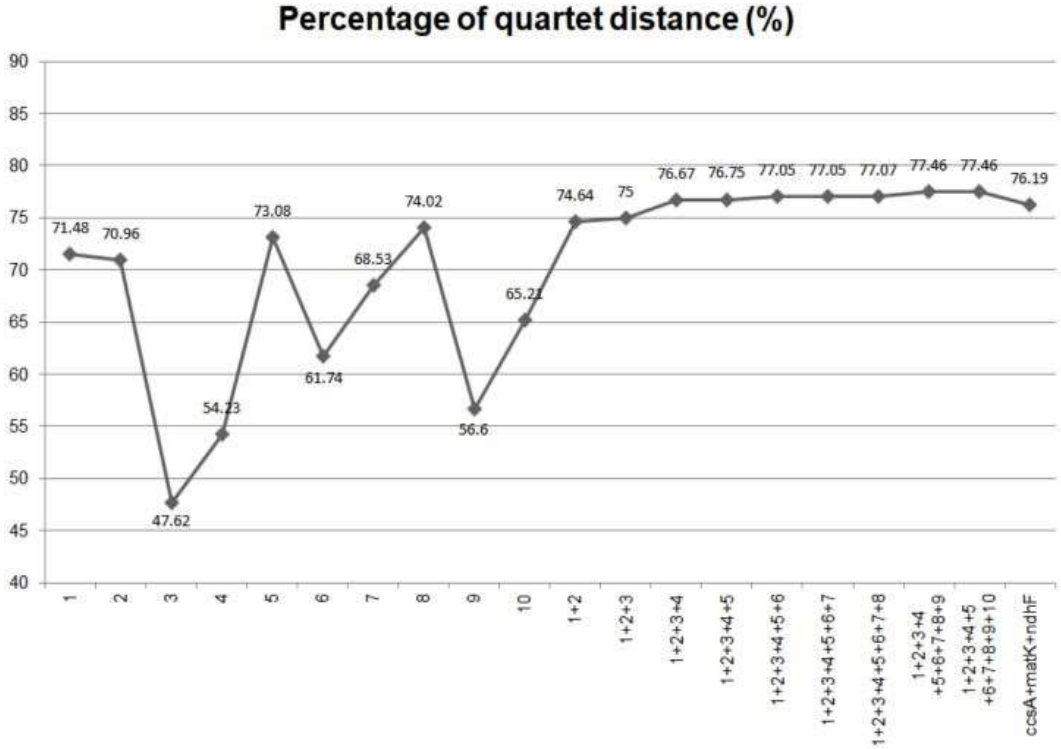


Fig. 17. A graph showing the percentage of quartet distance calculated from each phylogenetic tree. 1, *trnK–rps16*; 2, *ycf1*; 3, *ndhF*; 4, *petA–psbJ*; 5, *petN–psbM*; 6, *ccsA–ndhD*; 7, *trnT–trnL*; 8, *rpoB–trnC*; 9, *rpl32–ccsA*; 10, *rbcL–accD*.

IV. Discussion

1. Characteristics of chloroplast genomes in *Scutellaria*

This chapter is the first to use the complete chloroplast genomes of *Scutellaria* with global sampling. The complete chloroplast genomes of 49 taxa including two outgroup taxa were newly reported (Table 8). The structures of *Scutellaria* chloroplast genomes were highly conserved with no structural rearrangement and gene losses (Figs. 8 and 9).

In chloroplast genome, SSRs have been widely used for molecular systematics and population genetics (Dossa et al., 2017; Gupta et al., 1996; Liu et al., 2018; Peakall et al., 1998). In the subfamily Scutellarioideae, the mononucleotides were detected as the richest (Zhao et al., 2020). Similarly, the mononucleotides were also found the most common in *Scutellaria*, and penta- and hexanucleotide were rare (Fig. 10). Among repeat sequences of 30 bp or more, the complement repeat was not detected in the previous study for Scutellarioideae (Zhao et al., 2020), but complement repeat was confirmed in *S. heterophylla* (Fig. 11) in *Scutellaria*. *S. scutellarioides* showed the highest repeat length (215 bp) than other taxa of *Scutellaria* (Fig. 12). Besides, a total of 20 types of SSR repeat units were found in *Scutellaria* (Fig. 13), and it was confirmed that more types of repeat units were detected than those suggested by Zhao et al. (2020). Similar to the study by Zhao et al. (2020), the mononucleotide repeat unit (A/T),

dinucleotide repeat unit (AT/AT), trinucleotide repeat unit (AAT/ATT), and tetranucleotide (AAAG/CTTT and AAAT/ATTT) were commonly detected in *Scutellaria.* The trinucleotide repeat unit (AAG/CTT) was commonly indicated in all taxa by Zhao et al. (2020), but it was absent in chloroplast genome of *S. orthocalyx* included in this chapter. A/T repeats were also detected, which accounted for the largest percentage of the total SSRs. It was indicated mostly located in non-coding regions. According to previous studies (Menezes et al., 2018; Raubeson et al., 2007; Yang et al., 2013; Zhao et al., 2020), A/T bases are considered to be mainly found in non-coding regions in chloroplast genome of higher plants. Previous studies (Menezes et al., 2018; Raubeson et al., 2007) have suggested that A/T bases are relatively easier to separate strands than G/C bases during replication of chloroplast genome. It was also suggested that the molecular process that causes repetition is more likely to occur in the non-coding region than in the coding region. Repeat sequences cause illegal recombination and mispairing of slipped-strands, resulting in rearrangements and mutations in chloroplast genome. In addition to Zhao et al. (2020), SSRs found in this chapter could be a potential resource for genetic diversity and phylogenetic studies of *Scutellaria.*

2. Phylogenetic relationships based on super-matrix tree

This chapter provides a phylogenetic tree showing improved resolution with a large amount of data for *Scutellaria,* and global sampling was performed including the most taxa than previous studies (Safikhani et al., 2018; Zhao et al., 2017; Zhao et al., 2020). The topology and phylogenetic relationships between the super-matrix tree (Fig. 15) and the previous tree (Fig. 5) were showed similarly. In the super-matrix tree, subgen. *Apeltanthus* was also a monophyletic group, and subgen. *Scutellaria* was a paraphyletic group.

In clade #1, sects. *Apeltanthus, Lupulinaria,* and sp-gr. ORIENTALIS of subgen. *Apeltanthus* were still indicated as paraphyletic groups (Figs 15 and 16). In clade #2, *S. heterophylla* of sect. *Salviifoliae* was a taxon with conspicuous changes of position compared to the previous tree. In the previous tree, sect. *Salviifoliae* was a paraphyletic group, and *S. hetetophylla* was located at the base of the tree and was more closely related to sp-gr. ALBIDA of sects. *Anaspis,* unlike other taxa of sect. *Salviifoliae.* However, in super-matrix tree, *S. heterophylla* was relocated to the sister clade of subgen. *Apeltanthus (*clade #2), sect. *Salviifoliae* was indicated as a monophyletic group, and it was strongly supported. To clearly confirm the positional change of *S. heterophylla*, it was confirmed in the phylogenetic tree analyzed with 52 chloroplast genomes. In that tree (Appendix 10), *S.*

heterophylla was located in the sister clade with subgen. *Apeltanthus*. Additionally, as a result of comparing the three sequences of *S. heterophylla* obtained from Chapter 2 with the sequences of three genes in chloroplast genome, each sequence was also indicated identical. Therefore, it was considered that this positional change was due to the difference in the amount of data included: phylogenetic analyses were analyzed for partial sequences in Chapter 2, but in this chapter, the complete chloroplast genomes were included. *S. baicalensis* of sp-gr. STRIGILLOSA was included in clade #2 differently from the rest of the taxa in this group, and it was also included in clade #2 in the previous tree. Thus, the location of *S. baicalensis* and the classification system of sp-gr. STRIGILLOSA needs to be reexamined. In clade #3 and #4, the phylogenetic relationships between taxa were the same as previous tree. In clade #5, the positions between species groups and taxa of sect. *Scutellaria* were still unresolved. Sp-grs. RESINOSA and INCANA in this clade indicated as the monophyletic groups in both trees. Sp-gr. PARVULA, a monophyletic group in the previous tree, was changed to a paraphyletic group as another sample (*S. parvula*) was also included in the super-matrix tree. *S. mexicana* of sect. *Salazaria,* which was in a consensus position on the previous tree, formed sister clades with several taxa of sect. *Scutellaria* in the super-matrix tree, but was lowly supported. In clade #6, sp-gr. ALBIDA and sect. *Anaspis* were located at the base of the tree in both trees, and the location of *S. ariana* was still unresolved. The

locations of *S. urtriculata*, *S. purpurascens, S. oblonga*, and *S. floridana* were the same as in the previous tree.

In this chapter, a phylogenetic tree of improved resolution was obtained and reported the most data of *Scutellaria* complete chloroplast genomes than the previous studies (Safikhani et al., 2018; Zhao et al., 2017; Zhao et al., 2020). As a result, the phylogenetic relationships were more clearer that subgen. *Apeltanthus* is a monophyletic group and subgen. *Scutellaria* is a paraphyletic group. Also, it was reconfirmed that the basal group of *Scutellaria* was a clade consist of sect. *Anaspis* of subgen. *Scutellaria* and sp-gr. ALBIDA of sect. *Scutellaria*. However, some phylogenetic relationships among the other groups and taxa were still unresolved. This is considered to be due to the existence of completely uncovered data including missing data for each taxon. Therefore, further studies are needed to supplement the weakly supported nodes, including more chloroplast genomes of *Scutellaria*.

3. Screening of the best region for the phylogenetic study

Ten highly variable regions (*trnK-rps16*, *ycf1*, *ndhF*, *petA-psbJ*, *petN-psbM*, *ccsA-ndhD*, *trnT-trnL*, *rpoB-trnC*, *rpl32-ccsA*, and *rbcL-accD)* were confirmed as proper DNA regions for phylogenetic analyses in this chapter (Fig. 14, Table

10). In Chapter 2, three fast-evolving chloroplast genes (*ccsA, matK,* and *ndhF*) suggested in the previous two chloroplast genomes (*S. baicalensis* and *S. insignis*). These genes partially overlapped in only *ccsA* region out of ten regions suggested in this chapter. Therefore, it was needed to suggest proper regions by comparing and evaluating the resolution and topological difference of the phylogenetic tree based on the previous three genes and 10 regions of this chapter for 52 chloroplast genomes. The quartet distance analysis has been widely used in evolutionary research because it shows all topological information of phylogenetic trees and efficiently evaluates the degree of similarity between phylogenetic trees (Denk et al., 2019; Larridon et al., 2020; Pease et al., 2016; Wangner et al., 2019). The percentage of quartet distances showed that it is more efficient for phylogenetic analysis to include at least four regions among ten regions with high nucleotide diversity rather than using each region. It was confirmed that the percentage of quartet distance for the phylogenetic tree of three genes in Chapter 2 is higher than that of a tree using three regions out of ten regions in this chapter. Therefore, the reliability of the phylogenetic tree obtained in Chapter 2 is confirmed, and it is considered most useful to analyze the phylogenetic tree, including nine or more of the ten regions with high nucleotide diversity. The regions evaluated in this chapter will use as the basis for further phylogenetic studies and DNA barcoding studies of *Scutellaria*.

In Scutellarioideae, Zhao et al. (2020) also suggested ten highly variable regions (*psbA*−*trnH*, *trnK*−*rps16*, *petN*−*psbM*, *rbcL*−*accD*, *petA*−*psbJ*, *petB*−*petD*, *ndhF*, *rpl32*−*trnL*, *rps15*−*ycf1*, and *ycf1*). Among these regions, *trnK*−*rps16*, *petN*−*psbM*, *rbcL*−*accD*, *petA*−*psbJ*, *ndhF*, and *ycf1* are the same as this chapter. Future studies should evaluate the quartet distance of the phylogenetic trees based on these regions by additionally chloroplast genomes of Scutellarioideae.

CHAPTER 4.

Phylogeny of *Scutellaria* based on 272 nuclear single-copy genes using Hyb-Seq technique

I. Introduction

The development of next-generation sequencing (NGS) technology has allowed for the efficient sequencing of large amounts of data and reducing the sequencing costs (Cronn et al., 2012; Egan et al., 2012; Mamanova et al., 2010; Van et al., 2014). Hyb-Seq, combined with NGS technology, is defined as the hybridization-based target-enrichment sequencing (Weitemier et al., 2014). General NGS technology provides many nucleotide sequences at random regions, whereas the Hyb-Seq method produces specific probes (baits) for a specific gene region and collects only the target region to determine their nucleotide sequence. This approach has some advantages rather than randomly sequencing, 1) allows multiplex sequencing the entire genome, permitting more samples to be sequenced cost-effectively. Also, 2) the complexity of the analysis is greatly reduced by targeting only a portion of the genome that is necessary or beneficial for phylogenetic studies (Grover et al., 2012). Generally, Hyb-Seq is known that combines the target enrichment method with the genome skimming technique and efficiently generates both the low-copy nuclear genome and high-copy elements (e.g., organellar genomes) appropriate for phylogenomic analyses in plants (Dodsworth et al., 2019; Weitemier et al., 2014). The use of low-copy nuclear gene sequences is considered important in resolving the taxonomic relationship (Curto et al., 2012). The

low-copy nuclear loci are potentially rapid-evolving and represent the independent gene history, and typically targeted in Hyb-Seq analyses as probes (Curto et al., 2012; Dodsworth et al., 2019; Schmickl et al., 2016; Yuan et al., 2009). A general process for efficient Hyb-Seq is first finding a single-copy gene (SCG) for each taxon. MarkerMiner (Chamala et al., 2015) is an easy-to-use and effective tool for extraction of SCGs. It provides previously known genome data sets of 16 taxa and extracts SCGs from the transcriptome of specific taxa as a reference. Probes are produced based on the obtained SCGs, and the next process is performed to increase capture efficiency. In addition, through shearing, ligation, and capture processes, a large number of nucleotide sequences are obtained. These raw data are mapped to the previous probe sequences to obtain each sequence of the specific gene. Finally, after the sequence obtained for each gene is aligned, it concatenates to perform phylogenetic analysis. HybPiper pipeline (Johnson et al., 2016) is commonly used to assemble and align Hyb-Seq data in plants (Brewer et al., 2019; Chau et al., 2018; Couvreur et al., 2019; Herrando-Moraira et al., 2019; Johnson et al., 2019; Morales-Briones et al., 2018). This pipeline is a user-friendly platform that separately assembles each gene region, extracts exon and intron sequences, and identifies paralogous gene copies. The general process of Hyb-Seq is presented in Fig. 18. The data obtained using Hyb-Seq includes not only the exon sequence but also the intron

and intergenic spacer sequences. Therefore, this data can be used to provide a phylogenetic resolution between related taxa and can be applied to DNA barcoding based on this data, and it also provides high-quality data to detect evolutionary events such as ancient hybridization.

In this chapter, 641 nuclear SCGs were designed for the probes, and 96 taxa including outgroup taxa were analyzed. Through a phylogenetic analysis using the SCGs, this chapter aims to: 1) compare with the phylogenetic trees obtained in the previous chapters; 2) to investigate the discordance among the gene trees and the species trees; and 3) to reexamine the phylogenetic relationship in *Scutellaria* that has not been solved before.

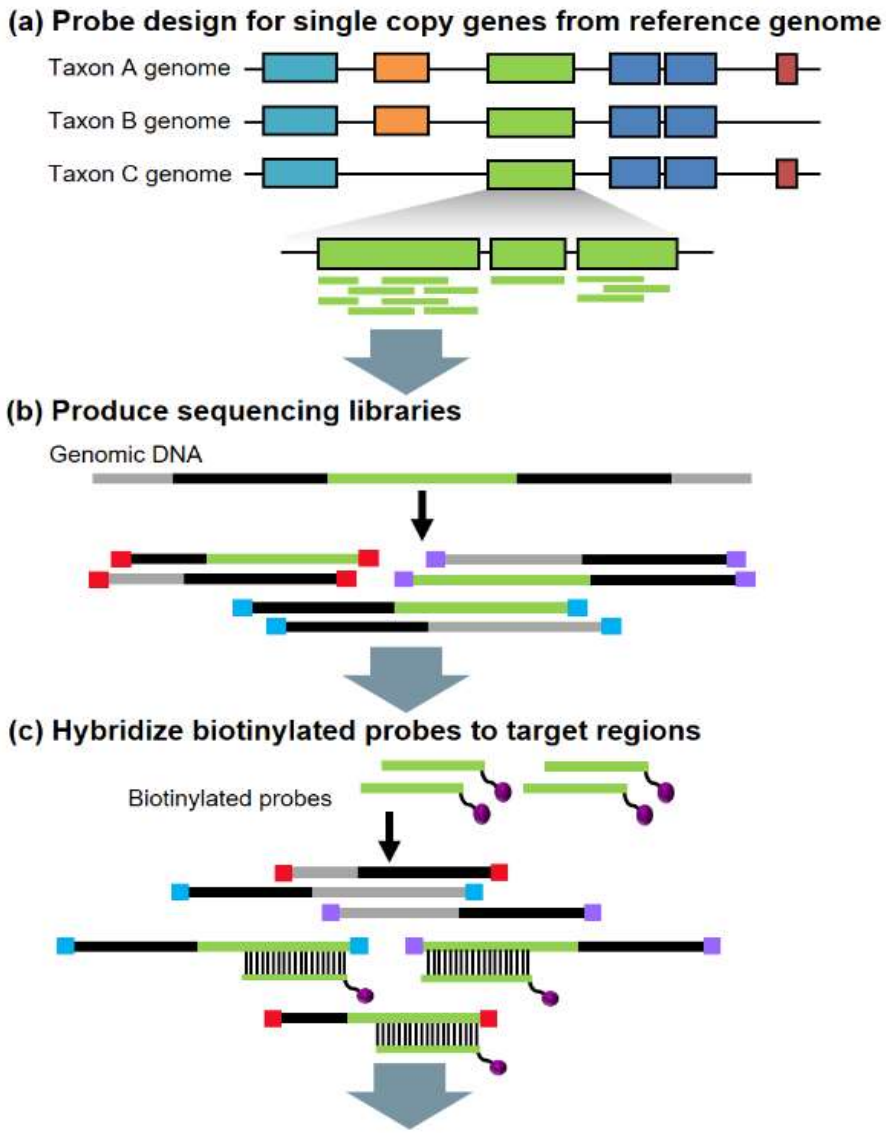


Fig. 18. Workflow of the Hyb-Seq technique. (a) Sigle-copy genes are derived from the closely related reference genome, and the probes are designed. (b) After the extracted DNAs are fragmented, index sequences distinguishing each taxon are ligated at both ends of the fragments. (c) Biotin is attached to the ends of the probes, and the biotinylated probes and DNA libraries are hybridized.

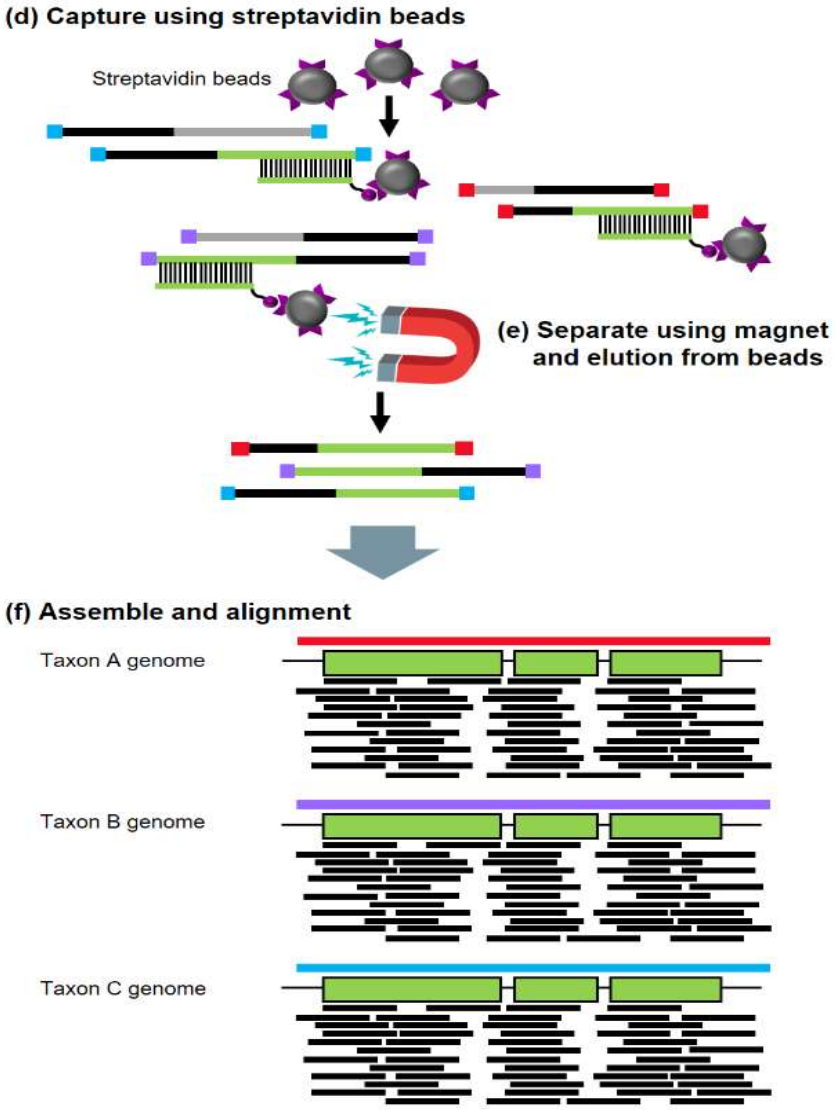


Fig. 18. Continued. (d) Streptavidin beads that have magnetic attraction are added, and it binds to biotin. (e) Only hybridized DNA fragments are captured by magnets and DNA fragments eluted from the beads. (f) A large number of nucleotide sequences are obtained from captured DNA fragments through NGS sequencing, and a matrix for phylogenetic analysis is obtained through assembly and alignment processes.

II. Materials and Methods

1. Probe design

To generate DNA probes used to capture orthologous genes, the previously published transcriptome data of *S. baicalensis* Georgi (Liu et al., 2015) were used, the following transcriptome data sets were obtained from the Sequence Read Archive (http://www.ncbi.nih.gov/sra): root (accession number SRX1700376), leaves (accession number SRX1700375), flower (accession number: SRX726670). The data sets were downloaded using the fastq-dump utility of the SRA toolkit from NCBI (http://www.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=toolkit_doc&f=fastq-dump). The total amount of raw data set was 13.8 G, and the adapter sequences and low-quality nucleotides were trimmed using Trimmomatic v.0.38 (Bolger et al., 2014). The trimming process was conducted with LEADING:25 TRAILING:25 SLIDINGWINDOW:4:25 MINLEN:30. The trimmed data were assembled using Trinity v.2.6.6 (Grabherr et al., 2011) with default settings, and a total of 71,666 genes were obtained.

The assembled data of *S. baicalensis* and the SCG sequences of previously published four taxa [Duarte et al., 2010; *Arabidopsis thaliana* (L.) Heynh., *Populus trichocarpa* Torr. & A.Gray ex Hook., *Oryza sativa* L., *Vitis vinifera* L.] were compared using MarkerMiner v.1.0 (Chamala et al., 2015). The relationship between the number of orthologous genes shared by five taxa was indicated in Fig. 19.

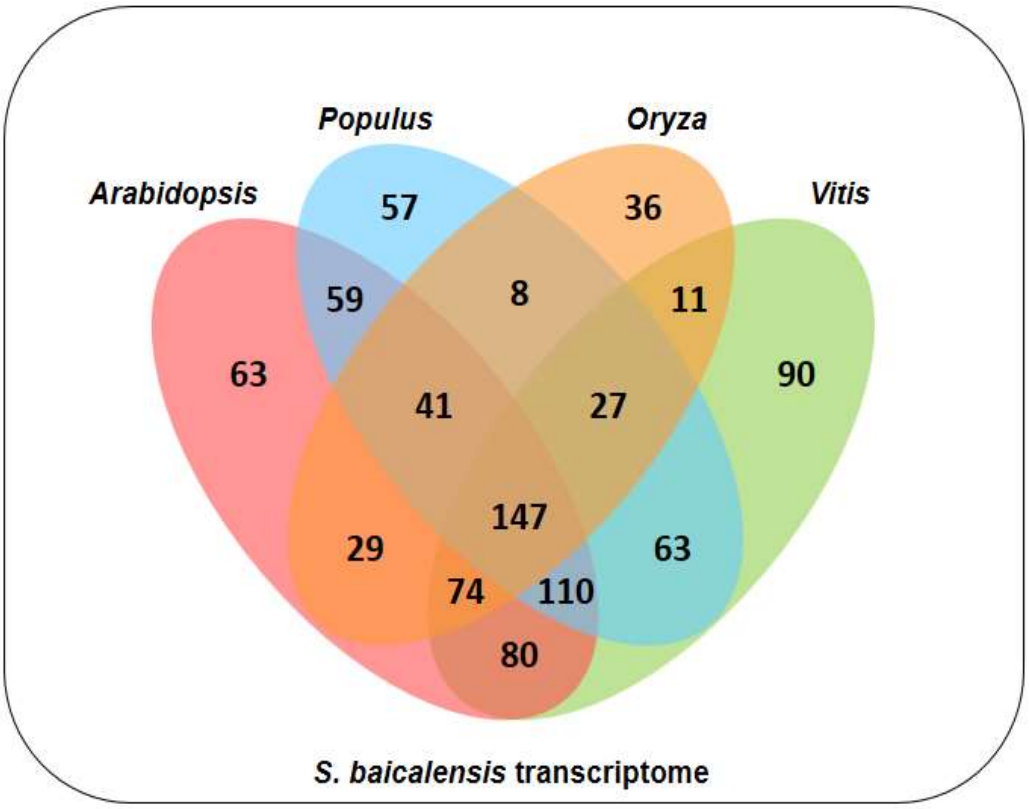


Fig. 19. A Venn diagram indicating the number of shared single-copy genes among *S. baicalensis*, *Arabidopsis thaliana*, *Populus trichocarpa*, *Oryza sativa*, and *Vitis vinifera*.

Among a total of 895 sequences shared by the five taxa, 650 SCGs shared by two or more taxa including *S. baicalensis* were selected. Open reading frames were found in each SCG using Geneious Prime (Kearse et al., 2012), and the sequences were extracted (938,758 bp). These SCG data was sent to RAPiD Genomics, LLC (Florida, USA) and used for probe design. By mapping the transcriptome data of *S. baicalensis* for 650 genes using GMAP (Wu et al., 2016), exon and intron boundaries and probe coverage were screened. Eight genes with a probe coverage of less than 50% were excluded from probe design. As a result, 120-mer probes based on 641 SCG sequences were designed. The probes were also designed by 3× tiling to increase the capture efficiency of probes and to obtain more sequences of the intron region close to the exon from the fragments containing the exon and the intron boundaries. Information of the 650 SCGs used in the probe design is shown in Appendix 11.

2. Taxon sampling, DNA extraction, and sequencing

A total of 96 taxa were sampled, including three outgroup taxa (*Holmskioldia sanguinea*, *Tinnea aethiopica*, and *Wenchengia alternifolia*) closely related to *Scutellaria* (Li et al., 2016). Compared with the Paton classification system (Paton, 1990), all subgenera (subgen.) and sections (sect.) were included, and 33 species groups (sp-gr.) were covered (86.8 %). The information of samples is

shown in Appendix 12.

DNA extraction and amplification were performed in the same methods in Chapter 3. The concentration of all samples was normalized from 30 ng/㎕ to 50 ng/㎕, and the sample volume was normalized from 35 ㎕ to 50 ㎕. To confirm the quantity and quality of gDNA, 1 ㎕ of gDNA was electrophoresed for 10 minutes at 135 V on a 1.3 % agarose gel (Appendix 13). Library preparation and hybrid sequencing of all samples were performed by RAPiD Genomics, LLC (Florida, USA).

The obtained gDNA of each sample was conducted random mechanical shearing process for preparing a DNA library, and some degraded samples as a result of electrophoresis reduced the shearing. The sheared DNA fragments were conducted end-repair reaction, and the adenine residue was ligated to the 3' end of the blunt end fragment, allowing the ligation of the barcode adapters. The barcode adapters for sequencing and identifying each taxon were ligated to the blunt-end, and PCR-amplification prepared the libraries. Libraries for 16 samples were pooled, and DNA fragments were hybridized with the biotinylated probe using SureSelect probes (Agilent Technologies: Santa Clara, CA). Subsequently, solution-based target capturing was performed using streptavidin and magnets. Data was sequenced with paired-end 100 bp reads of captured DNA fragments with Illumina Hiseq 3000 platform (Illumina Inc., San Diego, CA, USA).

3. Raw data analyses

The quality of raw data was confirmed using FastQC v.0.11.5 (Andrews et al., 2011). The raw data sequences with a Phred score of 20 or less were cutoff using Trimmomatic v.0.38 (Bolger et al., 2014) and trimmed with the following options: LEADING:20 TRAILING:20 SLIDINGWINDOW:4:20 MINLEN:30 (Johnson et al., 2016). All assembly processes for trimmed data were performed by HybPiper v1.3.1 pipeline (Johnson et al., 2016). First, paired reads and unpaired reads were mapped by BWA (Li and Durbin, 2009) to target genes used in probe design. Aligned reads were sorted in each gene directory by the Python wrapper of SAMtools (Li et al., 2009). For trimmed and filtered reads sorted in each gene directory, target enrichment efficiency was calculated as a percentage. To visualize mapping efficiency of each sample for target genes, a heatmap was obtained by executing gene_recovery_heatmap.R script in HybPiper using R studio v.1.2.5033 (Team, 2015). Reads were individually assembled based on *de-novo* assembly for each gene by SPAdes (Bankevich et al., 2012) and assembled into contigs. Subsequently, the following analyzes were conducted by Exonerate (Slater and Birney, 2005): 1) the coding sequences from which introns were removed were extracted, 2) supercontigs with assembled contigs combined exon and intron were generated, and 3) true intron-exon junctions were detected. Intron regions flanking the exons were identified by intronerate.py from the generated supercontigs. Supercontigs in

each gene were aligned by MAFFT v7.471 (Katoh and Standely, 2013), and FFT-NS-1 option was used as a progressive method with a –leaveappyregion option minimize unnecessary gaps. Each of the aligned sequences was concatenated as a concatenated matrix by AMAS (Borowiec, 2016), and statistics for the matrix were calculated.

4. Phylogenetic analyses

The phylogenetic analyses for 95 taxa were performed excluding one taxon (*S. oblonga*) which was few sequences mapped to target genes due to low capture efficiency on HybPiper pipeline. 369 genes detected as paralogs were excluded among 641 target genes, and loci for 272 genes were included in the phylogenetic analyses.

Two methods for phylogenetic analyses were conducted, the first is a concatenation approach using a concatenated matrix (concatenated ML tree), and the second is a coalescent approach using the individual gene trees obtained by phylogenetic analysis for each locus (coalescent ASTRAL tree). The concatenation approach was conducted using IQ-TREE v1.6.12 (Nguyen et al., 2015) and Standard RAxML v 8.2.12 (Stamatakis, 2014) based on Maximum Likelihood (ML) analyses. In IQ-TREE, the substitution model for analyzing the phylogenetic tree was selected as the GTR + F + G4 model by ModelFinder, and 1,000 ultrabootstraps were replicated. In RAxML, 500 rapid bootstraps were replicated with the GTR + I +

Γ model. In the coalescent approach, 272 gene trees were first obtained from each loci using RAxML with GTR + I + Γ model and replicated 200 bootstraps. To increase the accuracy by reducing the noise of the phylogenetic tree, the low support branches (bootstrap support value of 33 or less) in the obtained gene trees were collapsed (Zhang et al., 2018) using nw_ed in the Newick Utility package v.1.6.0 (Junier and Zdobnov, 2010). Based on the obtained gene trees, the coalescent ASTRAL tree was inferred using ASTRAL-III v.5.7.3 (Zhang et al., 2018). Phyparts (Smith et al., 2015) were used to confirm the discordance and concordance information of each node in the coalescent ASTRAL tree obtained from ASTRAL-III. Statistics for the information obtained using Phypart were visualized using phypartspiecharts.py (available from https://github.com/mossmatters/phyloscripts/tree/master/phypartspiecharts) in Phyloscripts package v.0.1 (Resl, 2015). To compare the phylogenetic tree obtained from the concatenation and coalescent approach simultaneously, a tanglegram was drawn using the R package (R Core Team, 2019) dendextend (Galili, 2015).

The number of shared quartets were calculated between the concatenated ML tree and the individual gene tree using the R package Quartet R package (Sande et al., 2014; Smith, 2019) to quantify the phylogenetic fidelity of each gene tree (Pease et al., 2018; Villaverde et al., 2020). From the obtained quartet distance statistics, the number of quartets shared with the concatenated ML

tree was sorted in the order of a large number. The phylogenetic analyses were performed based on loci with a shared quartet percentage of 80 % or more. Each loci matrices for shared quartet percentage of 80 % or more were accumulated in order, and IQtree analysis was performed with the same concatenation analysis option. The number of quartets shared by these subsets of loci and the concatenated ML tree was recalculated to investigate the best subset for the loci.

III. Results

1. Sequence analyses

The average number of raw reads per sample was 8,904,762 and ranged from 28,506 to 71,675,046 (Table 11, Appendix 14). The trimmed average number of reads per sample by Trimmomatic was 8,709,882 (97.98 %) and ranged from 27,806 (97.54 %) to 70,046,632 (97.73 %).

According to statistics obtained by the HybPiper pipeline, the average number of reads mapped to the target gene was 5,760,166 (44.34 %), and ranged from 4,150 (0.60 %) to 48,590,504 (87.50 %). The average number of genes mapped to the target genes was 632 genes (99.9 %), and the sample with the lowest mapping efficiency was mapped on 281 genes (43.84 %). Excluding 22

samples, 74 samples showed 100 % mapping efficiency. The average of the sum of captured exons per sample assembled by HybPiper was 615,224 bp, which corresponds to 65.5 % of 938,758 bp of 641 loci used for probe design (Table 11).

Table 11. Statistical summary of sequencing and assembly results using HybPiper

	Min.	Max.	Avg.
No. of paired reads	14,253	35,837,523	4,452,381
No. of reads	28,506	71,675,046	8,904,762
No. of trimmed reads	27,806	70,046,632	8,709,882
Percent of trimmed reads (%)	97.26	98.93	97.98
No. of reads on target	4,150	48,590,504	5,760,166
Percent of reads on target (%)	0.6	87.5	44.34
No. of genes mapped	281	641	632
No. of genes with contigs	9	641	513
No. of genes with sequences	1	641	508
ParalogWarnings	0	282	15
Total length of captured exons (bp)	147	929,499	615,224
Percent of captured exons (%)	0.02	99.01	65.54

In the heatmap visualized for the capture efficiency targeting 272 loci except for the 369 paralogs loci appeared that the capture efficiency of some samples significantly decreased (Fig. 20). Among 96 taxa, the capture efficiency of *S. oblonga* was only 0.6 %, and there was no data after excluding paralogs loci, so it was not included in phylogenetic analysis. As a result of alignment by MAFFT for phylogenetic analysis based on 95 taxa and 272 loci, the total length of the concatenated matrix was 965,469 bp (Table 12). According to statistics obtained using AMAS (Table 12, Appendix 15), the average length of each aligned loci was 3,550 bp, ranging from 1,259 bp to 10,724 bp. Of the total sequence of the concatenated matrix, 559,073 bp (57.90 %) were found as variable sites, and 309,071 bp (32.01 %) were parsimony informative sites.

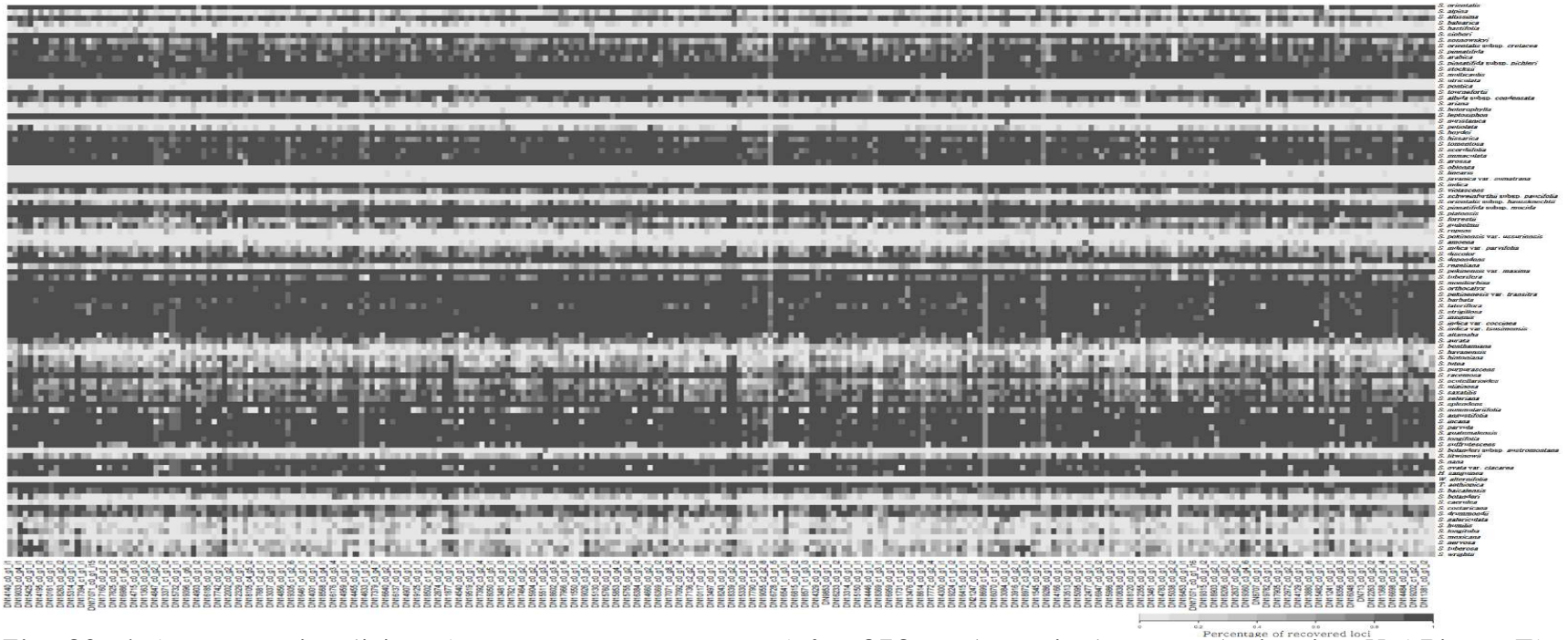


Fig. 20. A heatmap visualizing the percentage recovered for 272 nuclear single-copy loci using HybPiper. The x-axis shows each locus, and the y-axis shows each taxon name. It shows the ratio of the recovered length to the length of the target loci used in the probe design, and the darker grey the higher the recovery ratio.

Table 12. Statistical summary of concatenated matrix

	Concatenated matrix	Min.	Max.	Avg.
No. of taxa	95	33	83	66
Alignment length (bp)	965,469	1,259	10,724	3,550
Total matrix cells	64,624,371	43,593	804,300	237,590
Undetermined characters	18,236,867	10,997	256,753	67,047
Missing percent (%)	28.22	9.24	36.19	27.53
No variable sites	559,073	695	6,089	2,055
Proportion variable sites	0.58	0.43	0.78	0.58
Parsimony informative sites	309,071	398	3,238	1,140
Proportion parsimony informative	0.32	0.23	0.46	0.32
Proportion AT content	0.61	0.51	0.65	0.61
Proportion GC content	0.39	0.35	0.49	0.39

2. Phylogenetic analyses based on concatenation approach

As a result of the concatenated ML analysis based on 272 loci using IQ-TREE and RAxML, most of the branches are strongly supported (Fig. 21, Appendix 16). The topology of IQ-TREE and RAxML are identically, and the description focused on IQ-TREE (concatenated ML tree) because bootstrap support (BS) values of IQ-TREE are slightly higher.

In the concatenated ML tree, six major clades (clade #1–6) were recognized and strongly supported (BS ≥ 98). In clade #1, subgen. *Apeltanthus* was a monophyletic group with strongly supported value (BS = 100). Sects. *Lupulinaria* and *Apeltanthus* in clade #1 indicated also monophyletic group and highly supported (BS = 100), but species groups (sp–grs. ORIENTALIS and LINEARIS) in sect. *Lupulinaria* were not clearly recognized as a monophyletic group. In subgen. *Scutellaria,* sects. *Scutellaria, Salviifoliae,* and *Anaspis* not appeared monophyletic groups. Sect. *Salviifoliae,* two species (*S. heterophylla* and *S. pontica*) were located in different clade (clades #2 and #3). In clade #2, 10 taxa of subgen. *Scutellaria* showed a close relationship with subgen. *Apeltanthus* and strongly supported (BS ≥ 90). Especially, a clade including four taxa (*S. baicalensis*, *S. forrestii*, *S. orthocalyx*, and *S. schweinfurthii* subsp. *paucifolia)* of sect. *Scutellaria* was sister to subgen. *Apeltanthus* of clade #1, and *S. amoena* appeared as a sister clade of all other taxa included in clade #2. In clade #3, sp–grs. RACEMOSA and PARVULA were monophyletic groups with strong support values (BS = 100), and the rest of the species group was not clearly recognized as a monophyletic group. Sp–gr. RESINOSA was also a monophyletic group in clade #4 (BS = 100), and it was sister to a clade including sp–gr. INCANA (*S. incana* and *S. altamaha*)and *S. ovata* var. *clacarea* (BS = 100). The rest of the species group in clade #4 was mixed–up and not resolved. *S. mexicana* of sect. *Salazaria* formed a clade of all other taxa included

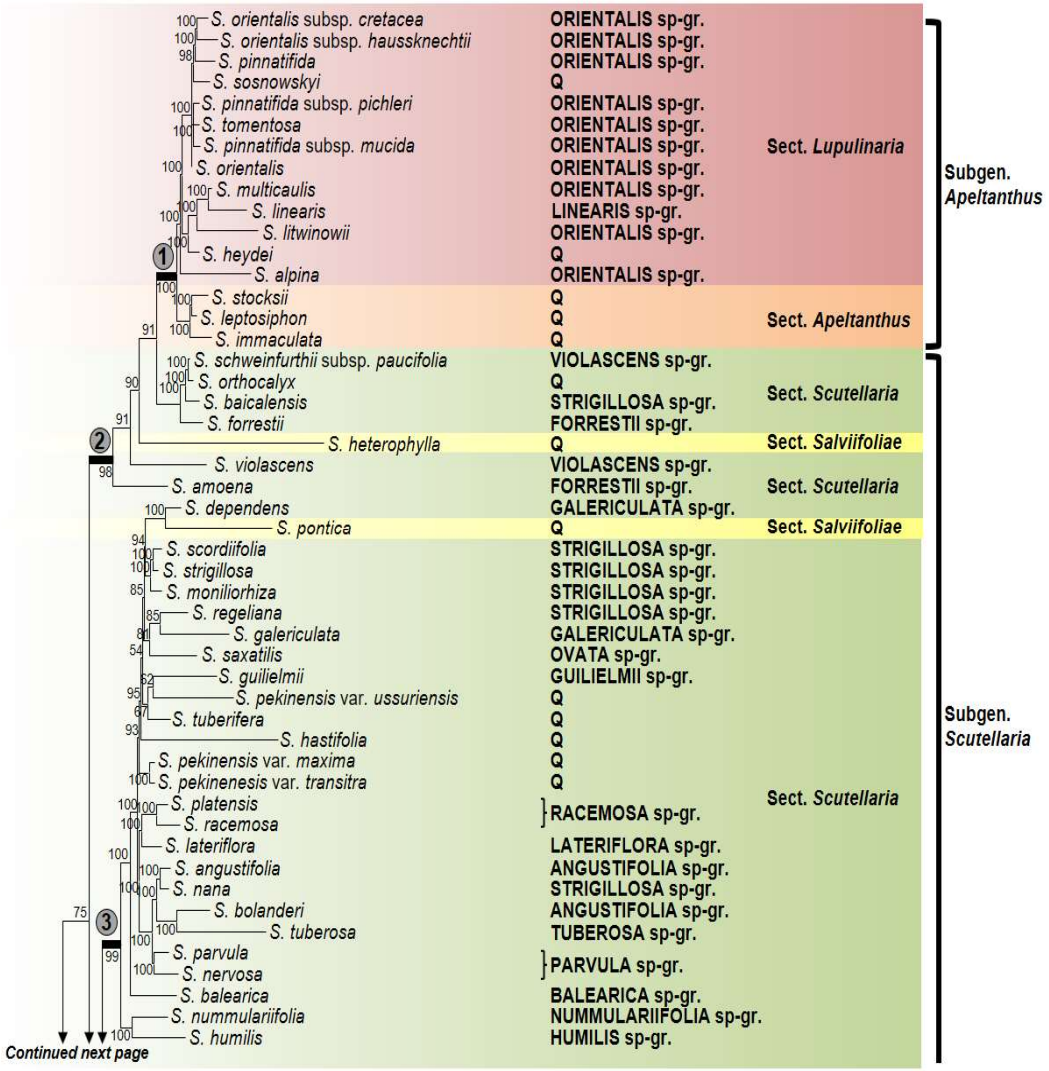


Fig. 21. A concatenated ML tree obtained using IQ-TREE based on 272 nuclear single-copy loci. The numbers above the branches are bootstrap values. Circled numbers (#1-6) correspond to major clades described in the text. Colored boxes are sections and capitalized names after the taxon names are "species groups" recognized in Paton (1990). Q" indicates a taxon of which taxonomic position (sections and "species group") have not been reported.

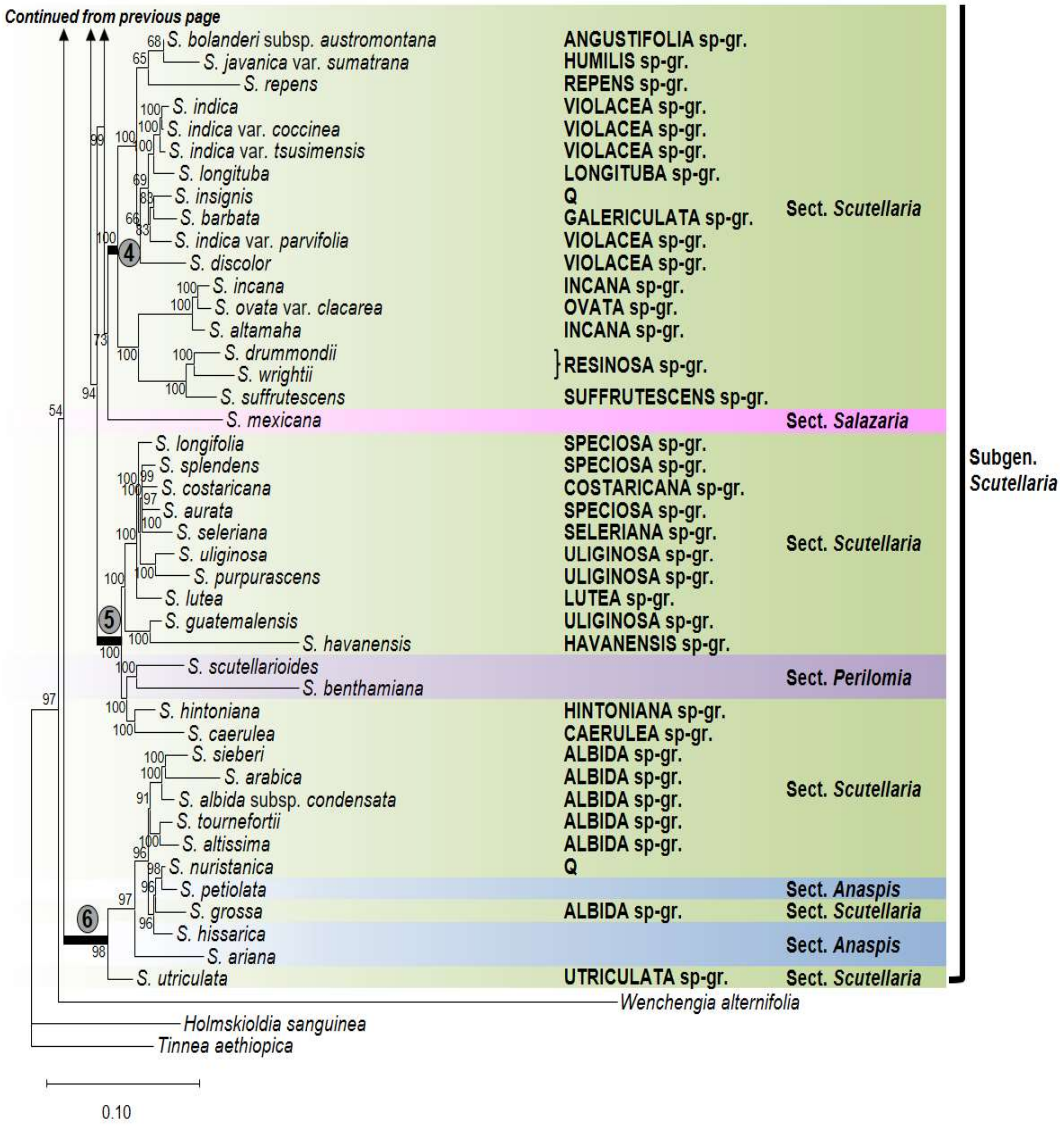


Fig. 21. Continued.

in clade #4 and highly supported (BS = 100). Sect. *Perilomia* was indicated a monophyletic group (BS = 100) and sister to a clade containing *S. hintoniana* (sp-gr. HINTONIANA) and *S. caerulea* (sp-gr. CAERULEA) of sect. *Scutellaria* (BS = 100). The rest of the species group in clade #5 was not clearly resolved. In clade #6, located at the base of the phylogenetic tree, it was indicated as a sister group of all other species but weakly supported (BS = 54). This clade included six taxa of sp-gr. ALBIDA, three taxa of sect. *Anaspis* and *S. utriculata* of sp-gr. UTRICULATA. sp-gr. ALBIDA and sect. *Anaspis* were not monophyletic groups (BS ≥ 91), and *S. utriculata* was a sister to the remaining species of clade #6 and highly supported (BS = 98).

The quartet distance between the concatenated ML tree and the individual gene trees was calculated in order to quantify the phylogenetic fidelity of each gene tree (Fig. 22; Appendix 17). The average shared quartet percentage for individual 272 loci was 59.73 % and ranged from 6.34 % to 87.27 %. The gene tree of the loci with the highest shared quartet percentage was the DN17712_c1_g5_i1 loci (Gene no. in *Arabidopsis*: AT3G54460), which was related to the protein containing the SNF2 domain, the helicase domain, and the F-box family protein. The rest of the seven loci were as follows (in order of highest shared quartet percentage): DN6748_c0_g1_i1(AT1G74460), DN18388_c0_g1_i1 (AT2G01320), DN18903_c0_g2_i1 (AT4G24270), DN10161_c0_g1_i1 (AT2G35720), DN9853_c0_g1_i1 (AT1G09300),

DN19026_c3_g5_i1 (AT1G14030), DN17366_c0_g12_i1 (AT2G38270). As a result of plotting each shared quartet percentage as a graph, an inflection point appeared at 80 % (Fig. 22), and the quartet distance was recalculated for subsets of loci with 80 % or more. The shared quartet percentage was recalculated by a subset of the top eight loci (Fig. 23), and the result of analyzing only one locus (DN17712_c1_g5_i1) indicates the highest shared quartet percentage.

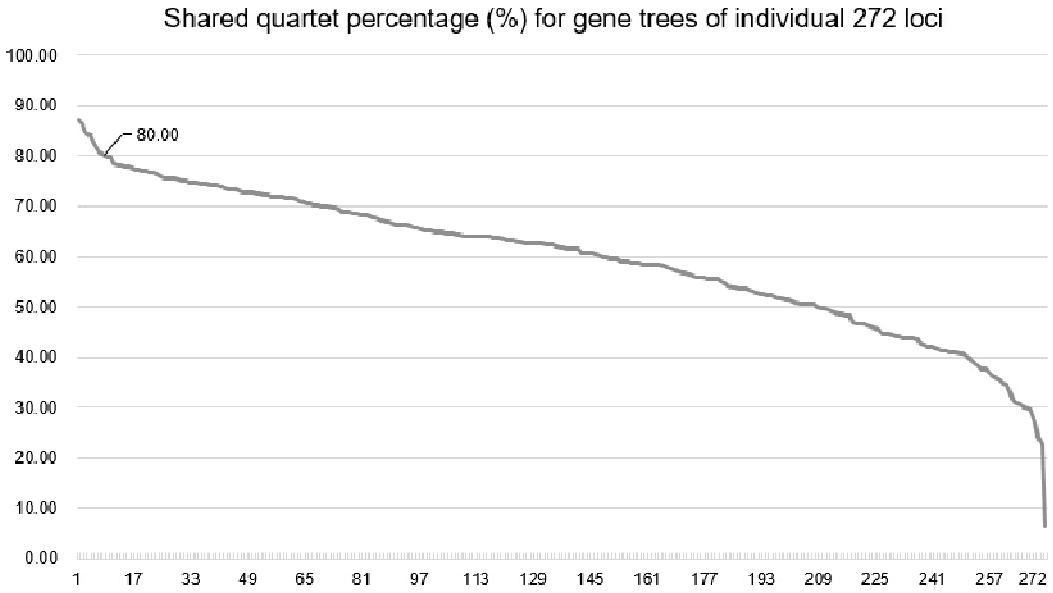


Fig. 22. A graph plotted by sorting in descending order of the shared quartet percentage of individual 272 loci with the concatenated ML tree. A point indicated with 80.00 % is the inflection point.

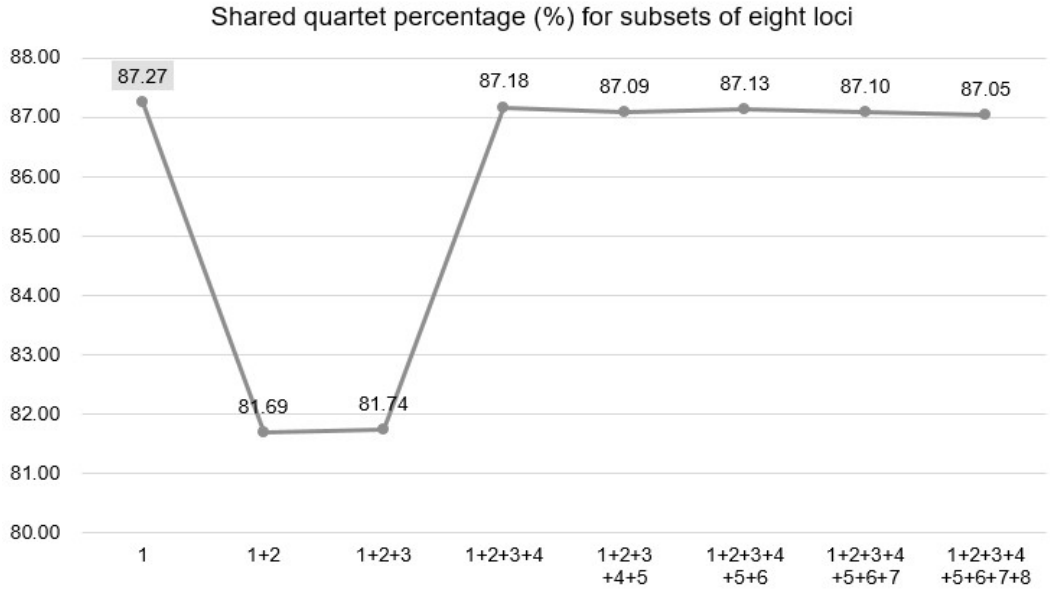


Fig. 23. A graph of recalculating the shared quartet percentage by combining the subset of top eight loci: 1, DN17712_c1_g5_i1 (AT3G54460 homolog); 2, DN6748_c0_g1_i1 (AT1G74460 homolog); 3, DN18388_c0_g1_i1 (AT2G01320 homolog); 4, DN18903_c0_g2_i1 (AT4G24270 homolog); 5, DN10161_c0_g1_i1 (AT2G35720 homolog); 6, DN9853_c0_g1_i1 (AT1G09300 homolog); 7, DN19026_c3_g5_i1 (AT1G14030 homolog); 8, DN17366_c0_g12_i1 (AT2G38270 homolog).

3. Phylogenetic analyses based on coalescent approach

In the coalescent ASTRAL tree (Fig. 24), it was confirmed that the coalescent ASTRAL tree has a lower local posterior probability (LPP $\leq$ 0.9) than the concatenated ML tree. In tanglegram (Fig. 25), the coalescent ASTRAL tree was shown concordance in most of the nodes comparing with the concatenated ML tree, but

discordance appeared in some nodes. The six major clades were commonly recognized in both trees. However, in the coalescent ASTRAL tree, the location of some taxa included in clades #2–4 was different from the concatenated ML tree (Fig. 21). *Wenchengia alternifolia*, included as an outgroup, was located in an unresolved position different from other species in the outgroup (LPP = 0.33; Fig. 24), so it was excluded from the tanglegram. The taxonomic relationships of sect. *Lupulinaria* of subgen. *Apeltanthus* showed the most conflict discordance. In subgen. *Scutellaria,* some confilct was indicated in the rest of the clades except for the base of the tree. The piechart obtained using PhyParts analysis shows the degree of discordance of individual gene trees along with node (Fig. 26). Of the three clades recognized identically to the concatenated ML tree, clades #1 and #5 confirmed that the relationship among the gene trees and the coalescent ASTRAL tree was congruent (blue proportion), and clade #6 confirmed that a large proportion of gene trees were not fully supported and uninformative (gray proportion). Sect. *Lupulinaria* of subgen. *Apeltanthus* and sect. *Scutellaria* of subgen. *Scutellaria* were mostly of nodes with a large proportion of red, which showed that most gene tree topologies conflict with the coalescent ASTRAL tree.

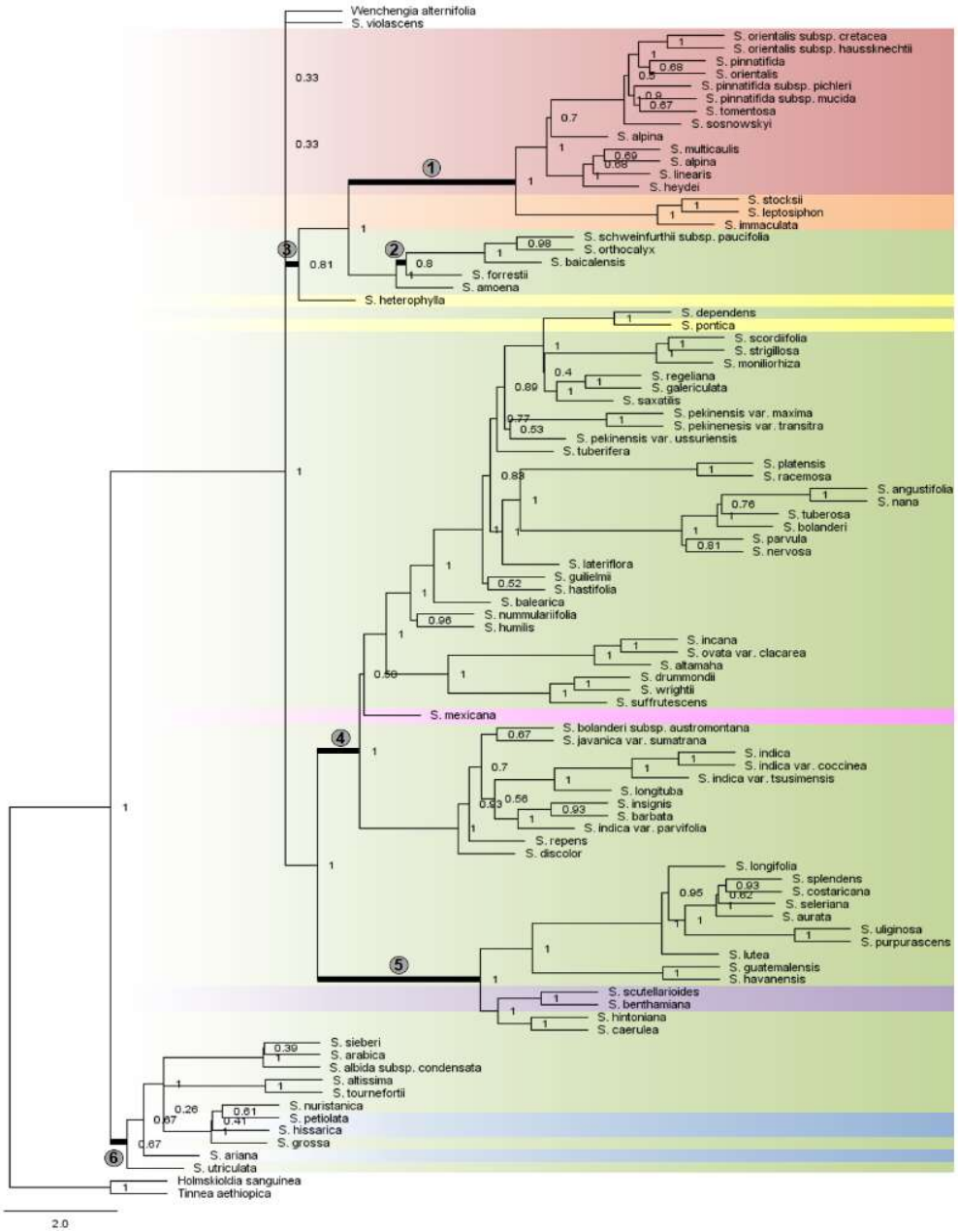


Fig. 24. A coalescent ASTRAL tree obtained using ASTRAL based on 272 nuclear single-copy loci trees. The numbers above the branches are local posterior probabilities (LPP). Bold lines and circled numbers indicate major clades recognized in the concatenated ML tree.

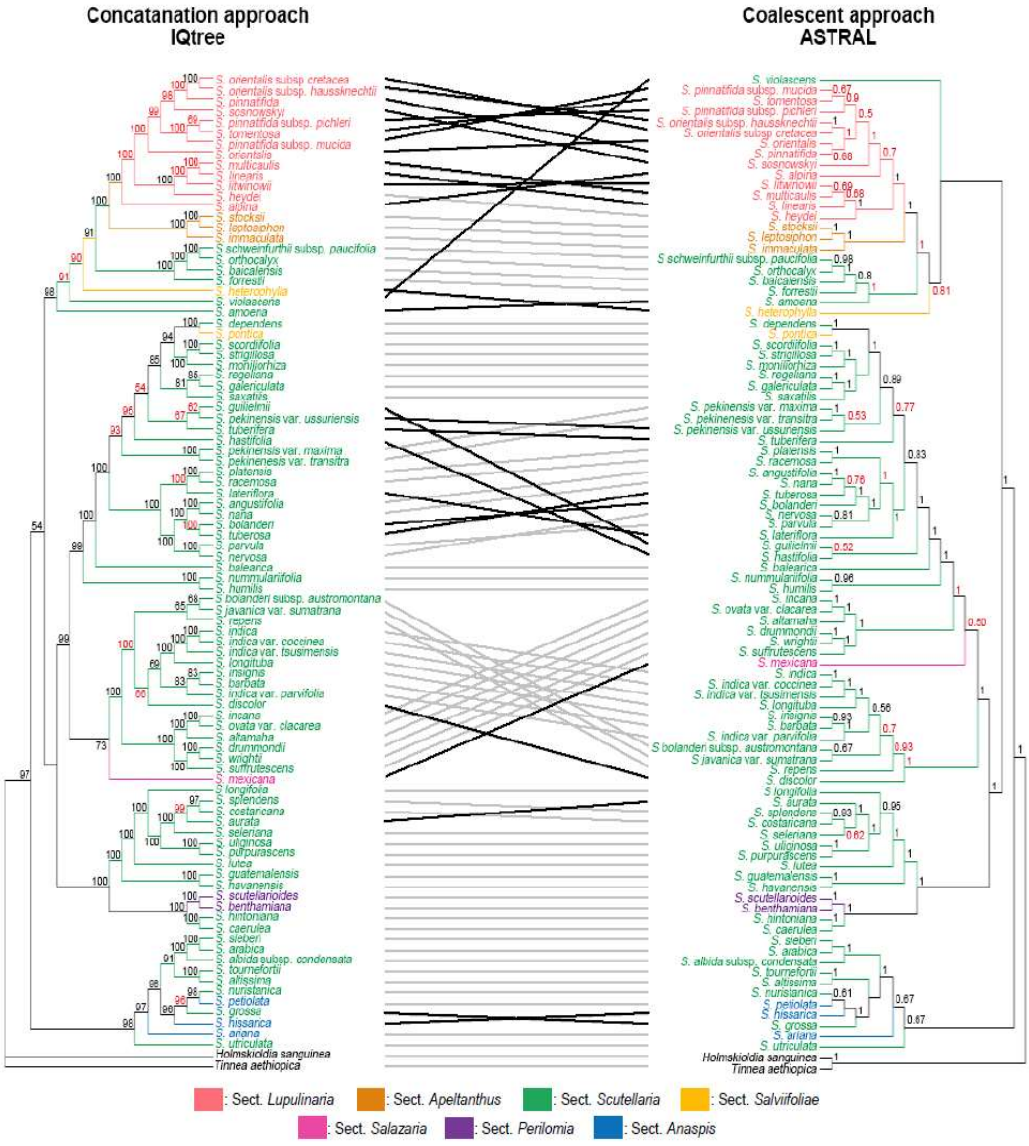


Fig. 25. A tanglegram for comparing the concatenated ML tree (left) and the coalescent ASTRAL tree (right). The numbers above each branch are bootstrap values in the left tree and local posterior probability in the right tree. The black connecting lines between the two trees indicate the discordance between two trees, and the grey connecting lines indicate the concordant. The BS and LPP in red also indicate the discordance between the two trees.

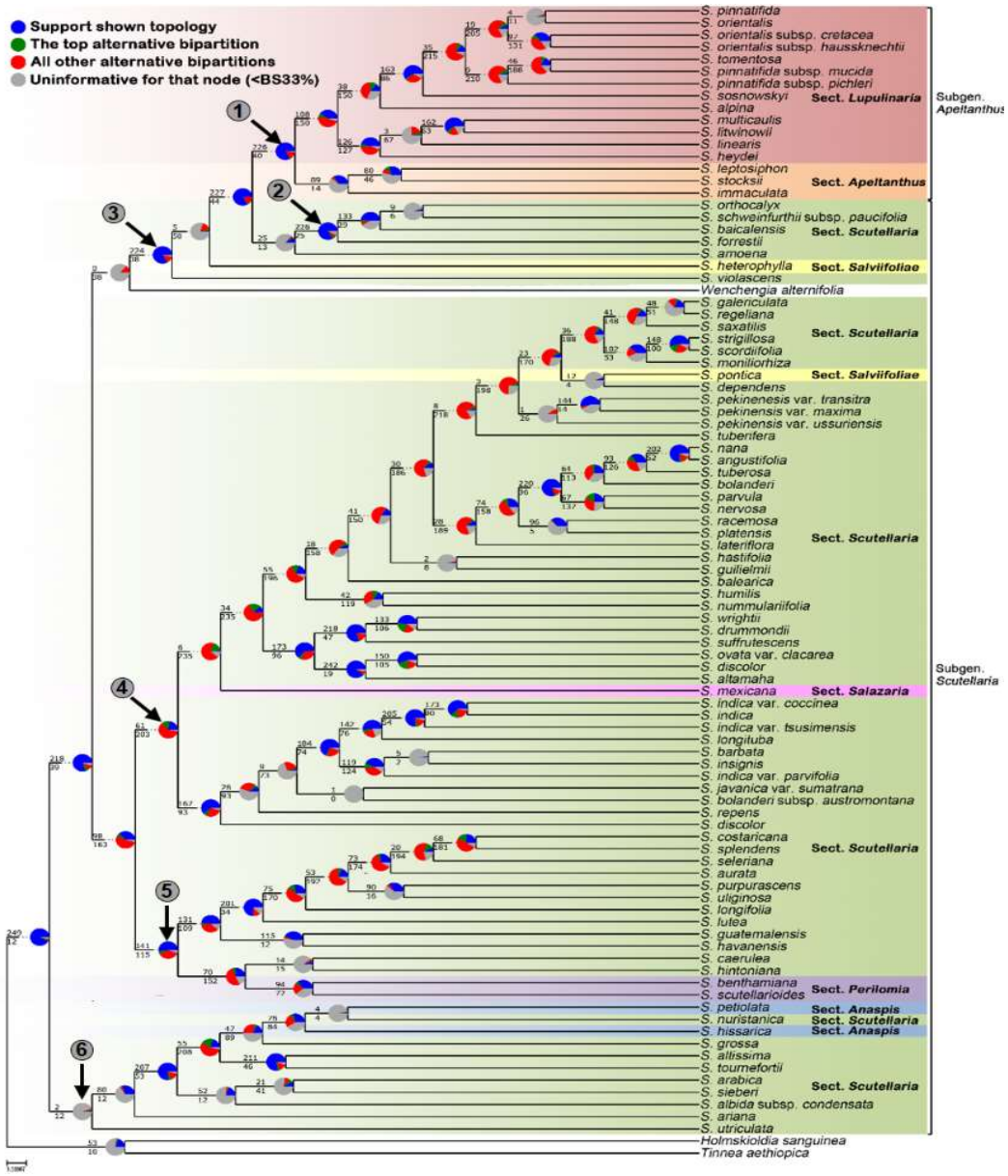


Fig. 26. Piechart on nodes showing conflict and congruent between individual gene trees and the coalescent ASTRAL tree based on PhyParts analysis. The proportion of four categories (blue, green, red, and gray) shows the degree of discordance of each node. Arrows (circled numbers) are the same clades as recognized in the concatenated ML tree.

IV. Discussion

1. Discordance between the concatenated ML tree and the coalescent ASTRAL tree

The tanglegram shows an easy-to-understand relationship between taxa of concatenated ML tree and coalescent ASTRAL tree (Fig. 25). In the coalescent ASTRAL tree (Fig. 24), the LPP values were lower than that of the concatenated ML tree (Fig. 21). In addition, through ASTRAL analysis for individual gene trees, discordance between the concatenated ML tree and the coalescent ASTRAL tree was confirmed. In the piechart obtained using PhyParts, it was also confirmed that some nodes showed discordance between individual gene trees and the coalescent ASTRAL tree through more red proportions. According to the previous studies (Herrando-Moraira et al., 2019; Long and Kubatko, 2018; Solis-Lemus et al., 2016; Morales-Briones et al., 2017; Villaverde et al., 2020), low LPP values and discordance were derived from 1) insufficient information for individual gene trees to resolve branches due to short length of loci, missing data by low capture efficiency or 2) low variable loci conflicts from evolutionary signals such as gene flow between ancestral lineage or ancient hybridization. The obtained tanglegram and piechart in this chapter indicated that the discordance clades with low LPP values of the coalescent ASTRAL tree, so it is considered that the cause is missing data due to low capture efficiency, not evolutionary signals

such as ancient hybridization. Brewer et al. (2019) suggested that genomic DNA quality is important for target enrichment efficiency, and mean exon and intron coverage are affected by sample age. Among the samples used in this chapter, most of the old specimens were collected from herbaria (Appendix 12). Some of the degraded DNA samples were also included (Appendix 13), which is considered to have affected the target enrichment efficiency. Therefore, for the analysis of conflict evolutionary signals based on the coalescent ASTRAL tree, additional sampling and sequencing are required to supplement the missing data in future studies.

2. Reexamination of phylogenetic relationships based on the concatenated ML tree

The phylogenomic analyses based on nuclear single-copy genes were first performed in *Scutellaria* using the Hyb-Seq technique. The probes for 641 loci were designed for 96 taxa including outgroup taxa, and the average capture efficiency per sample was 65.54 %, which was slightly low. Due to the low capture efficiency, the coalescence ASTRAL tree was inconsistent with the individual gene tree in most of the nodes (Fig. 24), and the LPP values were indicated also low. Therefore, the phylogenetic relationships of *Scutellaria* were inferred focused on the concatenated ML tree with high resolution.

As a result of phylogenetic analysis using a concatenated supercontig matrix for 272 loci excluding paralogous loci, BS values of most nodes in the concatenated ML tree were 90 % or more, and only 15 nodes had a BS value of 90 % or less (Fig. 21). It is considered that this was affected by missing data, however, the resolution of the concatenated ML tree was improved compared to the previous phylogenetic trees in Chapters 2 and 3. In addition, most of the taxa in which the phylogenetic position had been ambiguous were resolved. Therefore, it was confirmed that the total resolution of the concatenated ML tree was not significantly affected. The phylogenetic relationship, which was suggested in previous chapters between the monophyly of subgen. *Apeltanthus* and the paraphyly of subgen. *Scutellaria,* was clearly confirmed. Sect. *Salviifoliae* including two species (*S. heterophylla* and *S. pontica*), which has been considered a monophyletic group, is divided into different clades. Further phylogenetic analysis is needed adding more taxa included in sect. *Salviifoliae.* Sect. *Salazaria* has remained a polytomy with a low BS value of 50 or less in the trees of the previous chapters. In contrast, in this concatenated ML tree, it was located as a sister group of a clade containing some species of sect. *Scutellaria.* Sect. *Perilomia* was a sister clade to a clade including sp-grs. HINTONIANA and CAERULEA identically to the previous trees, and sect. *Anaspis* was also identically located at the base of the tree. In this chapter, although it is difficult to confirm the clear taxonomic delimitation

due to the lack of sampling compared to the previous chapters, it was possible to reexamine the phylogenetic relationship between subgeneric and sectional level.

To confirm the phylogenetic fidelity for the concatenated ML tree, the number of the shared quartets were calculated based on each gene tree for 272 loci and the concatenated ML tree. From the obtained statistics, it was confirmed that eight loci could cover more than 80% of the concatenated ML tree (Fig. 22). As a result of recalculating the shared quartet as a subset of the top eight loci, DN17712_c1_g5_i1 locus (Gene No. in *Arabidopsis*: AT3G54460) showed the largest shared quartet percentage (87.27 %; Fig. 23), and it was confirmed that this locus indicated the highest phylogenetic fidelity among 272 loci. Therefore, this locus will be the basis for selecting a proper DNA region in the phylogenetic analysis of *Scutellaria* to be performed later, and further study is needed to be analyzed including more loci.

CHAPTER 5.

Phylogenetic analyses based on nuclear ribosomal DNA sequences and the secondary structure of ITS2 in *Scutellaria*

I. Introduction

Nuclear ribosomal DNA (nrDNA) is tandemly repeated with consists of coding regions for 18S, 5.8S, and 26S rDNAs (Fig. 27; Cheon et al., 2018; Edger et al., 2014; Linder et al., 2000), and three spacer regions for internal transcribed spacers (ITS1 and ITS2), and external transcribed spacer (ETS). The intergenic spacer (IGS) region includes ETS and non-transcribed spacer (NTS). 5.8S, 18S, and 26S rDNA are relatively conserved regions. ITS1, ITS2, and ETS are highly diverse among species and exist multiple copies. ITS regions differ in length due to point mutations or indels, which is variable in length from 500-700 bp in angiosperms (Baldwin et al., 1995) and 1500-3700 bp in gymnosperms (Alvarez and Wendel, 2003). ITS1, ITS2, and ETS are informative and polymorphic regions, and it has been commonly used for phylogenetic studies (Chen et al., 2016; Safikhani et al., 2018; Zhao et al., 2017). Especially, the secondary structure of ITS2 was analyzed with phylogenetic

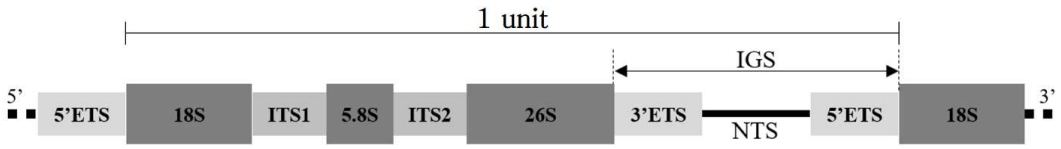


Fig. 27. General structure of nrDNA repeating unit in plants.

study and provided the key to distinguish taxa (Edger et al., 2014, Goertzen et al., 2003, Giudicelli, 2016, Zhang et al., 2015). The secondary structures are formed by the interaction of specific base pairs and are required for ribosome biogenesis and correct processing of rRNA (Hillis and Dixon, 1991; Nazar, 2004; Mullineux and Hausner, 2009). This structure is known to be well-conserved among taxa due to evolutionary constraints for own function (Freire et al., 2012). The conservation of secondary structures and specific nucleotides allows identifying taxa with sequences representing many polymorphisms in the sequences alignments (Goertzen et al., 2003). Recently, well-conserved common motifs in the secondary structure of ITS2 were confirmed in eukaryotes (Schultz et al., 2005). Schultz et al. (2005) suggested that all eukaryotes have shared characteristics within the ITS2 secondary structure (Fig. 28). These characteristics are 1) four helices with 2) helix III as the longest and 3) containing a UGGU motif 5' to the apex of the helix III (deviations like UGGGU, UGG, or GGU have been described) as well as 4) a U-U mismatch in the second helix. Therefore, the conservation of the secondary structure makes it possible to compare positional homology between sequences that are difficult to identify through sequence alignment (Goertzen et al., 2003). Morales-Briones and Tank (2019) and Zarrei et al. (2014) confirmed ITS2 secondary structure in Rosaceae and identified the same characteristics as those suggested by Schultz et al. (2005).

In this chapter, nrDNA sequence of *S. orientalis* was completed,

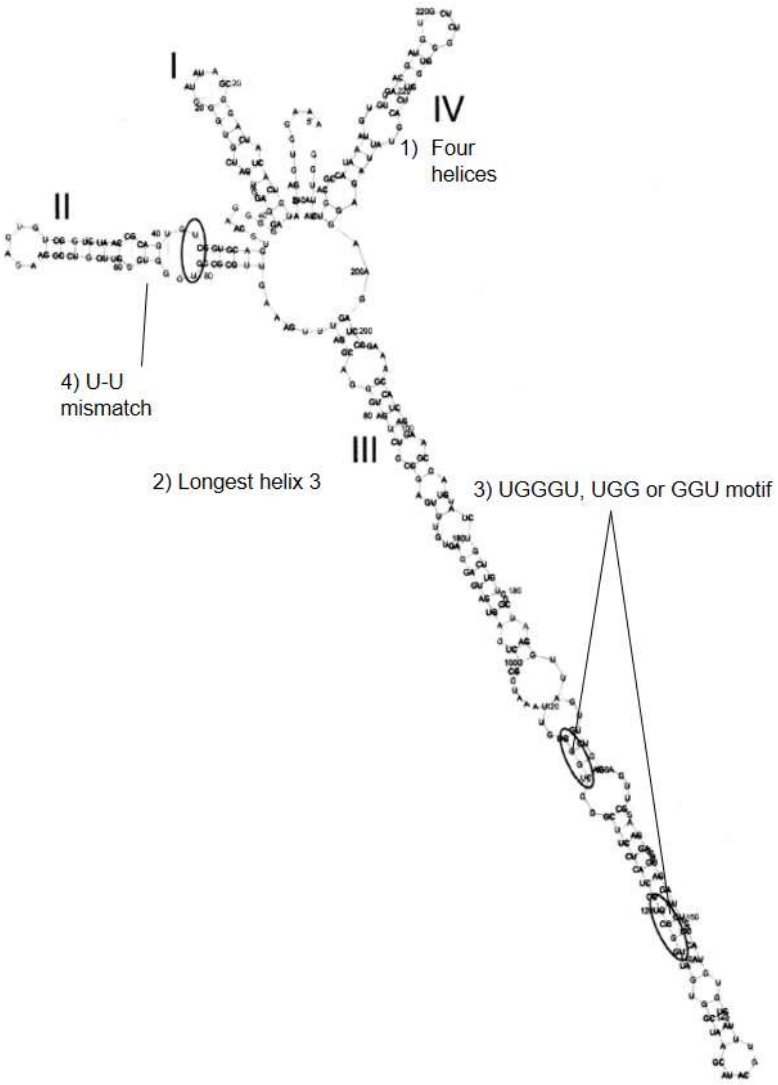


Fig. 28. A typical graphical visualization of the ITS2 secondary structure of eukaryotes (Schultz et al., 2005). 1)−4) shows the conserved features commonly found in eukaryotes (four helices indicate as I−IV).

and partial nrDNA sequences excluding NTS region of 50 taxa were obtained. Through a phylogenetic analysis using the nrDNA, the phylogenetic relationships were confirmed. Moreover, the various patterns of ITS2 secondary structure were compared for distinguishing the groups suggested by Paton (1990), and the shared characteristics of ITS2 secondary structures suggested by Schultz et al. (2005) were investigated in *Scutellaria.*

II. Materials and Methods

1. Taxon sampling and *de-novo* assembly

A total of 51 taxa were sampled, including *Holmskioldia sanguinea* and *Tinnea aethiopica* as outgroup taxa. According to the previous phylogenetic study of Lamiaceae, outgroup taxa were selected closely related to *Scutellaria* (Li et al., 2016). Compared to the current classification system (Paton, 1990), all subgenera (subgen.) and sections (sect.) were included, and 26 species groups (sp-gr.) were included (68.4 %). The information of samples is shown in Appendix 18. The *de-novo* assembly process was performed in the same method presented in Chapter 3.

2. Completion of nrDNA sequences by PCR and Sanger sequencing

Each filtered paired-end reads were mapped against a previously completed nrDNA sequence from *Perilla citriodora* (GenBank accession: KT220690) included in Lamiaceae (Cheon et al., 2018) using Geneious Prime (Kearse et al., 2012). In NTS region, the paired-end reads of all samples were not fully mapped due to high variability. To solve this problem in NTS, one of the samples (*S. orientalis*) was selected, and polymerase chain reaction (PCR) was performed in the unmapped region. The forward primer (5'-GATTCAGCCCTTTGTCGCTTC-3') was designed to 26S and the reverse primer (5'-CATACAAGCCAAACACCACTCAC-3') to 5' ETS, and the length of the PCR product was approximately 1,800 bp. The PCR was conducted with a total volume of 20 ㎕ containing 1 ㎕ template DNA, 10 ㎕ Master Taq (2X PCR Master mix Solution i-Taq, intron, Seoul, Korea), 1 ㎕ of each primer, and 7 ㎕ of distilled water. The PCR cycle performed using S1000 Thermal cycler (BioRad, Hercules, USA) with the following program: 3 min pre-denaturation at 94 ℃ followed by 35 cycles of 30 s denaturation at 95 ℃, 30 s annealing at 55 ℃ and 45 s extension at 72 ℃, and finished by a final extension of 7 min at 72 ℃. The PCR product was confirmed in 1.3 % agarose gel with 0.001 % ethidium bromide under the UV light of the Gel Doc XR+ System (BioRad, Hercules, USA). Sequencing was performed by a 3730xl DNA analyzer (Macrogen Inc., Korea), and sequences were

assembled and edited using Sequencher 4.9 (Gene Code Corporation, Ann Arbor, USA). The obtained sequence was re-mapped as reference with the paired-end reads of *S. orientalis* using Geneious. As a result, both sides of the consensus sequence were extended approximately 400 bp length, and the total length of the consensus sequence was approximately 1,500 bp. It was confirmed that this sequence overlapped with 26S region of the consensus sequence mapped with *P. citriodor*, but it did not completely cover the IGS region (extended sanger sequence). Subsequently, a partial ETS sequence was downloaded of *S. regeliana* Nakai from GenBank (MF193582.1) as a reference sequence, and mapped with paired-end reads of *S. orientalis*. After mapping, the consensus sequence and the extended sanger sequence were aligned, and it was confirmed that the partial sequence overlapped. Finally, *P. citriodor* sequence, *S. orientalis* sequence of mapped with *P. citriodor*, partial ETS sequence of *S. regeliana*, and partial ETS sequence of mapped with *S. regeliana* were aligned using MAFFT alignment in Geneious. As a result, it was confirmed that some of the sequences overlapped continuously, and the 18S, ITS1, 5.8S, ITS2, and 26S regions were annotated based on the *P. citriodor* sequence. Furthermore, the 5' ETS region was annotated based on partial ETS sequence of *S. regeliana*. The completed nrDNA sequence was additionally confirmed using RNAmmer v.1.2 (Lagesen et al., 2007). The rest of the taxa were mapped to complete nrDNA with the completed sequence of *S.*

orientalis as a reference.

3. Phylogenetic analyses

The methods of substitution model selection, sequence variation analysis, and phylogenetic analysis were performed in the same presented in Chapter 3. The GTR + I + Γ model was selected as a suitable model, and the bootstrap option was set in the same value as in Chapter 3. Phylogenetic tree was visualized in MEGA7 (Kumar et al., 2016).

4. Secondary structure analyses of ITS2

To analyze the secondary structure, each ITS2 sequence was extracted from nrDNA sequences. ITS2 sequences of 35 taxa were included, excluding the taxa containing ambiguous sequences. The secondary structure of the ITS2 was predicted by RNAfold web server (http://rna.tbi.univie.ac.at/) with default option, and the picture was plotted using PseudoViewer v.3.0 (Han et al., 2002). The specific characteristics in the secondary structure were investigated based on Schultz et al. (2005).

III. Results

1. Completion of nrDNA sequences and phylogeny based on nrDNA

In nrDNA, *S. orientalis* was successfully completed nrDNA sequence including NTS repeat region, and its length was 8,047 bp. However, the rest of the taxa were failed to assemble in NTS region of nrDNA sequences. The sequences of 48 taxa were obtained including partial 5' ETS, 18S, ITS1, 5.8S, ITS2, and 26S regions. The analyses of nrDNA sequences focused on these six regions, and the length of total sequences ranged from 6,240 bp to 6,358 bp (Table 13). The GC content of 49 taxa was 55.7 % to 57.5 %, and all sequences were uploaded in GenBank (Table 13).

The result of the sequence analyses for each region of nrDNA is summarized in Table 14. The length of the aligned sequences was 6,496 bp. The proportion of variable sites and informative characters in 18S, 5.8S, and 26S coding regions was highly conserved from 0.8 % to 8.9 %. Among the three spacer regions (ETS, ITS1, and ITS2), the proportion of variable sites was the highest in ITS1 region (54.2 %), and the region with the highest proportion of informative characters was ETS region (40.0 %). ITS2 region was indicated a slightly lower ratio of variable sites and informative characters than the rest of the two spacer regions.

In the phylogenetic tree (Fig. 29), some clades showed the lower bootstrap support (BS) value than previous trees (Chapters 2, 3,

and 4). Similar to previous trees, subgen. *Apeltanthus* was a monophyetic group and subgen. *Scutellaria* was a paraphyletic group. Also, sect. *Apeltanthus* of subgen. *Apeltanthus* was a clearly paraphyletic group, and sect. *Lupulinaria* was a monophyletic group. Sect. *Lupulinaria* was a sister group with some species of sect. *Apeltanthus*. Similar to chloroplast genome trees in Chapters 2 and 3, *S. forrestii* and *S. heterophylla* of subgen. *Scutellaria* appeared a close relationship with subgen. *Apeltanthus*. *S. dependen*s of sp-gr. GALERICULATA (BS = 67) was a sister to sp-gr. STIRIGILLOSA. Sp-grs. INCANA was a monophyetic group (BS =100) and closely related *S. ovata* of sp-grs. OVATA. *S. mexicana* of sect. *Salviifoliae* was a sister to some taxa of sect. *Scutellaria* (BS = 66). Sp-gr. VIOLACEA was indicated as a monophyletic group and highly supported (BS = 100). *S. scutellarioides* of Sect. *Perilomia* is close to *S. hintoniana*. Sp-grs. ALBIDA and *sect. Anaspis* were located at the base of the tree the same as previous trees in Chapters 3 and 4, and this clade was highly supported (BS = 100). The rest of the species groups of the sect. *Scutellaria* were also mixed-up.

Table 13. Statiscical summary of nrDNA sequences of 49 *Scutellaria* taxa and outgroup taxa

Taxa	Length (bp)							Total GC Contents (%)	GenBank accession No.
	Total	5' ETS	18S	ITS1	5.8S	ITS2	26S		
H. sanguinea (outgroup)	6,278	452	1,808	229	164	234	3,391	56.2	MW396590
S. albida subsp. *condensata*	6,282	453	1,808	246	164	219	3,392	57.5	MW396591
S. altamaha	6,260	443	1,808	239	164	214	3,392	55.7	MW396592
S. angustifolia	6,261	443	1,808	238	164	216	3,392	57.2	MW396593
S. barbata	6,272	442	1,808	202	164	245	3,392	56.9	MW396595
S. bolanderi	6,259	441	1,808	239	164	215	3,392	57.3	MW396596
S. costaricana	6,291	444	1,808	243	164	240	3,392	57	MW396597
S. dependens	6,267	443	1,808	238	164	222	3,392	57	MW396598
S. discolor	6,240	436	1,809	201	164	238	3,392	56.7	MW396599
S. drummondii	6,265	441	1,808	239	164	222	3,391	56.9	MW396600
S. forrestii	6,285	443	1,808	242	164	235	3,393	56.8	MW396601
S. franchetiana	6,246	442	1,809	202	164	237	3,392	56.8	MW396602
S. gaumeri	6,288	442	1,808	242	164	240	3,392	56.9	MW396603
S. guatemalensis	6,288	442	1,808	242	164	240	3,392	56.9	MW396604
S. hastifolia	6,268	442	1,808	239	164	223	3,392	56.6	MW396605
S. heterophylla	6,277	444	1,808	240	164	231	3,390	56.4	MW396606
S. heydei	6,275	443	1,808	242	164	226	3,392	57.1	MW396607
S. hintoniana	6,278	442	1,808	243	164	230	3,391	57.4	MW396608
S. hissarica	6,278	453	1,808	244	164	217	3,392	57.1	MW396609
S. immaculata	6,275	445	1,808	241	164	225	3,392	57.1	MW396610
S. incana	6,256	443	1,808	237	164	214	3,390	55.7	MW396611
S. indica	6,251	446	1,808	202	164	239	3,392	56.4	MW396612
S. indica var. *coccinea*	6,251	446	1,808	202	164	239	3,392	56.5	MW396613

S. insignis	6,250	441	1,808	202	164	244	3,391	56.5	MW396614
S. lateriflora	6,268	444	1,808	238	164	223	3,391	56.8	MW396615
S. leptosiphon	6,274	444	1,808	241	164	225	3,392	57.1	MW396616
S. linearis	6,283	444	1,808	241	164	234	3,392	56.9	MW396617
S. longifolia	6,288	443	1,808	243	164	239	3,391	57.4	MW396618
S. mexicana	6,259	443	1,808	238	164	214	3,392	56.8	MW396619
S. mollis	6,275	447	1,808	240	164	224	3,392	57.1	MW396620
S. moniliorhiza	6,270	444	1,808	239	164	223	3,392	57.1	MW396621
S. multicaulis	6,277	445	1,808	241	164	227	3,392	56.9	MW396622
S. nana	6,259	442	1,808	238	164	215	3,392	57.2	MW396623
S. nummulariifolia	6,277	448	1,808	240	164	225	3,392	57.2	MW396624
S. orientalis	8,047	443	1,803	243	164	224	3,392	56.1	MW396625
S. ovata	6,257	443	1,808	237	164	214	3,391	55.7	MW396626
S. parvula	6,261	443	1,808	238	164	216	3,392	57.3	MW396627
S. pekinensis var. *maxima*	6,270	444	1,808	239	164	223	3,392	56.7	MW396628
S. pinnatifida subsp. *mucida*	6,272	443	1,808	241	164	224	3,392	57	MW396629
S. pinnatifida subsp. *pichleri*	6,272	443	1,808	241	164	224	3,392	56.9	MW396630
S. platensis	6,257	443	1,808	238	164	214	3,390	57.2	MW396631
S. purpurascens	6,289	443	1,808	243	164	239	3,392	57.4	MW396632
S. rubicunda	6,283	453	1,808	246	164	220	3,392	57.5	MW396633
S. scutellarioides	6,358	442	1,808	243	164	237	3,464	57.4	MW396634
S. seleriana	6,292	443	1,808	243	164	242	3,392	57.4	MW396635
S. siphocampyloides	6,261	442	1,811	237	164	215	3,392	57.2	MW396636
S. splendens	6,282	444	1,808	242	164	232	3,392	57.2	MW396637
S. strigillosa	6,250	444	1,808	239	164	225	3,392	57.1	MW396638
S. suffrutescens	6,265	441	1,808	239	164	222	3,391	56.9	MW396639
S. tomentosa	6,273	444	1,808	241	164	224	3,392	56.9	MW396594
T. aethiopica (outgroup)	6,270	450	1,808	236	164	220	3,392	56	MW396640

Table 14. Statistical summary of nrDNA sequences in *Scutellaria* excluding heterogeneous sequences (35 taxa)

	ETS	18S	ITS1	5.8S	ITS2	26S	Total
Aligned length (bp)	478	1,840	273	164	275	3,466	6,496
No. of Variable sites (bp)	257	163	148	7	99	137	811
Proportion of Variable sites (%)	53.8	8.9	54.2	4.3	36.0	4.0	12.5
No. of Informative characters (bp)	191	15	89	5	76	83	459
Proportion of Informative characters (%)	40.0	0.8	32.6	3.0	27.6	2.4	7.1

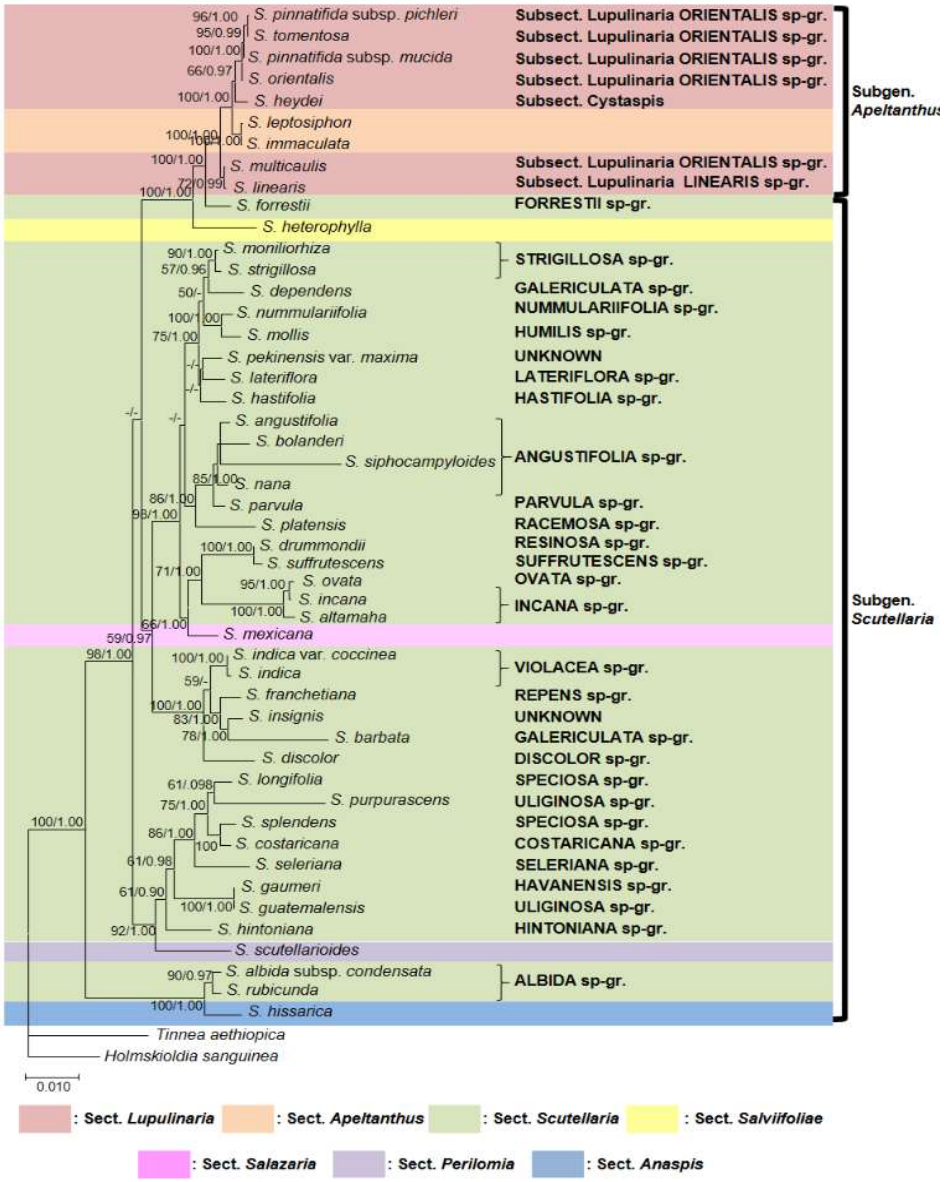


Fig. 29. Phylogenetic tree of *Scutellaria* based on nrDNA using RAxML (1,000 bootstrap replication). Numbers above the branches are bootstrap values. Numbers only BS ≥ 50% and PP ≥ 0.90 are indicated (BS/PP). Colored boxes are sections and capitalized names after the taxon names are "species groups" recognized in Paton (1990). Q" indicates a taxon of which taxonomic position (sections and "species group") have not been reported.

2. Analyses of ITS2 secondary structures

Thirty-five taxa included in the secondary structure analyses showed various patterns (Figs. 30 and 31; Table 15). Seven types (types I-VII) of secondary structure patterns were distinguished in *Scutellaria*, and the five taxa had individually different patterns (*S. purpurascens*, *S. splendens*, *S. heydei*, *S. scutellarioides*, and *S. platensis*). Each pattern was displayed on the phylogenetic tree in a different color (Fig. 30), but the patterns did not have the characteristics to distinguish subgroups of *Scutellaria.* Each pattern was found to be mixed in the tree.

As a result of investigating the secondary structure to compare the four characteristics suggested by Schultz et al. (2005; Fig. 31; Table 15; Appendix 19) in *Scutellaria,* taxa had three to six helices in secondary structures. Sixteen taxa with five helices were most common, followed by 14 taxa with four helices, four taxa with three helices, and one taxon with six helices. The helix III was generally indicated the longest, however, type V had the longest first helix, and type VI and *S. heydei* had the longest second helix. The helix III included UGGU motif at the 5' apex, but some taxa (type V and *S. scutellarioides*) was not indicated this motif. In addition, U-U mismatch at helix II was confirmed in only four taxa (*S. altamaha*, *S. heterophylla, S. heydei,* and *S. ovata*).

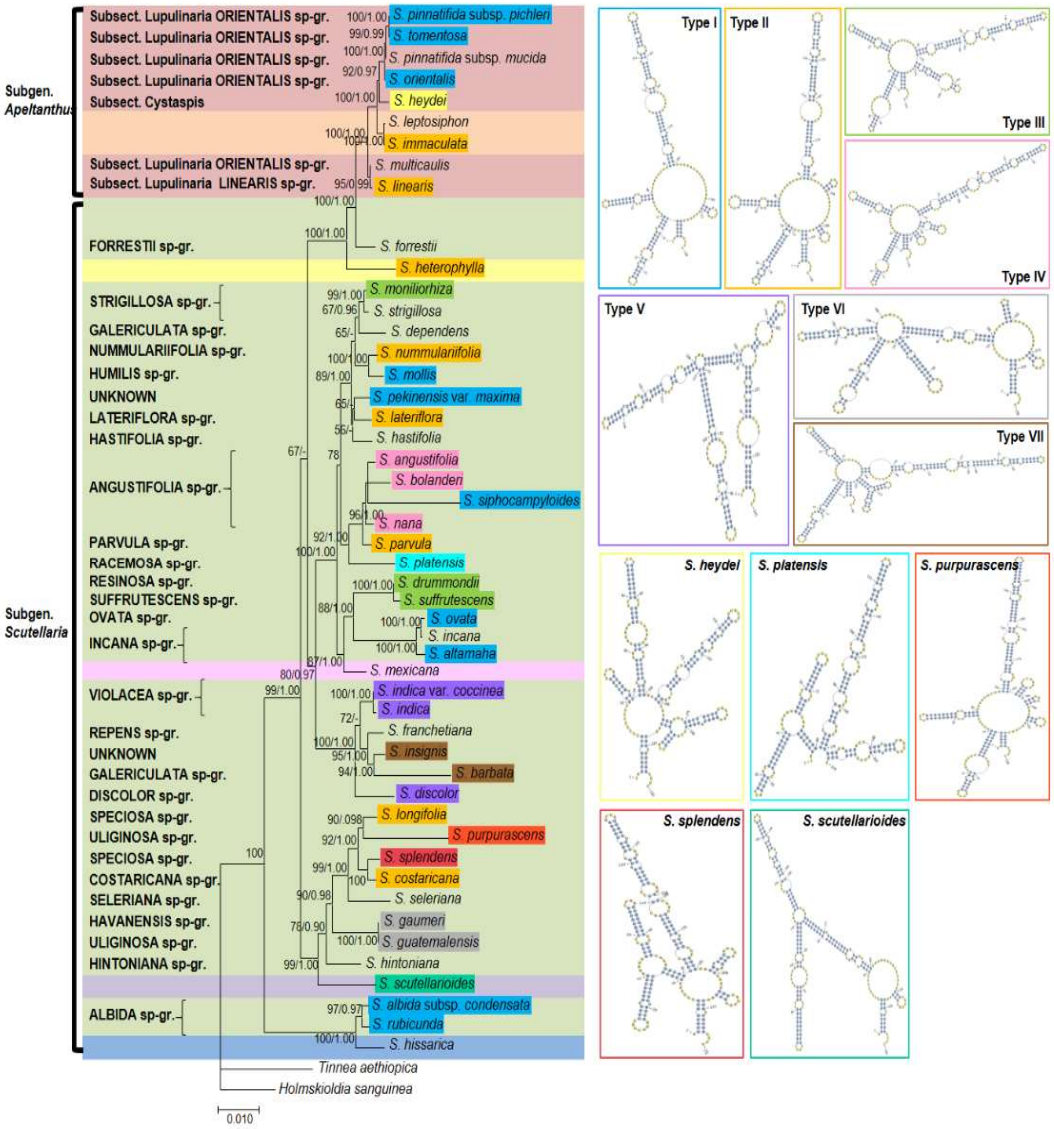


Fig. 30. Distribution of types of ITS2 secondary structure in the phylogenetic tree. Each type matches with different background color in the taxon name. No background color indicates taxon having heterogeneous sequences in nrDNA.

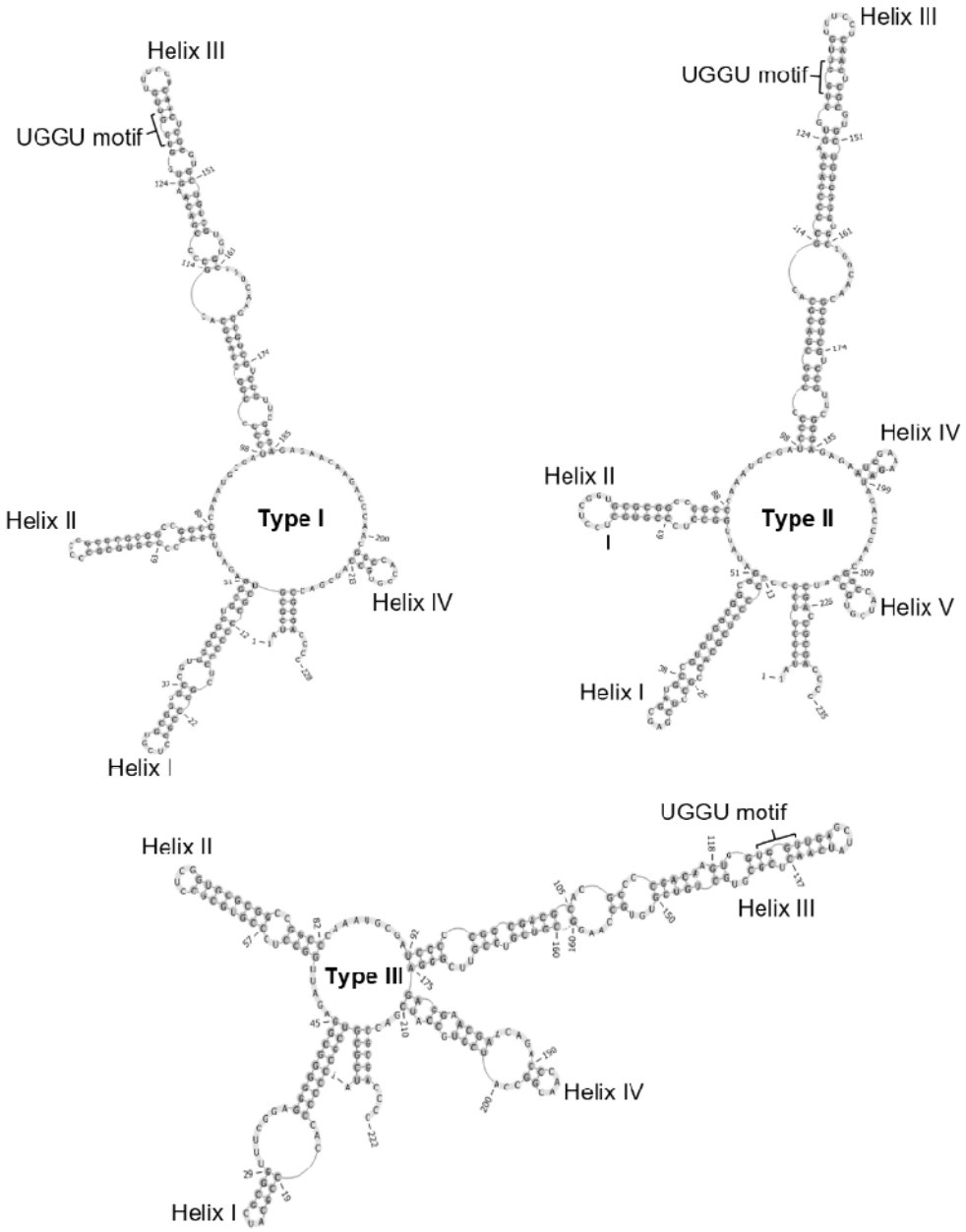


Fig. 31. Twelve ITS2 structural types of *Scutellaria*. Helix numbers (I–VI) and UGGU motif are followed the suggestions by Schultz et al. (2005).

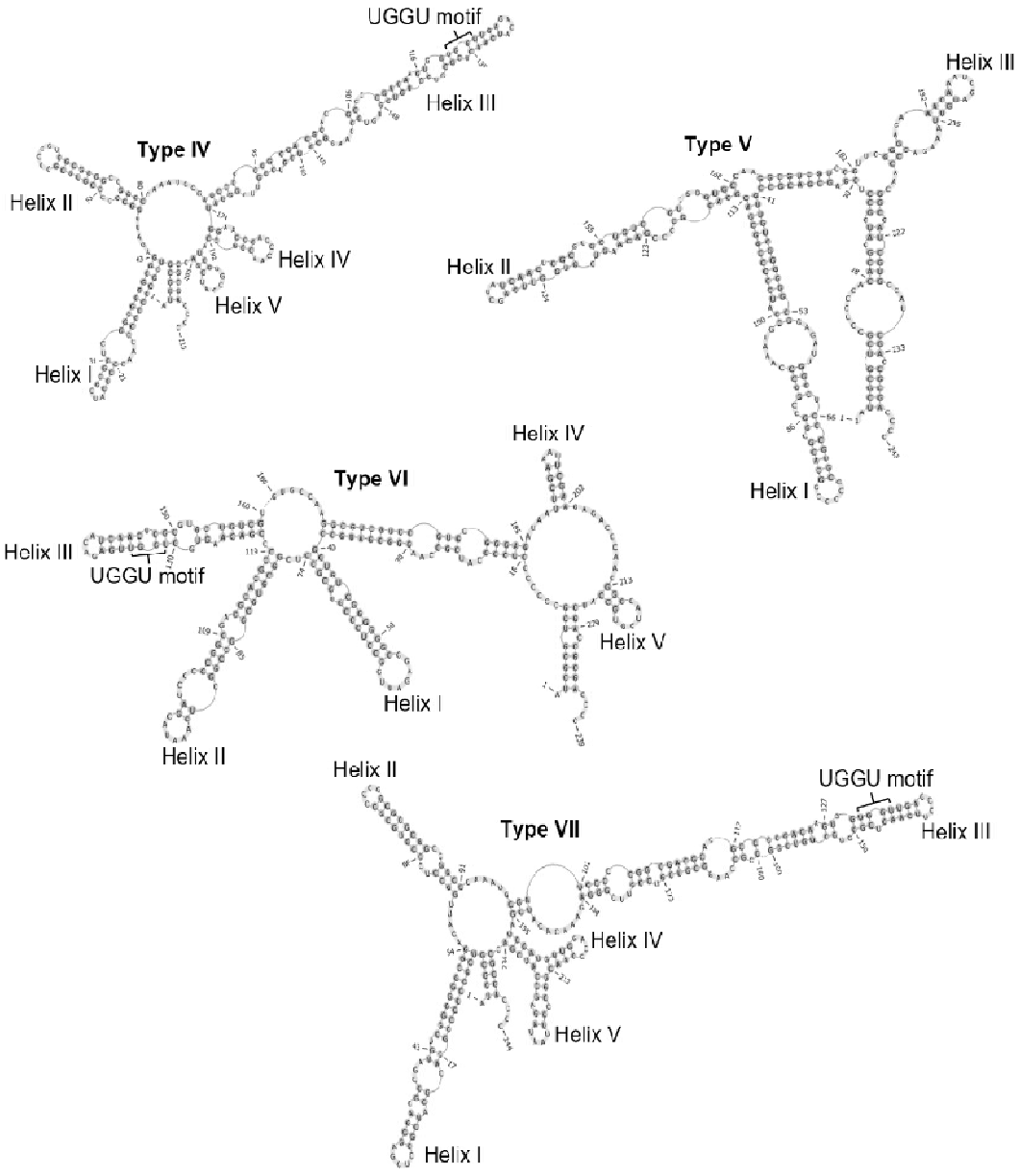


Fig. 31. Continued.

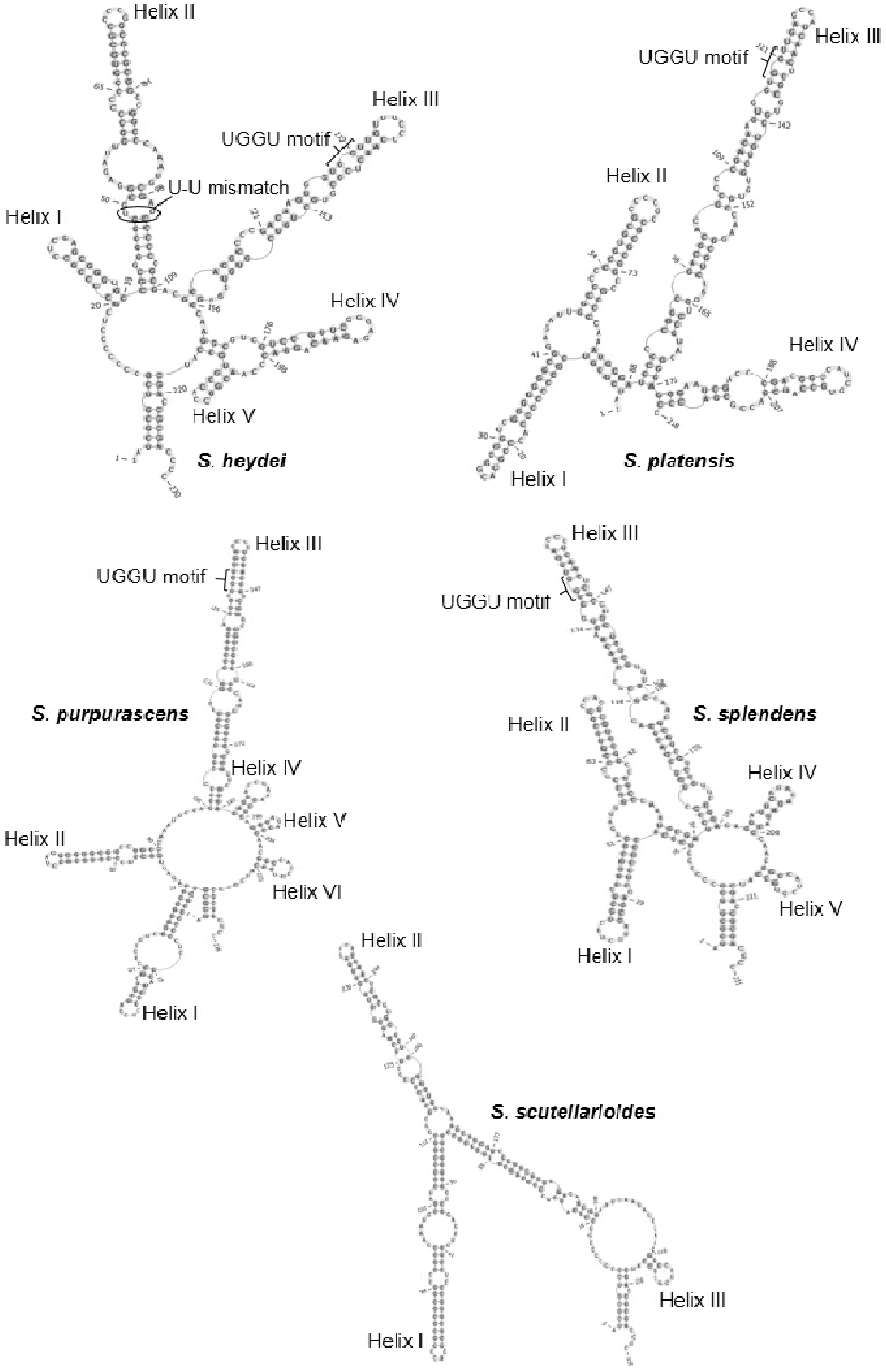


Fig. 31. Continued.

Table 15. Comparison between groups in the current classification system (Paton, 1990) and the characteristics of the ITS2 secondary structure of *Scutellaria*

secondary structure type	No. of helices	UGGUmotif at helix III	U-U mismatch at helix II	Taxa	Subgenus	Section	Subsection	Species-group
I	4	o	o	*S. orientalis*	*Apeltanthus*	*Lupulinaria*	*Lupulinaria*	ORIENTALIS
I	4	o	x	*S. pinnatifida* subsp. *pichleri*	*Apeltanthus*	*Lupulinaria*	*Lupulinaria*	ORIENTALIS
I	4	o	x	*S. tomentosa*	*Apeltanthus*	*Lupulinaria*	*Lupulinaria*	ORIENTALIS
I	4	o	x	*S. albida* subsp. *condensata*	*Scutellaria*	*Scutellaria*		ALBIDA
I	4	o	x	*S. rubicunda*	*Scutellaria*	*Scutellaria*		ALBIDA
I	4	o	x	*S. siphocampyloides*	*Scutellaria*	*Scutellaria*		ANGUSTIFOLIA
I	4	o	x	*S. mollis*	*Scutellaria*	*Scutellaria*		HUMILIS
I	4	o	o	*S. ovata*	*Scutellaria*	*Scutellaria*		OVATA
I	4	o	o	*S. altamaha*	*Scutellaria*			
I	4	o	x	*S. pekinensis* var. *maxima*	*Scutellaria*			
II	5	o	x	*S. immaculata*	*Apeltanthus*	*Apeltanthus*		
II	5	o	x	*S. heterophylla*	*Scutellaria*	*Salviifoliae*		
II	5	o	x	*S. costaricana*	*Scutellaria*	*Scutellaria*		COSTARICANA
II	5	o	x	*S. lateriflora*	*Scutellaria*	*Scutellaria*		LATERIFLORA
II	5	o	x	*S. nummulariifolia*	*Scutellaria*	*Scutellaria*		NUMMULARIIFOLIA
II	5	o	x	*S. parvula*	*Scutellaria*	*Scutellaria*		PARVULA
II	5	o	x	*S. longifolia*	*Scutellaria*	*Scutellaria*		SPECIOSA
III	4	o	x	*S. drummondii*	*Scutellaria*	*Scutellaria*		RESINOSA
III	4	o	x	*S. moniliorhiza*	*Scutellaria*	*Scutellaria*		STRIGILLOSA
III	4	o	x	*S. suffrutescens*	*Scutellaria*	*Scutellaria*		SUFFRUTESCENS
IV	5	x	x	*S. angustifolia*	*Scutellaria*	*Scutellaria*		ANGUSTIFOLIA
IV	5	x	x	*S. bolanderi*	*Scutellaria*	*Scutellaria*		ANGUSTIFOLIA
IV	5	x	x	*S. nana*	*Scutellaria*	*Scutellaria*		ANGUSTIFOLIA
V	3	o	x	*S. discolor*	*Scutellaria*	*Scutellaria*		DISCOLOR
V	3	o	x	*S. indica*	*Scutellaria*	*Scutellaria*		VIOLACEA
V	3	o	x	*S. indica* var. *coccinea*	*Scutellaria*	*Scutellaria*		VIOLACEA
VI	5	o	x	*S. gaumeri*	*Scutellaria*	*Scutellaria*		HAVANENSIS
VI	5	o	x	*S. guatemalensis*	*Scutellaria*	*Scutellaria*		ULIGINOSA
VII	5	o	x	*S. barbata*	*Scutellaria*	*Scutellaria*		GALERICULATA
VII	5	o	x	*S. insignis*	*Scutellaria*			
	5	o	o	*S. heydei*	*Apeltanthus*	*Lupulinaria*	*Cystaspis*	
	3	x	x	*S. scutellarioides*	*Scutellaria*	*Perilomia*		
	4	o	x	*S. platensis*	*Scutellaria*	*Scutellaria*		RACEMOSA
	5	o	x	*S. splendens*	*Scutellaria*	*Scutellaria*		SPECIOSA
	6	o	x	*S. purpurascens*	*Scutellaria*	*Scutellaria*		ULIGINOSA

IV. Discussion

The structure of nrDNA sequences among *Scutellaria* taxa was confirmed highly conserved. Among the three coding regions, it was founded that ETS and ITS1 regions contained relatively high variable sites and informative characters, but ITS2 region appeared slightly lower.

The phylogenetic analysis showed a similar relationship to Chapters 3 and 4; 1) *S. heterophylla* was close to subgen. *Apeltanthus,* 2) sect. *Perilomia* was close to *S. hintoniana,* 3) sp-gr. ALBIDA of sect. *Scutellaria* and sect. *Anaspis* were located at the base of the phylogenetic tree, and 4) most species groups of sect. *Scutellaria* are located at different clades. Conspicuously, sp-gr. STRIGILLOSA was a monophyletic group in nrDNA tree, unlike the other trees in previous chapters. However, other taxa of sp-gr. STRIGILLOSA were not sufficiently included in this tree compared to the previous trees. In addition, *S. immaculata* and *S. leptosiphon* of sect. *Apeltanthus* showed as monophyletic clade and highly supported (BS=100). This relationship was also found in phylogenetic trees using single-copy nuclear genes in Chapter 4. This chapter will help establishing their phylogenetic relationships by comparing chloroplast and nuclear trees.

The secondary structure of ITS2 showed variable patterns (Figs 30 and 31; Table 15). As a result of applying the secondary structure pattern in nrDNA tree, type of ITS2 secondary structure was not correspond to phylogenetic clades in *Scutellaria*. The

number of helices in the secondary structure of the taxa was formed three to six, and this number was also considered to not correlate with the phylogenetic relationships. Schultz et al. (2005) suggested that ITS2 secondary structure of Eukaryota has four shared characteristics, and recently Morales-Briones and Tank (2018) and Zarrei et al. (2014) confirmed this in Rosaceae. However, these shared structures in eukaryotes were not found in *Scutellaria* (Table 15). One shared characteristic that UGGU motif was located at 5' to the apex of helix III was found in most taxa of *Scutellaria,* but four taxa (Type V and *S. scutellarioides*) included this motif in helix II. Also, the third helix was the longest in most structures, but some taxa were not. U-U mismatch also confirmed only four taxa, All the characteristics suggested by Schultz et al. (2005) were not commonly found in *Scutellaria,* and were not related to the phylogenetic relationship. Moreover, these helices patterns were not related to sequences alignment or the phylogenetic relationship of taxa. Therefore, it is considered that the four shared characteristics suggested by Schultz et al. (2005) need to be also confirmed not only for additional taxa of *Scutellaria* but also for other genera of Lamiaceae or Lamiales.

CHAPTER 6.

Conclusion

I. Comparison between chloroplast and nuclear phylogenetic trees

The discordance between phylogenetic trees of nuclear and chloroplast has been known to derive from different evolutionary characteristics. First, chloroplast genome is inherited maternally, and recombination rarely occurs and is conserved (Twyford and Ness, 2017). In contrast, the nuclear genome is generally biparentally inherited with recombination (Badro et al., 2019; Lin et al., 2019; Rieseberg and Soltis, 1991; Neiman and Taylor, 2009). Therefore, a comparison of chloroplast and nuclear phylogeny possibly detect signals for evolutionary history such as ancient hybridization (Bruun-Lund et al., 2017; Morales-Briones et al., 2018; Vargas et al., 2019).

To compare the phylogenetic tree based on the nuclear dataset obtained in Chapter 4 and chloroplast phylogenetic tree obtained in Chapter 3, a tanglegram was drawn using R package (R Core Team, 2019) dendextend (Galili, 2015) noted in Chapter 4 (Fig. 32). In the tanglegram, four major clades were recognized, and these clades were completely different from major groups in the current classification system (Paton, 1990). These four major clades are potential subgenera in the new classification system later. The relationships among the taxa in both phylogenetic trees were shown as highly diverged lineages. In clades #1 and #3, most nodes were highly

supported [bootstrap support (BS) value ≥ 90] in both trees but conflict each other. It is considered that hybridization origins by diversification of many taxa occurred. In clade #2, most taxa were indicated conflict with each other. However, since most nodes in the chloroplast tree were weakly supported, the diversification process could not be inferred. A clade containing section (sect.) *Perilomia, S. hintoniana,* and *S. caerulea* also showed concordance relationships in both trees. These clades were relatively strongly supported (BS ≥ 98), appearing in the same location in both trees. In addition, a clade containing sect. *Anaspis* and species group (sp−gr.) ALBIDA located at the base of both trees showed almost concordance relationships. Therefore, it is considered that these clades were unaffected by hybridization origin by diversification within *Scutellaria.*

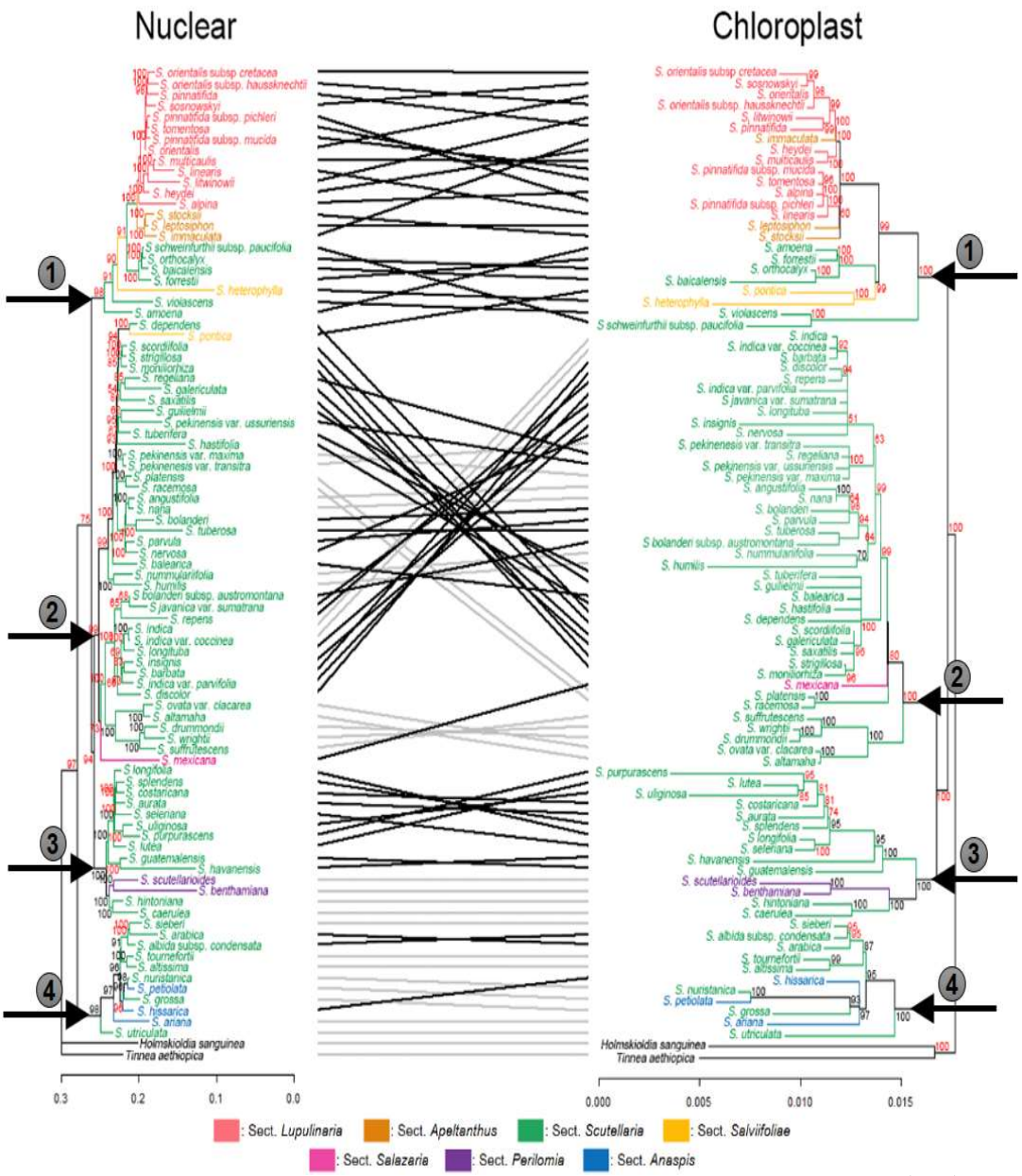


Fig. 32. A tanglegram showing phylogenetic trees based on nuclear (left) and chloroplast (right) datasets. The numbers above each branch in both trees are BS values. The black connecting lines indicate the concordant. The BS in red also indicate the discordance between the two trees. Arrows (circled numbers) indicate conclusive four major clades in *Scutellaria* recognized in this study, and these four major clades are identical in both trees although their relationships are different.

II. Comprehensive review of phylogenetic relationships in *Scutellaria*

Compare to the previous molecular phylogenetic studies (Safikhani et al., 2018; Zhao et al., 2017; Zhao et al., 2020), the monophyly of subgenus (subgen.) *Apentanthus* and paraphyly of subgen. *Scutellaria* is commonly showed the previous studies by Zhao et al. (2017; 2020) and all chapters in this study. According to Safikhani et al. (2018), both subgenera were considered monophyletic groups in ITS tree. However, in *trnL−F* tree, subgen. *Scutellaria* was paraphyletic group. Especially, *S. galericulata* was located in an independent clade, and it was suggested that it is possible to be classified in a new section. However, only eight taxa of subgen. *Scutellaria* were included in the phylogenetic tree of Safikhani et al. (2018). All phylogenetic trees of all chapters in this study included more taxa than the previous studies, and sp−gr. GALERICULATA was divided into different clades in sect. *Scutellaria* with high support values. Based on these results, it is strongly demonstrated that subgen. *Scutellaria* is a paraphyletic group and *S. glaericulata* is not possible to be suggested as an independent section. Therefore, it needs to reexamine the taxonomic level of these subgenera and to clarify their classification with morphological characters.

In this study, the phylogenetic studies of *Scutellaria* were performed based on 1) three regions (*ccsA*, *matK* and *ndhF*) of chloroplast (Chapter 2, Fig. 5), 2) combined with three regions of

chloroplast and complete chloroplast genome (Chapter 3, Fig. 15), 3) nuclear single-copy loci (Chapter 4, Fig. 21), and 4) four regions of the nuclear ribosomal DNA (Chapter 5, Fig. 29). As a result, a total of four phylogenetic trees were obtained. The relationships among taxa in *Scutellaria* were possible to summarized common in the four trees by applying the classification system suggested by Paton (1990). First, subgen. *Apeltanthus* was a monophyletic group, and subgen. *Scutellaria* was a paraphyletic group. Second, in subgen. *Apeltanthus,* sects. *Apeltanthus* and *Lupulinaira* appeared paraphyletic groups in chloroplast trees. However, in the tree based on nuclear single-copy loci, both sections were monophyletic groups. Similarly, in nrDNA tree, sect. *Apeltanthus* was a monophyletic group, and sect. *Lupulinaria* was paraphyletic group. However, since the trees based on the nuclear single-copy loci and the nrDNA region do not include sufficient taxa compared to chloroplast trees, the relationships between the two sections are still considered unresolved. Third, the sister clade of subgen. *Apeltanthus* included *S. schweinfurthii* subsp. *paucifolia, S. orthocalyx, S. baicalensis*, and *S. forrestii*. This phylogenetic position appeared similarly in four trees, and it is considered that a clade containing these taxa could be suggested as an independent clade and could be suggested a new classification system in further study in further study. Fourth, Sect. *Salviifoliae* appeared as a sister clade of subgen. *Apeltanthus* in super-matrix tree, but the location is still unresolved because of a different relationship in the

nuclear trees. It has been clearly confirmed by all trees that *S. heterophylla* is a sister group with subgen. *Apeltanthus*. Fifth, in sect. *Scutellaria,* it is difficult to compare the phylogenetic relationship between species groups due to the difference in the number of taxa included in each tree, but most of the species groups were mixed-up. Sixth, sect. *Salazaria,* which is a monotypic section, was sister to a clade consisting of some taxa in sect. *Scutellaria.* According to previous morphological studies (Torrey, 1858; Barkely, 1975), sect. *Salazaria* had remained unclear, but it was included in subgen. *Scutellaria* based on morphological reexamination by Paton (1990). Therefore, the molecular phylogenetic relationship between sects. *Scutellaria* and *Salazaria* support the current classification system (Paton, 1990). In addition, sect. *Salazaria* was suggested that it had a habitat and corolla form similar to sp-gr. RESINOSA of sect. *Scutellaria* by Paton (1990), but this relationship between both groups is not shown in all trees. Fifth, sect. *Perilomia* indicated as a monophyletic group in all trees and was shown to be closely related to *S. hintoniana* of sp-gr. HINTONIANA and *S. caerulea* of sp-gr. CAERULEA. According to Paton (1990), sect. *Perilomia* was evolved from species groups including *S. caerulea*, *S. lutea, S. hintoniana,* and *S. nummularifolia.* In phylogenetic trees for all chapters, sect. *Perilomia* was showed a close relationship with these species. However, *S. nummularifolia* was included in another clade. Therefore, sect. *Perilomia* and a clade including *S. hintoniana* and *S. caerulea,* could be considered as

an independent clade and could be suggested a new classification system in further study. Lastly, sp-gr. ALBIDA and sect. *Anaspis* were commonly located at the base of all trees, and *S. utriculata* was located as the most basal taxa of a clade including sp-gr. AlBIDA and sect. *Anaspis.* Similar to these relationships, Paton (1990) suggested that the pericarp characteristics of sp-gr. ALBIDA were similar to sect. *Anaspis,* and sp-gr. UTRICULATA also noted that the nutlet morphology was similar to sp-gr. ALBIDA. Therefore, a clade including sect. *Anaspis,* sp-grs. ALBIDA, and UTRICULATA could also be considered as an independent clade and a new phylogenetic classification system.

In this study, the molecular phylogenetic studies were first performed based on chloroplast and nuclear regions of *Scutellaria* in the world, which had not been performed until now. Also, this study included the largest amount of data than the previous studies. In conclusion, although this study was not possible to compare clear phylogenetic relationships of *Scutellaria* with the classification system suggested by Paton (1990), various diversification and variation in chloroplast and nuclear regions in *Scutellaria* were extensively identified for the first time. Moreover, it was confirmed that highly supported common clades in both chloroplast and nuclear trees could be named in the concept of the phylogenetic classification.

III. Further studies

Further studies to be performed after this study of *Scutellaria* are as follows. First, additional sampling is required for species groups suggested in current classification system (Paton, 1990) that has not been covered. Second, it is necessary to exclude the taxa that obtained insufficient data due to low genomic DNA quality among the obtained samples and obtain genomic DNA with improved quality through additional sampling and DNA extraction. Third, it is necessary to complete more chloroplast genomes and obtain more nuclear single-copy data by performing additional Hyb-Seq for taxa with improved quality genomic DNA. Through this large amount of data, it is necessary to obtain chloroplast and nuclear trees with high resolution and more taxa. By comparing the both chloroplast and nuclear trees, it is possible to investigate the origin of taxa by evolutionary events such as ancient hybridization. Evolutionary events for subgroups in *Scutellaria* can also be investigated by comparing the coalescent ASTRAL tree and the concatenated ML tree with high resolution. Moreover, through additional quartet analysis, it is possible to confirm whether DN17712_c1_g5_i1 loci (Gene No. in *Arabidopsis:* AT3G54460) loci previously presented is the most proper region for phylogenetic study and to suggest additional loci as proper DNA region. Fifth, the biogeographic analysis should be performed by measuring the divergence time of *Scutellaria* in the world, which has not been performed until now. Finally, the morphological studies of nutlets suggested as the

important keys that distinguish taxa of *Scutellaria* should be performed, and a comprehensive review is needed comparing with the phylogenetic tree. It is expected that the morphological characteristics of nutlets can be applied to the obtained phylogenetic tree, and a new phylogenetic classification could be suggested through comparison with the current classification system (Paton, 1990).

REFERENCES

Andrews, S. (2010). A quality control tool for high throughput sequence data. *Fast QC*, 532.

Alvarez, I. J. F. W., and Wendel, J. F. (2003). Ribosomal ITS sequences and plant phylogenetic inference. *Molecular phylogenetics and evolution*, 29(3), 417-434.

Badro, H., Furtado, A., and Henry, R. (2019). Relationships between Iraqi Rice Varieties at the Nuclear and Plastid Genome Levels. *Plants,* 8(11), 481.

Baldwin, B. G., Sanderson, M. J., Porter, J. M., Wojciechowski, M. F., Campbell, C. S., and 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(2), 247-277.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., and Pevzner P. A., (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology,* 19(5), 455-477.

Beier, S., Thiel, T., Munch, T., Scholz, U., and Mascher, M. (2017) MISA-web: a web server for microsatellite prediction. *Bioinformatics,* 33, 2583-2585.

Bendiksby, M., Thorbek, L., Scheen, A. C., Lindqvist, C., and Ryding, O. (2011). An updated phylogeny and classification of Lamiaceae subfamily Lamioideae. *Taxon,* 60(2), 471-484.

Bentham, G. (1834). *Scutellaria* and *Perilomia. In*: Bentham, G., (ed). Labiatarum Genera and Species. *London: Ridgeway & Sons,* 416-446.

Benthem, G. (1876) *Scutellaria, Salazaria* and *Perilomia. In*: Bentham, G., Hooker, J. D., (eds). Genera Plantarum. *London: Lovell Reeve & Co,* , 1201-1203.

Bentham, G. (1876). Verbenaceae and Labiatae. *In* Genera Plantarum. Bentham, G., and Hooker, J. D., (eds), *Lovell Reeve & Co., Williams & Norgate, London,*2, 1131-1123.

Berkley, W. F. A. (1975). A note concerning two flowering plants. *Phytologia,* 32: 304.

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics,* 30(15), 2114-2120.

Borowiec, M. L. (2016). AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ*, 4, e1660.

Bramley, G. L. C., Forest, F., and De Kok, R. P. J. (2009) Troublesome tropical mints: re-examining generic limits of *Vitex* and relations (Lamiaceae) in South East Asia. *Taxon,* 58, 500-510.

Brewer, G. E., Clarkson, J. J., Maurin, O., Zuntini, A. R., Barber, V., Bellot, S., Biggs N., Cowan R. S., Davies N. M., Dodsworth S., Edwards S. L., Eiserhardt W. L., Epitawalage N., Frisby S., Grall A., Kersey P. J., Pokorny L., Leitch I. J., Forest F., and Baker W. (2019). Factors affecting targeted sequencing of 353 nuclear genes from herbarium specimens spanning the diversity of angiosperms. *Frontiers in Plant Science*, *10*, 1102.

Briquet, J. (1896). *Scutellaria, Salazaria* and *Perilomia.* In: Engler, A., Prantl, K. A. E. Die, (eds). *Naturlichen Pflanzenfamilien* ed. 1. 4(3a):[*Scutellaria* and *Salzaria:* 224-227;*Perilomia*: 232-233].

Briquet, J. (1895-1897). Labiatae. *Engler and Prantle's. Die Naturlichen Pflanzenfamilien 4.*

Bruun-Lund, S., Clement, W. L., Kjellberg, F., and Ronsted, N. (2017). First plastid phylogenomic study reveals potential cyto-nuclear discordance in the evolutionary history of *Ficus* L.(Moraceae). *Molecular Phylogenetics and Evolution*, 109, 93-104.

Bunsawat, J., Elliott, N. E., Hertweck, K. L., Sproles, E., and Alice, L. A. (2004). Phylogenetics of *Mentha* (Lamiaceae): evidence from chloroplast DNA sequences. *Systematic Botany,* 29(4), 959-964.

Chau, J. H., Rahfeldt, W. A., and Olmstead, R. G. (2018). Comparison of taxon-specific versus general locus sets for targeted sequence capture in plant phylogenomics. *Applications in Plant Sciences*, *6*(3), e1032.

Cantino, P. D. (1982). A monograph of the genus *Physostegia* (Labiatae). *Contributions from the grey Herbarium of Harvard University*, (211), 1-105.

Cantino, P. D. (1992a). Toward a phylogenetic classification of the Labiatae. *Advances in Labiatae science,* 27-37.

Cantino, P. D. (1992b). Evidence for a polyphyletic origin of the Labiatae. *Annals of the Missouri Botanical Garden,* 79(2), 361-379.

Cantino, P., Harley, R. M., and Wagstaff, S. J. (1992). Genera of Labiatae: status and classification. *Advances in labiate science*, 11, 511-522.

Chamala, S., Garcia, N., Godden, G. T., Krishnakumar, V., Jordon-Thaden, I. E., De Smet, R., Barbazuk, W. B., Soltis, D. E., and Soltis, P. S. (2015). MarkerMiner 1.0: A new application for phylogenetic marker development using angiosperm transcriptomes. *Applications in Plant Sciences*, 3(4), 1400115.

Chau, J. H., Rahfeldt, W. A., and Olmstead, R. G. (2018). Comparison

of taxon-specific versus general locus sets for targeted sequence capture in plant phylogenomics. *Applications in Plant Sciences*, *6*(3), e1032.

Cheon, K. S., Jeong, I. S., Kim, K. H., Lee, M. H., Lee, T. H., Lee, J. H., Yoon, U. H., Chandra, R., Lee, Y. J., and Kim, T. H. (2018). Comparative SNP analysis of chloroplast genomes and 45S nrDNAs reveals genetic diversity of *Perilla* species. *Plant Breeding and Biotechnology*, 6(2), 125-139.

Cantino, P. D., and Sanders, R. W. (1986). Subfamilial classification of Labiatae. *Systematic Botany,* 11(1), 163-185.

Couvreur, T. L., Helmstetter, A. J., Koenen, E. J., Bethune, K., Brandao, R. D., Little, S. A., Sauquet, H., and Erkens, R. H. (2019). Phylogenomics of the major tropical plant family Annonaceae using targeted enrichment of nuclear genes. *Frontiers in plant science*, *9*, 1941.

Cronn, R., Knaus, B. J., Liston, A., Maughan, P. J., Parks, M., Syring, J. V., and Udall, J. (2012). Targeted enrichment strategies for next-generation plant biology. *American Journal of Botany*, 99(2), 291-311.

Cronquist, A., and Takhtadzhi̇an, A. L. (1981). An integrated system of classification of flowering plants. *Columbia university press.*

Cronquist, A. (1988). The evolution and classification of flower plants. *The New York Botanical Garden, Bronx, NY.*

Curto, M. A., Puppo, P., Ferreira, D., Nogueira, M., and Meimberg, H. (2012). Development of phylogenetic markers from single-copy nuclear genes for multi locus, species level analyses in the mint family (Lamiaceae). *Molecular Phylogenetics and Evolution*, 63(3), 758-767.

Darling, A. C., Mau, B., Blattner, F. R., and Perna, N. T. (2004).

Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome research*, 14(7), 1394-1403.

Denk, T., Hipp, A. L., Manos, P. S., Hahn, M., Avishai, M., Bodenes, C., and Gailing, O. (2019). Genomic landscape of the global oak phylogeny. *New Phytologist*, 1-15.

Diaz-Tapia, P., Maggs, C. A., West, J. A., and Verbruggen, H. (2017). Analysis of chloroplast genomes and a super-matrix inform reclassification of the Rhodomelaceae (Rhodophyta). *Journal of Phycology*, 53(5), 920-937.

Dodsworth, S., Pokorny, L., Johnson, M. G., Kim, J. T., Maurin, O., Wickett, N. J., Forest, F., and Baker, W. J. (2019). Hyb-Seq for flowering plant systematics. *Trends in plant science,* 24(10), 887-891.

Dossa, K., Yu, J., Liao, B., Cisse, N., and Zhang, X. (2017). Development of highly informative genome-wide single sequence repeat markers for breeding applications in sesame and construction of a web resource: SisatBase. *Frontiers in plant science*, 8, 1470.

Duarte, J. M., Wall, P. K., Edger, P. P., Landherr, L. L., Ma, H., Pires, P. K., Leebens-Mack, J., and Claude, W. D. (2010). Identification of shared single copy nuclear genes in *Arabidopsis, Populus, Vitis* and *Oryza* and their phylogenetic utility across various taxonomic levels. *BMC Evolutionary Biology,* 10(1), 61.

Edger, P. P., Tang, M., Bird, K. A., Mayfield, D. R., Conant, G., Mummenhoff, K., Koch, M. A., and Pires, J. C. (2014). Secondary structure analyses of the nuclear rRNA internal transcribed spacers and assessment of its phylogenetic utility across the Brassicaceae (Mustards). *PloS one*, 9(7), e101341.

Edmondson J. R. (1982). *Scutellaria. In*: Davis, P. H., (ed). Flora of Turkey. 7. *Edinburgh: University Press*. 78-100.

Egan, A. N., Schlueter, J., and Spooner, D. M. (2012). Applications of next-generation sequencing in plant biology. *American journal of botany*, 99(2), 175-185.

Erdlman, G. (1945). Pollen morphology and plant taxonomy. IV. Labiatae, Verbenaceae and Avicenniaceae. *Svensk botanisk tidskrift*, 39, 276-285.

Fragoso-Martinez, I., Salazar, G. A., Martinez-Gordillo, M., Magallon, S., Sanchez-Reyes, L., Lemmon, E. M., Lemmon, A. R., Sazatornil, F., and Mendoza, C. G. (2017). A pilot study applying the plant Anchored Hybrid Enrichment method to New World sages (*Salvia* subgenus *Calosphace;* Lamiaceae). *Molecular Phylogenetics and Evolution*, 117, 124-134.

Freire, M. C. M., da Silva, M. R., Zhang, X., Almeida, A. M. R., Stacey, G., and de Oliveira, L. O. (2012). Nucleotide polymorphism in the 5.8S nrDNA gene and internal transcribed spacers in *Phakopsora pachyrhizi* viewed from structural models. *Fungal Genetics and Biology*, 49(2), 95-100.

Galili, T. (2015). dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics,* 31(22), 3718-3720.

Giudicelli, G. C., Mader, G., Silva-Arias, G. A., Zamberlan, P. M., Bonatto, S. L., and Freitas, L. B. (2017). Secondary structure of nrDNA Internal Transcribed Spacers as a useful tool to align highly divergent species in phylogenetic studies. *Genetics and molecular biology*, 40(1), 191-199.

Goertzen, L. R., Cannone, J. J., Gutell, R. R., and Jansen, R. K. (2003). ITS secondary structure derived from comparative analysis: implications for sequence alignment and phylogeny of the Asteraceae. *Molecular phylogenetics and evolution*, 29(2), 216-234.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., Palma, F. di., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., and Regev, A. (2011). Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature biotechnology,* 29(7), 644.

Grover, C. E., Salmon, A., and Wendel, J. F. (2012). Targeted sequence capture as a powerful tool for evolutionary analysis. *American journal of botany*, 99(2), 312-319.

Gupta, P. K., Balyan, H. S., Sharma, P. C., and Ramesh, B. (1996). Microsatellites in plants: a new class of molecular markers. *Current Science*, 70, 45-54.

Han, K., Lee, Y., and Kim, W. (2002). PseudoViewer: automatic visualization of RNA pseudoknots. *Bioinformatics,* 18(supple_1), S321-S328

Harley, R. M., Atkins, S., Budantsev, A. L., Cantino, P. D., Conn, B. J., Grayer, R., Harley, M. M., Kok, R. de, Krestovskaja, T., Morales, R., Paton, A. J., Ryding, O., and Upson, T. (2004). Labiatae. *In* Flowering Plants · Dicotyledons (pp. 167-275). *Springer, Berlin, Heidelberg.*

Hamilton, A. (1832). Esquisse d'une monographie du genre Scutellariao utoque. *Lyon: Louis Perrin.*

Herrando-Moraira, S., Calleja, J. A., Galbany-Casals, M., Garcia-Jacas, N., Liu, J. Q., Lopez-Alvarado, J., Lopez-Pujol, J., Mandel, J. R., Masso, S., Mehregan, I., Montes-Morenos, N., Pyak, E., Roquet, C., Saez, L., Sennikow, A., Susanna, A., and Roquet, C. (2019). Nuclear and plastid DNA phylogeny of tribe Cardueae (Compositae) with Hyb-Seq data: A new subtribal classification and a temporal diversification framework. *Molecular phylogenetics and evolution*,

137, 313-332.

Hillis, D. M., and Dixon, M. T. (1991). Ribosomal DNA: molecular evolution and phylogenetic inference. *The Quarterly review of biology*, 66(4), 411-453.

Johnson, M. G., Gardner, E. M., Liu, Y., Medina, R., Goffinet, B., Shaw, A. J., Zerega, N. J. C., and Wickett, N. J. (2016). HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Applications in plant sciences*, 4(7), 1600016.

Larridon, I., Villaverde, T., Zuntini, A. R., Pokorny, L., Brewer, G. E., Epitawalage, N., Fairlie, I., Hahn, M., Kim, J., Maguilla, E., Maurin, O., Xanthos, M., Hipp, A. L., Forest, F., and Baker, W. J. (2019). Tackling rapid radiations with targeted sequencing. *Frontiers in plant science*, 10, 1655.

Li, B., Xu, W., Tu, T., Wang, Z., Olmstead, R. G., Peng, H., Francisco-Ortega, J., Cantino, P. D., and Zhang, D. (2012). Phylogenetic position of *Wenchengia* (Lamiaceae): a taxonomically enigmatic and critically endangered genus. *Taxon,* 61(2), 392-401.

Johnson, M. G., Pokorny, L., Dodsworth, S., Botigue, L. R., Cowan, R. S., Devault, A., Eiserhardt, W. L., Epitawalage, N., Forest, F., Kim, J. T., Leebens-Mack, J. H., Leitch, I. J., Maurin, O., Soltis, D. E., Soltis, P. S., Wong, G. K., Baker, W. J., and Wickett, N. J. (2019). A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-medoids clustering. *Systematic Biology*, *68*(4), 594-606.

Judd, W. S., Campbell, C. S., Kellogg, E. A., Stevens, P. F., and Donoghue, M. J. (1999). Lamiaceae. *In:* Plant systematics: a phylogenetic approach. *Sinauer Associates. Inc., Sunderland, U.S.A.* 492-494.

Junier, T., and Zdobnov, E. M. (2010). The Newick utilities: high-throughput phylogenetic tree processing in the UNIX shell. *Bioinformatics,* 26(13), 1669-1670.

Juzepczuk, S. V. (1954). *Scutellaria* L. *In*: Shishkin, B. K., Juzepczuk, S. V., (eds). Flora of the U.S.S.R. *Botanical Institute of the Academy of Sciences of the USSR. Moscow: Leningrad.* 20, 50-150.

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., von Haeseler, A., and Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature methods*, 14(6), 587-589.

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, 30(4), 772-780.

Kaufmann, M., and Wink, M. (1994). Molecular systematics of the Nepetoideae (family Labiatae): phylogenetic implications from *rbcL* gene sequences. *Zeitschrift fur Naturforschung C*, 49(9-10), 635-645.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., and Drummond, A. (2012). Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics.* 28(12): 1647-1649.

Kim, S. T. (2018). *Scutellaria. In*: Park, C. W., (eds). Flora of Korea. *National Institute of Biological Resources*. 6a, 86-90.

Kim, S. T., and Lee, S. T. (1995a). Fruit surface morphology of Scutellaria (Lamiaceae) in Korea and its taxonomic implication. Korean Journal of Plant Taxonomy, 25(3), 165-176

Kim, S. T., and Lee, S. T. (1995b). Taxonomy of the genus *Scutellaria* (Lamiaceae) in Korea. *Korean Journal of Plant Taxonomy*, 25(2),

71-102.

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. M*olecular biology and evolution,* 33, 1870-1874.

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular biology and evolution*, 35(6), 1547-1549.

Kurtz, S., Choudhuri, J. V., Ohlebusch, E., Schleiermacher, C., Stoye, J., and Giegerich, R. (2001). REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic acids research*, 29(22), 4633-4642.

Lagesen, K., Hallin, P., Rodland, E. A., Stærfeldt, H. H., Rognes, T., and Ussery, D. W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic acids research*, 35(9), 3100-3108.

Lane, T. M. (1983). Mericarp micromorphology of Great Plains *Scutellaria* (Labiatae). *The Southwestern Naturalist,* 28(1), 71-79.

Li, B., Cantino, P. D., Olmstead, R. G., Bramley, G. L., Xiang, C. L., Ma, Z. H., Tan, Y. H., and Zhang, D. X. (2016). A large-scale chloroplast phylogeny of the Lamiaceae sheds new light on its subfamilial classification. *Scientific reports*, 6(1), 1-18.

Li, B., and Olmstead, R. G. (2017). Two new subfamilies in Lamiaceae. *Phytotaxa,* 313(2), 222-226.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. *Bioinformatics,* 25(16), 2078-2079.

Lin, H. Y., Hao, Y. J., Li, J. H., Fu, C. X., Soltis, P. S., Soltis, D. E., and Zhao, Y. P. (2019). Phylogenomic conflict resulting from

ancient introgression following species diversification in *Stewartia sl.* (Theaceae). *Molecular phylogenetics and evolution*, 135, 1-11.

Linder, C. R., Goertzen, L. R., Heuvel, B. V., Francisco-Ortega, J., and Jansen, R. K. (2000). The complete external transcribed spacer of 18S-26S rDNA: amplification and phylogenetic utility at low taxonomic levels in Asteraceae and closely allied families. *Molecular Phylogenetics and Evolution*, 14(2), 285-303.

Liu, J., Hou, J., Jiang, C., Li, G., Lu, H., Meng, F., and Shi, L. (2015). Deep sequencing of the *Scutellaria baicalensis* Georgi transcriptome reveals flavonoid biosynthetic profiling and organ-specific gene expression. *PloS one*, 10(8), e0136397.

Liu, S., An, Y., Li, F., Li, S., Liu, L., Zhou, Q., Zaho, S., and Wei, C. (2018). Genome-wide identification of simple sequence repeats and development of polymorphic SSR markers for genetic studies in tea plant (*Camellia sinensis*). *Molecular Breeding*, 38(5), 59.

Long, C., and Kubatko, L. (2018). The effect of gene flow on coalescent-based species-tree inference. *Systematic biology*, 67(5), 770-785.

Mamanova, L., Coffey, A. J., Scott, C. E., Kozarewa, I., Turner, E. H., Kumar, A., Howard, E., Shendure, J., and Turner, D. J. (2010). Target-enrichment strategies for next-generation sequencing. *Nature methods*, 7(2), 111-118.

Menezes, A. P., Resende-Moreira, L. C., Buzatti, R. S., Nazareno, A. G., Carlsen, M., Lobo, F. P., Kalapothakis E., and Lovato, M. B. (2018). Chloroplast genomes of *Byrsonima* species (Malpighiaceae): comparative analysis and screening of high divergence sequences. *Scientific Reports*, 8(1), 1-12.

Mint Evolutionary Genomics Consortium. (2018). Phylogenomic mining of the mints reveals multiple mechanisms contributing to the

evolution of chemical diversity in Lamiaceae. *Molecular plant*, 11(8), 1084-1096.

Morales-Briones, D. F., Liston, A., and Tank, D. C. (2018). Phylogenomic analyses reveal a deep history of hybridization and polyploidy in the Neotropical genus *Lachemilla* (Rosaceae). *New Phytologist*, 218(4), 1668-1684.

Mullineux, T., and Hausner, G. (2009). Evolution of rDNA ITS1 and ITS2 sequences and RNA secondary structures within members of the fungal genera *Grosmannia* and *Leptographium*. *Fungal Genetics and Biology*, 46(11), 855-867.

Nazar, R. (2004). Ribosomal RNA processing and ribosome biogenesis in eukaryotes. *IUBMB life*, 56(8), 457-465.

Neiman, M., and Taylor, D. R. (2009). The causes of mutation accumulation in mitochondrial genomes. *Proceedings of the Royal Society B: Biological Sciences*, 276(1660), 1201-1209.

Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular biology and evolution*, 32(1), 268-274.

Paton, A. (1990). A global taxonomic investigation of *Scutellaria* (Labiatae). *Kew Bulletin*, 45(3), 399-450.

Pease, J. B., Brown, J. W., Walker, J. F., Hinchliff, C. E., and Smith, S. A. (2018). Quartet Sampling distinguishes lack of support from conflicting support in the green plant tree of life. *American journal of botany*, 105(3), 385-403.

Pedersen, J. A. (2000). Distribution and taxonomic implications of some phenolics in the family Lamiaceae determined by ESR spectroscopy. *Biochemical systematics and Ecology*, 28(3), 229-253.

Philippe, H., Brinkmann, H., Lavrov, D. V., Littlewood, D. T. J., Manuel, M., Worheide, G., and Baurain, D. (2011). Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS Biol*, 9(3), e1000602.

Rambaut, A., and Drummond, A. J. (2018). FigTree v1. 4. 4. Institute of Evolutionary Biology. *University of Edinburgh, Edinburgh.*

R Core Team. (2019). Core R: A Language and Environment for Statistical Computing, Version 3. 5. 3. *Vienna: R Foundation for Statistical Computing*.

Raubeson, L. A., Peery, R., Chumley, T. W., Dziubek, C., Fourcade, H. M., Boore, J. L., and Jansen, R. K. (2007). Comparative chloroplast genomics: analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. *BMC genomics*, 8(1), 174.

Rechinger, K. H. *Scutellaria. In*: Rechinger, K. H. (ed). Flora Iranica; 1982. 150, 48-84.

Resl, P. (2015). phylo-scripts: Python scripts for phylogenetics.

Rieseberg, L. H., and Soltis, D. E. (1991). Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary Trends in Plants*.

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Hohna, S., Larget, B., Liu, L., Suchard, M. A., and Huelsenbeck, J. P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic biology*, 61(3), 539-542.

Roure, B., Baurain, D., and Philippe, H. (2013). Impact of missing data on phylogenies inferred from empirical phylogenomic data sets. *Molecular biology and evolution*, 30(1), 197-214.

Rozas, J., Ferrer-Mata, A., Sanchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., and Sanchez-Gracia, A. (2017).

DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular biology and evolution*, 34(12), 3299-3302.

Ryding, O. (2007). Amount of calyx fibres in Lamiaceae, relation to calyx structure, phylogeny and ecology. *Plant systematics and evolution*, 268(1-4), 45-58.

Safikhani, K., Jamzad, Z., and Saeidi, H. (2018). Phylogenetic relationships in Iranian *Scutellaria* (Lamiaceae) based on nuclear ribosomal ITS and chloroplast *trnL-F* DNA data. *Plant Systematics and Evolution,* 304, 1077-1089.

Salmaki, Y., Zarre, S., Ryding, O., Lindqvist, C., Brauchler, C., Heubl, G., Barber, J., and Bendiksby, M. (2013). Molecular phylogeny of tribe Stachydeae (Lamiaceae subfamily Lamioideae). *Molecular Phylogenetics and Evolution*, 69(3), 535-551.

Sand, A., Holt, M. K., Johansen, J., Brodal, G. S., Mailund, T., and Pedersen, C. N. (2014). tqDist: a library for computing the quartet and triplet distances between binary or general trees. *Bioinformatics,* 30(14), 2079-2080.

Schmickl, R., Liston, A., Zeisek, V., Oberlander, K., Weitemier, K., Straub, S. C., Cronn, R. C., Dreyer, L. L., and Suda, J. (2016). Phylogenetic marker development for target enrichment from transcriptome and genome skim data: the pipeline and its application in southern African *Oxalis* (Oxalidaceae). *Molecular Ecology Resources*, 16(5), 1124-1135.

Schultz, J., Maisel, S., Gerlach, D., Muller, T., and Wolf, M. (2005). A common core of secondary structure of the internal transcribed spacer 2 (ITS2) throughout the Eukaryota. *Rna*, 11(4), 361-364.

Silvestro, D., and Michalak, I. (2012). raxmlGUI: a graphical front-end for RAxML. *Organisms Diversity & Evolution*, 12(4), 335-337.

Slater, G. S. C., and Birney, E. (2005). Automated generation of

heuristics for biological sequence comparison. *BMC bioinformatics*, 6(1), 31.

Smith, M. R. (2019). Quartet: comparison of phylogenetic trees using quartet and bipartition measures. *Zenodo* (10.5281/zenodo. 2536318), Geneva: Zenodo.

Smith, S. A., Moore, M. J., Brown, J. W., and Yang, Y. (2015). Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. *BMC evolutionary biology,* 15(1), 150.

Solis-Lemus, C., Yang, M., and Ane, C. (2016). Inconsistency of Species Tree Methods under Gene Flow. *Systematic biology*, 65(5), 843-851.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics,* 30(9), 1312-1313.

Steane, D. A., Scotland, R. W., Mabberley, D. J., Wagstaff, S. J., Reeves, P. A., and Olmstead, R. G. (1997). Phylogenetic relationships of *Clerodendrum* s.l. (Lamiaceae) inferred from chloroplast DNA. *Systematic Botany*, 22(2), 229-243.

Stevens, P. F. (2015). Angiosperm Phylogeny Website (version 14). Retrieved July 2017, from http://www.mobot.org/MOBOT/research/APweb/.

Suh, Y. B., S. P. Hong and S. J. Park. (2007). Lamiaceae. *In* The Genera of Vascular Plants of Korea. Park, C. W., (ed.), *Seoul (Korea): Academy Publishing Co.*

Team, R. (2015). RStudio: integrated development for R. RStudio, *Inc. Boston, MA, USA*, 42.

The Angiosperm Phylogeny Group, Chase, M. W., Christenhusz, M. J. M., Fay, M. F., Byng, J. W., Judd, W. S., Soltis, D. E., Mabberley,

D. J., Sennikow, A. N., Soltis, P. S., and Stevens, P. F. (2016). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society,* 181(1), 1-20.

The Plant List (2013). Version 1.1. Published on the Internet. Accessed 1st January. from http://www.theplantlist.org/.

Thorne, R. F. (1992). Classification and geography of the flowering plants. *The botanical review,* 58(3), 225-327.

Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2003) Multiple sequence alignment using ClustalW and ClustalX. *Current protocols in bioinformatics,* 1, 2-3.

Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E. S., Fischer, A., Bock, R., and Greiner, S. (2017). GeSeq - versatile and accurate annotation of organelle genomes. *Nucleic Acids Research*, 45, 6-11.

Torrey, J. (1858). *Salazaria* in Emory, W. H. Rep. U. S. Mex. Bound. 2(1), *Botany of the Boundary,* 133-135.

Twyford, A. D., and Ness, R. W. (2017). Strategies for complete plastid genome sequencing. *Molecular ecology resources*, 17(5), 858-868.

Uribe-Convers, S., Carlsen, M. M., Lagomarsino, L. P., and Muchhala, N. (2017). Phylogenetic relationships of *Burmeistera* (Campanulaceae: Lobelioideae): combining whole plastome with targeted loci data in a recent radiation. *Molecular phylogenetics and evolution*, 107, 551-563.

Van Dijk, E. L., Auger, H., Jaszczyszyn, Y., and Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends in genetics*, 30(9), 418-426.

Vargas, O. M., Heuertz, M., Smith, S. A., and Dick, C. W. (2019).

Target sequence capture in the Brazil nut family (Lecythidaceae): Marker selection and in silico capture from genome skimming data. *Molecular phylogenetics and evolution*, 135, 98-104.

Villaverde, T., Jimenez-Mejias, P., Luceno, M., Waterway, M. J., Kim, S., Lee, B., Rincon-Barrado, M., Hahn, M., Maguilla, E., Roalson, E. H., Hipp, A. L., and The Global Carex Group (2020). A new classification of *Carex* (Cyperaceae) subgenera supported by a HybSeq backbone phylogenetic tree. B*otanical Journal of the Linnean Society*, 194(2), 141-163.

Wagner, N. D., He, L., and Horandl, E. (2019). Relationships and genome evolution of polyploid *Salix* species revealed by RAD sequencing data. *bioRxiv,* 864504.

Wagstaff, S. J., Hickerson, L., Spangler, R., Reeves, P. A., and Olmstead, R. G. (1998). Phylogeny in Labiatae s.l., inferred from cpDNA sequences. *Plant Systematics and Evolution,* 209(3-4), 265-274.

Wagstaff, S. J., and Olmstead, R. G. (1997). Phylogeny of Labiatae and Verbenaceae inferred from *rbcL* sequences. *Systematic Botany,* 165-179.

Wagstaff, S. J., Olmstead, R. G., and Cantino, P. D. (1995). Parsimony analysis of cpDNA restriction site variation in subfamily Nepetoideae (Labiatae). A*merican Journal of Botany*, 82(7), 886-892.

Weitemier, K., Straub, S. C., Cronn, R. C., Fishbein, M., Schmickl, R., McDonnell, A., and Liston, A. (2014). Hyb-Seq: Combining target enrichment and genome skimming for plant phylogenomics. *Applications in Plant Sciences,* 2(9), 1400042.

Williams, A. V., Miller, J. T., Small, I., Nevill, P. G., and Boykin, L. M. (2016). Integration of complete chloroplast genome sequences with

small amplicon datasets improves phylogenetic resolution in *Acacia*. *Molecular Phylogenetics and Evolution*, 96, 1-8.

Wink, M., and Kaufmann, M. (1996). Phylogenetic relationships between some members of the subfamily Lamioideae (family Labiatae) inferred from nucleotide sequences of the *rbcL* gene. *Botanica Acta*, 109(2), 139-148.

Wortley, A. H., Rudall, P. J., Harris, D. J., and Scotland, R. W. (2005). How much data are needed to resolve a difficult phylogeny? Case study in Lamiales. *Systematic biology*, 54(5), 697-709.

Wu, C. Y., and Li, X. W. (1977) *Scutellaria* L. In: Wu C. Y., and Li X. W., (ed). Flora reipublicae popularis sinicae. 65. *Beijing: Science Press,* 124-248.

Wu, T. D., Reeder, J., Lawrence, M., Becker, G., and Brauer, M. J. (2016). GMAP and GSNAP for genomic sequence alignment: enhancements to speed, accuracy, and functionality. *In Statistical genomics,* 283-334. *Humana Press, New York, NY.*

Wunderlich, R. (1967). Ein Vorschlag zu einer nattrlichen Gliederung der Labiaten auf Grund der Pollenkorner, der Samenentwicklung und des reifen Samens. *Oesterreichische Botanische Zeitschrift*, 114(4-5), 383-483.

Yang, J. B., Yang, S. X., Li, H. T., Yang, J., and Li, D. Z. (2013). Comparative chloroplast genomes of *Camellia* species. *PLoS One*, 8(8), e73053.

Yuan, Y. W., Liu, C., Marx, H. E., and Olmstead, R. G. (2009). The pentatricopeptide repeat (PPR) gene family, a tremendous resource for plant phylogenetic studies. *New Phytologist*, 182(1), 272-283.

Zarrei, M., Stefanović, S., and Dickinson, T. A. (2014). Reticulate evolution in North American black-fruited hawthorns (*Crataegus* section *Douglasia*; Rosaceae): evidence from nuclear ITS2 and

plastid sequences. *Annals of Botany*, 114(2), 253-269.

Zhang, C., Rabiee, M., Sayyari, E., and Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC bioinformatics*, 19(6), 153.

Zhang, W., Yuan, Y., Yang, S., Huang, J., and Huang, L. (2015). ITS2 secondary structure improves discrimination between medicinal "Mu Tong" species when using DNA barcoding. *PLoS One*, 10(7), e0131185.

Zhao, F., Li, B., Drew, B. T., Chen, Y. P., Wang, Q., Yu, W. B., Liu, E. D., Salmaki, Y., Peng, H., and Xiang, C. L. (2020). Leveraging plastomes for comparative analysis and phylogenomic inference within Scutellarioideae (Lamiaceae). *PloS one*, 15(5), e0232602.

Zhao, F., Liu, E., Peng, H., and Xiang, C. (2017). A new species of *Scutellaria* (Scutellarioideae, Lamiaceae) from Si-chuan Province in southwest China. *PeerJ,* 5, e3624.

국문초록

이윤경
생물학과
성신여자대학교 대학원

꿀풀과에 속하는 골무꽃속(*Scutellaria*)은 과 내에서 두 번째로 큰 속으로, 세계적으로 약 470 분류군이 포함된다. 많은 식물들이 약용식물 또는 향신료로 쓰이며 주로 온대 지역에 널리 분포한다. 일년생 또는 다년생 초본식물로 합판악의 숙존성, 자방병의 존재, 무배유성 등의 특징으로 꿀풀과의 다른 속들과 구분된다.

골무꽃속 내의 형태학적 분류체계는 여러 학자에 의해 제시되어왔지만, 대부분 지역적으로 한정된 종들을 대상으로 연구되었다. 전 세계적으로 분포하는 골무꽃속에 대한 연구는 Paton (1990)에 의해 수행되었는데, 골무꽃속에 속하는 110종을 대상으로 외부형태학적인 고찰을 통해 2아속 7절 2아절 38 종그룹(비공식 아절 이하 범주)으로 속 내 분류체계가 정립된 바 있고, 이는 현재까지 널리 받아들여지고 있다. 최근 골무꽃속에 대한 분자계통학적인 연구는 Zhao et al. (2017), Safikhani et al. (2018)에 의해 연구된 바 있지만, 이들은 지역적으로 한정된 종들을 그 대상으로 하여 골무꽃 속 내 분류군들의 유연관계를 이해하는 데에는 부족하였다.

본 연구에서는, 현재까지 수행된 바 없는 전 세계 골무꽃속 식물의 계통유전체학적인 연구를 통해 이들에 대한 진화와 종 분화를 이해하고자 하였다. 본 연구의 목적은 1) 골무꽃속의 분자 계통학적 정보에 의해 골무꽃속

내의 계통을 명확히 하며 이를 기존의 분류체계와 비교하고, 2) 분자계통학적 연구에 의해 얻어진 속 내 주요 그룹이 갖는 외부 형질을 재검토하고, 3) 비교유전체학적 정보로부터 미래의 새로운 분류체계 제시를 위한 분자계통학적 증거를 제공하고자 한다.

연구의 첫 번째 단계로, 이 전에 보고되었던 엽록체 유전체들(*S. baicalensis*, *S. insignis*)을 사용한 분석을 기반으로 빠르게 진화하는 DNA 유전자들을 선별하였다(*ccsA*, *matK, ndhF*). 계통 분석에는 골무꽃속 내의 주요 그룹들을 대표하는 분류군들을 포함하여 총 200 분류군이 분석에 포함되었다. 계통 분석 결과, 여섯 개의 주요 분기군이 인식되었으며 *Apeltanthus* 아속은 단계통군, *Scutellaria* 아속은 측계통군으로 나타났다. 또한, *Apeltanthus* 아속에 포함되는 *Apeltanthus, Lupulinaria* 절은 모두 측계통군이었으며, *Scutellaria* 아속에 포함되는 절 중 *Perilomia* 절만 단계통군이었고, 나머지 절과 종그룹들은 모두 측계통군으로 나타났다.

골무꽃속 식물들의 계통학적 관계를 해결하기 위한 두 번째 단계로 골무꽃속 식물들의 전체 엽록체 유전체를 시퀀싱하였고, 군외군을 포함하여 47 분류군이 전체 엽록체 유전체 시퀀싱에 포함되었다. 골무꽃속 식물들의 엽록체 전체 길이는 151,030 bp에서 152,594 bp를 나타냈고, 이들은 한 쌍의 IR 구간 (25,190−25,314 bp)과 LSC구간(83,008−84,622 bp), SSC 구간 (17,292−17,600 bp)으로 구성되었다. 모든 엽록체 유전체는 80개의 단백질 암호화 유전자, 30개의 tRNA 유전자와 4개의 rRNA 유전자로 이루어져 있었고, 매우 보존적인 구조를 보여주었다. 또한, 전체 엽록체 유전체에 대한 염기서열 분석을 통해 빠르게 진화하는 엽록체 구간을 조사하였고, 상위 10개의 구간을 선별한 결과 *trnK−rps16, ycf1*, *ndhF, petA−psbJ*, *petN−psbM*, *ccsA−ndhD*, *trnT*−UGU−*trnL*−UAA, *rpoB−trnC*, *rpl32−ccsA* and *rbcL−accD* 구간 순으로 다양성이 높게 나타났다. 계통 분석은 super−matrix 접근법을 기반으로 이전의 세 엽록체

구간(*ccsA*, *matK*, *ndhF*) 시퀀스들과 50분류군의 엽록체 유전체 시퀀스들을 결합하여 총 213 분류군을 대상으로 계통 분석이 수행되었다. 계통 분석 결과, 분류군들의 계통학적 관계는 *Salviifoliae* 절을 제외하고 세 엽록체 유전자를 사용한 이전의 계통수와 유사하게 나타났고, 신뢰도가 대부분 높아진 것을 확인하였다.

세 번째 단계로, 272개의 핵 구간을 포함한 핵 유전체를 기반으로 골무꽃속 식물들에 대한 계통 분석이 수행되었으며, 분석에는 군외군을 포함하여 95 분류군이 포함되었다. 각 구간에서 정렬된 염기서열들을 하나의 matrix로 연결한 전체 길이는 965,469 bp였다. 연결된 matrix를 기반으로 한 Maximum-Likelihood (ML) 계통수는 높은 신뢰도와 함께 해결된 계통 유연관계를 보여주었다. 이전 super-matrix 엽록체 계통수와 비교하였을 때, 핵 유전체 기반 ML 계통수는 해상력이 매우 개선되었음을 확인하였다. 또한, *Apeltanthus* 아속의 단계통성과 *Scutellaria* 아속의 측계통성이 명확하게 재인식되었다. 이전 엽록체의 super-matrix 계통수와는 달리, *Salviifoliae* 절에 속하는 두 종 (*S. heterophylla*, *S. pontica*)이 측계통군으로 나타남에 따라 이 종들이 모계 관계에 더 가까움을 확인하였다. 이전의 엽록체 계통수들과 유사하게, *Salazaria* 아절은 *Scutellaria* 아절에 속하는 일부 종들을 포함하는 분기군과 자매 관계를 나타냈다. 또한, *Perilomia* 절은 단계통군으로 나타났고, *Anaspis* 절은 계통수의 기부에 위치하였다. 272개 구간들 중 추후 계통 분석에 유용하게 쓰일 최상의 핵 DNA 구간을 찾기 위해, 개별 유전자 계통수와 핵 유전체 ML 계통수 사이에서 공유된 quartet의 수를 계산하였다. 계산 결과, DN17712_c1_g5_i1 (*Arabidopsis* AT3G54460: SNF2 도메인, helicase 도메인, F-box family 단백질 관여 유전자) 구간은 이 구간 하나만으로도 전체 자료와 87.27 %에 해당하는 공유도를 갖고 있음이 확인되었다.

위의 연구와는 독립적인 연구로서 골무꽃속 49 분류군에 대한 핵 리보

솜 DNA 염기서열이 결정되었고, ITS2 구간에 대한 2차 구조의 구조적 변이를 확인하였다. *S. orientalis*를 대상으로, 핵 리보솜 DNA의 한 반복 단위 전체의 서열이 결정되었고, 그 길이는 8,047 bp였다. 핵 리보솜 DNA 구간을 기반으로 계통 분석한 결과, 분류군들은 이전 연구의 계통수들과 유사한 계통 유연관계를 보였다. 진핵생물에서 ITS2 2차 구조는 네 가지 공통적인 특징을 보인다고 제시된 바 있지만(Schultz et al., 2005), 골무꽃속 식물들은 공통적인 특징을 공유하지 않았다. 또한 ITS2 2차 구조는 계통수에서 골무꽃속 내 하위그룹들을 구분할 수 있는 특별한 구조를 가지고 있지 않았으며, 이는 골무꽃속 내의 각각의 분류군에서 독립적인 구조적 진화가 일어났음을 제시하고 있다.

본 연구에서는 최초로 골무꽃속 식물들의 전 세계적인 샘플링과 대량의 엽록체 데이터와 핵 데이터에 의한 분석을 통해 골무꽃속 내 계통학적 유연관계를 보여주었다: 52 분류군의 엽록체 유전체 데이터(150 kbp)와 272개의 핵 구간 데이터(960 kbp). 엽록체와 핵 유전체 데이터를 기반으로 한 두 계통수에서 공통적으로 네 개의 주요 분기군들이 인식되었는데, 이는 Paton에 의해 제시된 분류체계(Paton, 1990)와는 매우 달랐으며, 이들은 추후 새로운 아속으로서 제시될 가능성이 있음을 확인하였다. 네 분기군 중 두 개의 분기군에 있어서, 각각의 분류군들은 양쪽 계통수에서 서로 다른 위치를 보이고 이러한 관계는 각각 높은 신뢰도를 나타내고 있어, 이는 이들 분류군들이 hybridization에 의한 기원으로 형성되었음을 제시하고 있다. 또한, 모든 계통수에서 공통적으로 나타나고 강하게 지지 되는 분기군들은 추후 새로운 계통학적 분류체계(phylogenetic classification)에서 인식될 수 있는 그룹으로 정의될 수 있을 것이다. 따라서, 본 연구 결과는 골무꽃속 내 새로운 분류체계를 확립하는 데에 중요한 기초 자료가 될 것이다.

APPENDICES FOR CHAPTER 2

Appendix 1. The list of species included in this chapter with voucher information and GenBank accession numbers.

Appendix 2. A phylogenetic tree using BI analysis based on three chloroplast genes (*ccsA*, *matK*, and *ndhF*).

Appendix 1. The list of species included in this chapter with voucher information and GenBank accession numbers.

Taxon name	Country	Collector and collector number (Herbarium code)	GenBank accession no. for *ccsA*	GenBank accession no. for *matK*	GenBank accession no. for *ndhF*
Holmskioldia sanguinea Retz.	Republic of Korea (cultivated)	Y. Lee 2019-0001 (SWU)	MN725115	MN725316	MN725517
Scutellaria adsurgens Popov	Ukbekistan	M. Popov & A. Vvedensky 6755 (K)	MN725116	MN725317	MN725518
Scutellaria agrestis A. St. -Hil. ex Benth.	Ecuador	T. Delgado 75 (MO)	MN725117	MN725318	MN725519
Scutellaria alabamensis Alexander	USA	R. Kral 53165 (MO)	MN725118	MN725319	MN725520
Scutellaria albida subsp. *colchica* (Rech. f.) J. R. Edm.	Turkey	Davis & Hedge D.32112 (NHM)	MN725119	MN725320	MN725521
Scutellaria albida subsp. *condensata* (Rech. f.) J. R. Edm.	Iraq	R. W. Haines s. n. (K)	MN725120	MN725321	MN725522
Scutellaria albituba A. Pool	Costa Rica	V. J. Dryer 1675 (MO)	MN725121	MN725322	MN725523
Scutellaria alborosea Lem.	Ecuador	T. Croat 89282 (MO)	MN725122	MN725323	MN725524
Scutellaria alpina L.	Spain	Mason & Smallcombe s. n. (K)	MN725123	MN725324	MN725525
Scutellaria altamaha Small	USA	R. Kral 51180 (MO)	MN725124	MN725325	MN725526
Scutellaria altissima L.	Macedonia	J. D. A. Stainton 7826 (K)	MN725125	MN725326	MN725527
Scutellaria amabilis H. Hara	Japan	s. n. (KYO)	MN725126	MN725327	MN725528
Scutellaria amoena C. H. Wright	China	Y. D. Zhu 65-0185 (KUN)	MN725127	MN725328	MN725529
Scutellaria angustifolia Pursh	USA	R. Olmstead 499 (WTU)	MN725129	MN725330	MN725531
Scutellaria angustifolia subsp. *micrantha* Olmstead	North America	J. B. Leiberg 2260 (NHM)	MN725128	MN725329	MN725530
Scutellaria antirrhinoides Benth.	USA	R. Olmstead 421 (WTU)	MN725130	MN725331	MN725532
Scutellaria arabica Jaub. & Spach	Saudi Arabia	S. Collenette 6659 (K)	MN725131	MN725332	MN725533
Scutellaria arenicola Small	USA	R. Kral 58563 (MO)	MN725132	MN725333	MN725534
Scutellaria ariana Hedge	Afghanistan	P. Furse 5873 (K)	MN725133	MN725334	MN725535
Scutellaria aurata Lem.	Ecuador	D. Neill & H. Vargar 12251 (MO)	MN725134	MN725335	MN725536
Scutellaria aurea B. L. Rob. & Greenm.	Mexico	D. E. Breedlove 39998 (MO)	MN725135	MN725336	MN725537
Scutellaria baicalensis Georgi	USA	R. Olmstead 98-54 (WTU)	MN725136	MN725337	MN725538
Scutellaria balearica Barcel?	Spain	Marcos 67 (K)	MN725137	MN725338	MN725539
Scutellaria bambusetorum C. Y. Wu	China	D. D. Tao 256 (KUN)	MN725138	MN725339	MN725540
Scutellaria barbata D. Don	China	W. T. Tsang 20415 (KUN)	MN725139	MN725340	MN725541
Scutellaria benthamiana (Mansf.) Epling	Peru	I. Shonle 2 (MO)	MN725140	MN725341	MN725542

Scutellaria bolanderi A. Gray	USA	R. Olmstead 395 (WTU)	MN725142	MN725343	MN725544
Scutellaria bolanderi subsp. *austromontana* Epling	USA	T. S. Eilias 10156 (PE)	MN725141	MN725342	MN725543
Scutellaria brachyspica Nakai & H. Hara	Japan	S. Kiyono no. 1 (KYO)	MN725143	MN725344	MN725545
Scutellaria brevibracteata subsp. *subvelutina* (Rech. f.) Greuter & Burdet	Jordan	K. Abulaia & al. 2008JOR1-2 (K)	MN725144	MN725345	MN725546
Scutellaria bucharica Juz.	Uzbekistan	G. S. Grigorjev s. n. (K)	MN725145	MN725346	MN725547
Scutellaria bushii Britton	USA	G. Allington 13-49 (MO)	MN725146	MN725347	MN725548
Scutellaria caerulea Moc. & Sesse ex Benth.	Mexico	M. Egger & D. Tank 1191 (WTU)	MN725147	MN725348	MN725549
Scutellaria calcarata C. Y. Wu & H. W. Li	China	C. W. Fan 879 (KUN)	MN725148	MN725349	MN725550
Scutellaria columnae All.	North Africa	E. Reverchon s. n. (NHM)	MN725149	MN725350	MN725551
Scutellaria comosa Juz.	Kyrgyzstan	Judanov s. n. (K)	MN725150	MN725351	MN725552
Scutellaria cordifrons Juz.	Uzbekistan	M. Popov & M. Abdullaova 6756 (K)	MN725151	MN725352	MN725553
Scutellaria costaricana H. Wendl.	USA	R. Olmstead s. n. (WTU)	MN725152	MN725353	MN725554
Scutellaria cypria Rech. f.	Cyprus	E. W. Kenedy 815 (K)	MN725153	MN725354	MN725555
Scutellaria dependens Maxim.	Japan	K. Yonekura 12880 (KYO)	MN725154	MN725355	MN725556
Scutellaria discolor Colebr.	Myanmar	J. Murata & al. 25593 (KYO)	MN725155	MN725356	MN725557
Scutellaria drummondii Benth.	USA	J. Bacon 2117 (WTU)	MN725156	MN725357	MN725558
Scutellaria dumetorum Schltdl.	Mexico	J. C. Ledesma-Corral 1803 (MO)	MN725157	MN725358	MN725559
Scutellaria ebracteata A. Pool	Nicaragua	R. Rueda & al. 13382 (MO)	MN725158	MN725359	MN725560
Scutellaria edelbergii Rech. f.	Afghanistan	C. Grey-Wilson & T. F. Hewer 1048 (K)	MN725159	MN725360	MN725561
Scutellaria elliptica Muhl.	USA	J. R. Abbot 2172 (MO)	MN725160	MN725361	MN725562
Scutellaria eplingii Legname	Bolivia	J. Balcazar 326 (MO)	MN725161	MN725362	MN725563
Scutellaria farsistanica Rech. f.	Iran	K-H. Rechinger 47342 (K)	MN725162	MN725363	MN725564
Scutellaria floridana Chapm.	USA	H. N. Moldenke 1147 (K)	MN725163	MN725364	MN725565
Scutellaria forrestii Diels	China	L. Zhenwen & Zhoujing 1041 (KH)	MN725164	MN725365	MN725566
Scutellaria franchetiana H. Lev.	China	D. Z. Ji 431 (PE)	MN725165	MN725366	MN725567
Scutellaria galericulata L.	USA	R. Olmstead s. n. (WTU)	MN725166	MN725367	MN725568
Scutellaria galerita Epling	Costa Rica	R. L. Wilber & K. Williams 66836 (MO)	MN725167	MN725368	MN725569
Scutellaria gardoquioides (Benth.) Benth.	Peru	E. M. Ortiz V. 1165 (MO)	MN725168	MN725369	MN725570
Scutellaria gaumeri Leonard	Mexico	B. Mogensen 1130 (MO)	MN725169	MN725370	MN725571
Scutellaria glabriuscula Fernald	USA	R. Kral 92043 (MO)	MN725170	MN725371	MN725572
Scutellaria glutinosa Benth.	Afghanistan	K-H. Rechinger 37191 (K)	MN725171	MN725372	MN725573
Scutellaria grandiflora Sims	Mongolia	F. Pipe-Wolferstan & B. Phillips FPW 235	MN725172	MN725373	MN725574

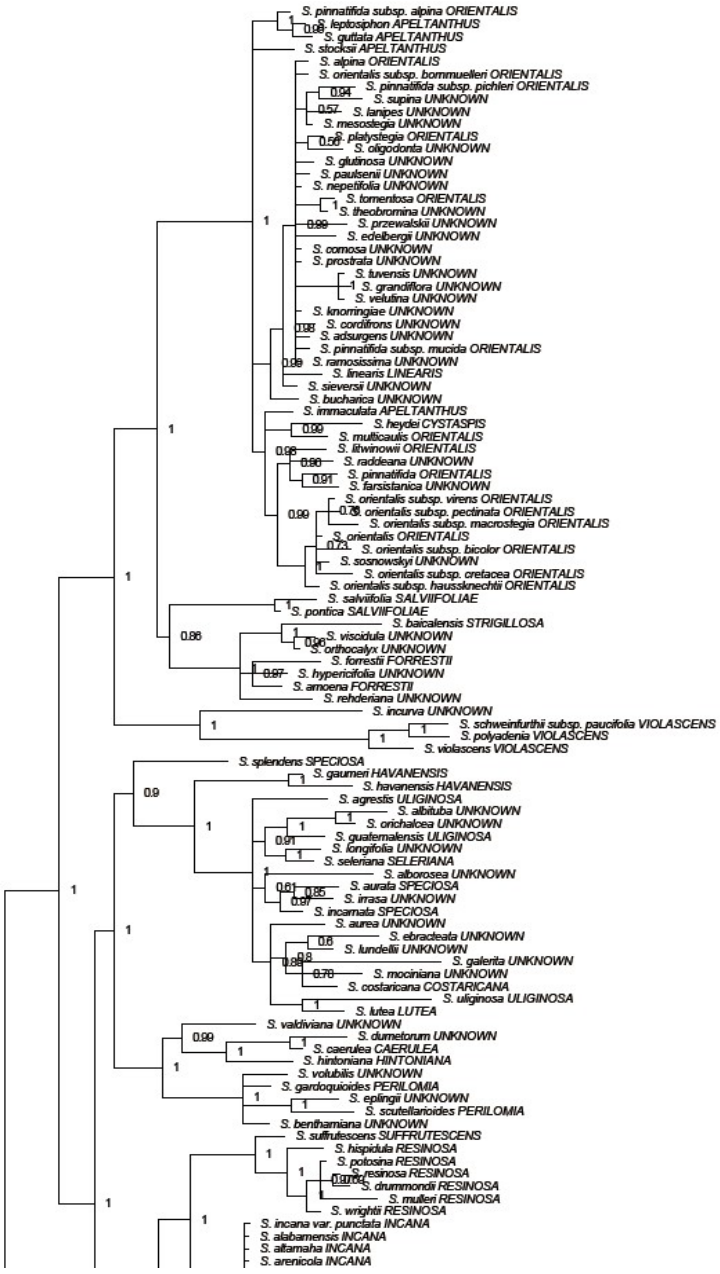
		(K)			
Scutellaria grossa Wall.	Nepal	C. Grey-Wilson & Phillips 359 (K)	MN725173	MN725374	MN725575
Scutellaria guatemalensis Leonard	Mexico	D. E. Breedlove 37309 (MO)	MN725174	MN725375	MN725576
Scutellaria guilielmii A. Gray	Japan	T. Miyazaki 605186 (KH)	MN725175	MN725376	MN725577
Scutellaria guttata Nevski ex Juz	Uzbekistan	I. Vassilczenko 7440 (K)	MN725176	MN725377	MN725578
Scutellaria hastifolia L.	Czech Republic	M. Dvorakova & al. 1552 (K)	MN725177	MN725378	MN725579
Scutellaria havanensis Jacq.	Bahamas	D. S. Correll 46132 (MO)	MN725178	MN725379	MN725580
Scutellaria heterophylla Montbret & Aucher ex Benth.	Turkey	B. Balansa 498 (K)	MN725179	MN725380	MN725581
Scutellaria heydei Hook. f.	Afghanistan	C. Grey-Wilson & T. F. Hewer 1625 (K)	MN725180	MN725381	MN725582
Scutellaria hintoniana Epling	Mexico	A. de Avila 662 (MO)	MN725181	MN725382	MN725583
Scutellaria hirta Sm.	Greece	F. Lemperg 521 (K)	MN725182	MN725383	MN725584
Scutellaria hispidula B. L. Rob.	Mexico	J. N. Rose 2555 (K)	MN725183	MN725384	MN725585
Scutellaria hissarica B. Fedtsch.	Tajikistan	L. P. Markova & al. s. n. (K)	MN725184	MN725385	MN725586
Scutellaria humilis R. Br.	Australia	B. Lepschi & Lally 1717 (WTU)	MN725185	MN725386	MN725587
Scutellaria hypericifolia H. L?v.	China	L. Zhenwen & Zhoujing 1098 (KH)	MN725186	MN725387	MN725588
Scutellaria immaculata Nevski ex Juz.	Uzbekistan	Vatolkina 4757 (K)	MN725187	MN725388	MN725589
Scutellaria incana Spreng.	USA	R. Olmstead 90-10 (WTU)	MN725189	MN725390	MN725591
Scutellaria incana var. *punctata* (Chapm.) Mohr	China	D. E. Boufford & al. 34560 (KUN)	MN725188	MN725389	MN725590
Scutellaria incarnata Vent.	Ecuador	W. Palacios 487 (MO)	MN725190	MN725391	MN725592
Scutellaria incurva Wall.	Thailand	K. Larsen 2368 (K)	MN725191	MN725392	MN725593
Scutellaria indica L.	Taiwan	B. J. Conn 4109 (K)	MN725194	MN725395	MN725596
Scutellaria indica var. *coccinea* S. Kim & S. Lee	Republic of Korea	S. Kim 2015-0298 (SWU)	MN725192	MN725393	MN725594
Scutellaria indica var. *parvifolia* Makino	Arkansas	F. H. Utech M.Ohara 83-446 (KUN)	MN725193	MN725394	MN725595
Scutellaria insignis Nakai	Republic of Korea	S. Kim 2015-0110 (SWU)	MN725195	MN725396	MN725597
Scutellaria integrifolia L.	USA	R. Olmstead 88-05 (WTU)	MN725196	MN725397	MN725598
Scutellaria irrasa Epling	Peru	H. E. Stork & O. B. Horton 9459 (MO)	MN725197	MN725398	MN725599
Scutellaria javanica Jungh.	Sumatra	De Wilde & De Wilde-Duyfjes 14708 (K)	MN725199	MN725400	MN725601
Scutellaria javanica var. *sumatrana* (Miq.) Bock	Malaysia	A. Lamb & A. Surat 375 (K)	MN725198	MN725399	MN725600
Scutellaria knorringiae Juz.	Kyrgyzstan	G.A. Lazkov s. n. (K)	MN725200	MN725401	MN725602
Scutellaria krasevi Kom. & I. Schischk. ex Juz.	Russia	P. G. Gorovoy s. n. (KH)	MN725201	MN725402	MN725603
Scutellaria laeteviolacea Koidz.	Japan	S. Yukawa 1239 (KYO)	MN725203	MN725404	MN725605

Scutellaria laeteviolacea var. *maekawae* (H. Hara) H. Hara	Japan	S. Kinoshita 12337 (KYO)	MN725202	MN725403	MN725604
Scutellaria lanipes Juz.	Kyrgyzstan	G. A. Lazkov s. n. (K)	MN725204	MN725405	MN725606
Scutellaria lateriflora L.	USA	R. Olmstead s. n. (WTU)	MN725205	MN725406	MN725607
Scutellaria leptosiphon Nevski	Tadjikistan	R. Markova & al. s. n. (K)	MN725206	MN725407	MN725608
Scutellaria linearis Benth.	India	R. W. Parker 3396 (K)	MN725207	MN725408	MN725609
Scutellaria litwinowii Bornm. & Sint.	Soviet	M. Popov 233 (PE)	MN725208	MN725409	MN725610
Scutellaria longifolia Benth.	Guatemala	J. J. Castillo & J. M. Vargas 2722 (MO)	MN725209	MN725410	MN725611
Scutellaria longituba Koidz.	USA	R. Olmstead s. n. (WTU)	MN725210	MN725411	MN725612
Scutellaria lundellii Epling	Belize	C. Whitefoord & R. Pacheco 9375 (MO)	MN725211	MN725412	MN725613
Scutellaria lutea J. D. Sm.	Guatemala	M. Perez 1532 (MO)	MN725212	MN725413	MN725614
Scutellaria mesostegia Juz.	Kyrgyzstan	J. Osborne 481 (K)	MN725213	MN725414	MN725615
Scutellaria mexicana (Torr.) A. J. Paton	USA	W. Messinger 257 (WTU)	MN725214	MN725415	MN725616
Scutellaria minor Huds.	Ireland	R. Sealy 1536 (K)	MN725215	MN725416	MN725617
Scutellaria mociniana Benth.	Mexico	F. Ventura 4515 (MO)	MN725216	MN725417	MN725618
Scutellaria mollis R. Br.	Australia	R. G. Coveny 9162 (K)	MN725217	MN725418	MN725619
Scutellaria moniliorhiza Kom.	China	J. Kim & Y. Lee 2019-002 (SWU)	MN725218	MN725419	MN725620
Scutellaria mulleri B. L. Turner	Mexico	J. A. Duke 3969 (MO)	MN725219	MN725420	MN725621
Scutellaria multicaulis Boiss.	Afghanistan	Rechinger 19308 (K)	MN725220	MN725421	MN725622
Scutellaria multiglandulosa (Kearney) Small ex Harper	USA	R. Clindbell 1047 (MO)	MN725221	MN725422	MN725623
Scutellaria muramatsui H. Hara	Japan	S. Tsugaru & T. Sawada 33272 (KYO)	MN725222	MN725423	MN725624
Scutellaria nana A. Gray	USA	R. Olmstead 588 (WTU)	MN725223	MN725424	MN725625
Scutellaria nepetifolia Benth.	Iran	Rechinger 48882 (K)	MN725224	MN725425	MN725626
Scutellaria nervosa Pursh	USA	S. Wagstaff 88-60 (WTU)	MN725225	MN725426	MN725627
Scutellaria nummulariifolia Hook. f.	Argentina	T. M. Pedersen 14491 (MO)	MN725226	MN725427	MN725628
Scutellaria nuristanica Rech. f.	Afghanistan	Podlech 16150 (K)	MN725227	MN725428	MN725629
Scutellaria oblonga Benth.	Ceylon	Wight 2141 (K)	MN725228	MN725429	MN725630
Scutellaria ochotensis Prob.	Russia	S. Joneson 2934 (WTU)	MN725229	MN725430	MN725631
Scutellaria oligodonta Juz.	Kyrgyzstan	S. Landrein 117 (K)	MN725230	MN725431	MN725632
Scutellaria orichalcea J. D. Sm.	Guatemala	R. Avila 3088 (MO)	MN725231	MN725432	MN725633
Scutellaria orientalis L.	Georgia	P. Davis 33797 (K)	MN725239	MN725440	MN725641
Scutellaria orientalis subsp. *bicolor* (Hochst.) J. R. Edm.	Turkey	P. Davis 22037 (K)	MN725232	MN725433	MN725634

Scutellaria orientalis subsp. *bornmuelleri* (Hausskn. ex Bornm.) J. R. Edm.	Turkey	J. M. Watson 1577 (K)	MN725233	MN725434	MN725635
Scutellaria orientalis subsp. *cretacea* (Boiss. & Hausskn.) J. R. Edm.	Iraq	J. B. Gillett 10946 (K)	MN725234	MN725435	MN725636
Scutellaria orientalis subsp. *haussknechtii* (Boiss.) J. R. Edm.	Turkey	P. Sintenis 1134 (K)	MN725235	MN725436	MN725637
Scutellaria orientalis subsp. *macrostegia* (Hausskn. ex Bornm.) J. R. Edm.	Turkey	J. M. Watson 3106 (K)	MN725236	MN725437	MN725638
Scutellaria orientalis subsp. *pectinata* (Montbret & Aucher ex Benth.) J. R. Edm.	Turkey	Hennipman & al. 1560a (K)	MN725237	MN725438	MN725639
Scutellaria orientalis subsp. *virens* (Boiss. & Kotschy) J. R. Edm.	Turkey	Davis & Hedge 29233 (K)	MN725238	MN725439	MN725640
Scutellaria orthocalyx Hand. −Mazz.	Korea	Y. Lee 2019−003 (SWU)	MN725240	MN725441	MN725642
Scutellaria ovata Hill	USA	R. Kral 53065 (MO)	MN725242	MN725443	MN725644
Scutellaria ovata subsp. *bracteata* (Benth.) Epling	USA	F. R. Waller & J. A. Bauml s. n. (MO)	MN725241	MN725442	MN725643
Scutellaria parvula Michx.	USA	S. R. Turner 13−065 (MO)	MN725244	MN725445	MN725646
Scutellaria parvula var. *australis* Fassett	USA	B. Summers & J. Pelton 8520 (MO)	MN725243	MN725444	MN725645
Scutellaria paulsenii Briq.	Tajikistan	O.A. & .B. Fedtschenko s. n. (K)	MN725245	MN725446	MN725647
Scutellaria pekinensis var. *maxima* S. Kim & S. Lee	Korea	S. Kim 2015−0108 (SWU)	MN725246	MN725447	MN725648
Scutellaria pekinensis var. *transitra* (Makino) H. Hara	Korea	S. Kim 2015−0109 (SWU)	MN725247	MN725448	MN725649
Scutellaria pekinensis var. *ussuriensis* (Regel) Hand.−Mazz.	Russia	P. G. Gorovoy s. n. (KH)	MN725248	MN725449	MN725650
Scutellaria petiolata Hemsl. ex Lace & Prain	Pakistan	Crookshank 155 (K)	MN725249	MN725450	MN725651
Scutellaria pinnatifida A. Ham.	Turkey	P. Davis 18802 (K)	MN725253	MN725454	MN725655
Scutellaria pinnatifida subsp. *alpina* (Boiss.) Rech. f.	Persia	K. H. Rechinger 55451 (K)	MN725250	MN725451	MN725652
Scutellaria pinnatifida subsp. *mucida* (Stapf) Rech. f.	Persia	K. H. Rechinger 43494 (K)	MN725251	MN725452	MN725653
Scutellaria pinnatifida subsp. *pichleri* (Stapf) Rech. f.	Iraq	A. Rawi 12254 (K)	MN725252	MN725453	MN725654
Scutellaria platensis Speg.	Argentina	B. Sparre 341 (K)	MN725254	MN725455	MN725656
Scutellaria platystegia Juz.	Armenia	Karajagin 4838 (K)	MN725255	MN725456	MN725657
Scutellaria polyadenia Rech. f.	Zambia	D. B. F F.1999 (K)	MN725256	MN725457	MN725658

Scutellaria pontica K. Koch	Turkey	P. Davis & I. Hedge s. n. (K)	MN725257	MN725458	MN725659
Scutellaria porphyrantha Rech. f.	Turkey	Davis 45660 (K)	MN725258	MN725459	MN725660
Scutellaria potosina Brandegee	Mexico	M. Egger 1140 (WTU)	MN725259	MN725460	MN725661
Scutellaria prostrata Jacquem. ex Benth.	India	F. Billet & J. Leonard 6759 (K)	MN725260	MN725461	MN725662
Scutellaria przewalskii Juz.	Kyrgyzstan	L. Janczenko 5392 (K)	MN725261	MN725462	MN725663
Scutellaria pseudotenax C. Y. Wu	China	M. Fr. Ducloux N6438 (KUN)	MN725262	MN725463	MN725664
Scutellaria purpurascens Sw.	Bolivia	J. S. Solomo 8558 (MO)	MN725263	MN725464	MN725665
Scutellaria quadrilobulata Y. Z. Sun	China	M. Fr. Ducloux M6674 (KUN)	MN725264	MN725465	MN725666
Scutellaria racemosa Pers.	Paraguay	M. Pena-Chocarro & al. 2256 (MO)	MN725265	MN725466	MN725667
Scutellaria raddeana Juz.	Georgia	T. Kurdadze 1117 (K)	MN725266	MN725467	MN725668
Scutellaria ramosissima Popov	Uzbekistan	Kosongitugay s. n. (KH)	MN725267	MN725468	MN725669
Scutellaria regeliana Nakai	Korea	s. n. (KYO)	MN725269	MN725470	MN725671
Scutellaria regeliana var. *ikonnikovii* (Juz.) C. Y. Wu & H. W. Li	China	s. n. (PE)	MN725268	MN725469	MN725670
Scutellaria rehderiana Diels	China	s. n. (KIB)	MN725270	MN725471	MN725672
Scutellaria repens Buch. -Ham. ex D. Don	Nepal	C. S. Chang & al. NE020279 (KH)	MN725271	MN725472	MN725673
Scutellaria resinosa Torr.	USA	T. Wendt 6110 (WTU)	MN725272	MN725473	MN725674
Scutellaria rubicunda Hornem.	Turkey	B. Balansa 499 (K)	MN725273	MN725474	MN725675
Scutellaria rubropunctata Hayata	Japan	s. n. (KYO)	MN725274	MN725475	MN725676
Scutellaria salviifolia Benth.	Turkey	Davis 15903 (K)	MN725275	MN725476	MN725677
Scutellaria sapphirina (Barneby) Olmstead	USA	R. Olmstead 596 (WTU)	MN725276	MN725477	MN725678
Scutellaria saxatilis Riddell	USA	Q. Metheny s. n. (MO)	MN725277	MN725478	MN725679
Scutellaria scandens D. Don	Nepal	A. Schilling 738 (K)	MN725278	MN725479	MN725680
Scutellaria schweinfurthii subsp. *paucifolia* (Baker) A. J. Paton	Zambia	Sanane 844 (K)	MN725279	MN725480	MN725681
Scutellaria scordiifolia Fisch. ex Schrank	Mongolia	C. Jeffey 1427 (K)	MN725280	MN725481	MN725682
Scutellaria scutellarioides (Kunth) Harley	Ecuador	J. A. Hart 1616 (MO)	MN725281	MN725482	MN725683
Scutellaria seleriana Loes.	Mexico	N. M. Melendez 1126 (MO)	MN725282	MN725483	MN725684
Scutellaria serrata Andrews	USA	A. Floden 1578 (MO)	MN725283	MN725484	MN725685
Scutellaria sessilifolia Hemsl.	China	s. n. (KIB)	MN725284	MN725485	MN725686
Scutellaria shikokiana Makino	Japan	K. Kawabata 9158 (KYO)	MN725285	MN725486	MN725687

Scutellaria sibthorpii (Benth.) Hal?csy	Cyprus	Davis 2982 (K)	MN725286	MN725487	MN725688
Scutellaria sieberi Benth.	Greece	J. K. Morton s. n. (K)	MN725287	MN725488	MN725689
Scutellaria sieversii Bunge	North Asia	Karelin & Kiriloff s. n. (NHM)	MN725288	MN725489	MN725690
Scutellaria siphocampyloides Vatke	USA	R. Olmstead 794 (WTU)	MN725289	MN725490	MN725691
Scutellaria sosnowskyi Takht.	Turkey	Mitchell 3205 (K)	MN725290	MN725491	MN725692
Scutellaria splendens Link Klotzsch & Otto	Mexico	F. J. Rendon Sandoval 571(MO)	MN725291	MN725492	MN725693
Scutellaria stocksii Boiss.	Afghanistan	RB & E. Gibbons 242 (K)	MN725292	MN725493	MN725694
Scutellaria strigillosa Hemsl.	Korea	S. Kim 20140705-1 (SWU)	MN725293	MN725494	MN725695
Scutellaria suffrutescens S. Watson	Mexico	J. L. Reveal 3390 (K)	MN725294	MN725495	MN725696
Scutellaria supina L.	Kazakstan	Kotukhov s. n. (K)	MN725295	MN725496	MN725697
Scutellaria tenax W. W. Sm.	China	M. L. Delavay s. n. (KUN)	MN725296	MN725497	MN725698
Scutellaria theobromina Rech. f.	Azerbaijan	K. H. Rechinger 49306 (K)	MN725297	MN725498	MN725699
Scutellaria tomentosa Bertol.	Iran	Rechinger 46863 (K)	MN725298	MN725499	MN725700
Scutellaria tournefortii Benth.	Iran	T. F. Hewer 3967 (K)	MN725299	MN725500	MN725701
Scutellaria tsusimensis H. Hara	Japan	Y. Hassegawa 14590 (KYO)	MN725300	MN725501	MN725702
Scutellaria tuberifera C. Y. Wu & C. Chen	Korea	S. Kim 2019-025 (SWU)	MN725301	MN725502	MN725703
Scutellaria tuberosa Benth.	USA	R. Olmstead 660 (WTU)	MN725302	MN725503	MN725704
Scutellaria tutensis A. Pool	North Asia	S. Timokhina & M. Sakowicz s. n. (NHM)	MN725303	MN725504	MN725705
Scutellaria tuvensis Juz.	Russia	S. Timolchina & M. Sakowicz 6593 (K)	MN725304	MN725505	MN725706
Scutellaria uliginosa A. St. -Hil. ex Benth.	Paraguay	E. M. Zardini & L. Guerrero 56594 (MO)	MN725305	MN725506	MN725707
Scutellaria utriculata Labill.	Lebanon	G. Samulsson 1586 (K)	MN725306	MN725507	MN725708
Scutellaria valdiviana (Clos) Epling	South America	P. Hechenleitner & C. Zamorano 231 (NHM)	MN725307	MN725508	MN725709
Scutellaria velutina Juz. & Vved.	North Asia	N .U. Yberaohe s. n. (NHM)	MN725308	MN725509	MN725710
Scutellaria violascens Gurke	Ethiopia	Friis & al. 12332 (K)	MN725309	MN725510	MN725711
Scutellaria viscidula Bunge	China	N. H. Cowdry 1367 (K)	MN725310	MN725511	MN725712
Scutellaria volubilis Kunth	Ecuador	F. A. Werner & K. Ledesma 1638 (MO)	MN725311	MN725512	MN725713
Scutellaria weishanensis C. Y. Wu & H. W. Li	China	s. n. (KIB)	MN725312	MN725513	MN725714
Scutellaria wrightii A. Gray	USA	J. Bacon 2117A (WTU)	MN725313	MN725514	MN725715
Scutellaria yezoensis Kudo	Russia	S. Gage SG-3021 (WTU)	MN725314	MN725515	MN725716
Scutellaria yunnanensis H. Lev.	China	s. n. (KIB)	MN725315	MN725516	MN725717



Appendix 2. A phylogenetic tree using BI analysis based on three chloroplast genes (*ccsA*, *matK*, and *ndhF*). The number next to each node represents the posterior probability, and the capital letters next to the taxa name represent each species group.



Appendix 2. Continued.

APPENDICES FOR CHAPTER 3

Appendix 3. The list of taxa included in this chapter with voucher information.

Appendix 4. A summary of the sequencing results for the newly completed 49 chloroplast genome in this chapter.

Appendix 5. Comparison of the repeat structure and repeat length for long repeats in the *Scutellaria* chloroplast genomes.

Appendix 6. Comparison of the repeat unit types, sequences, and length for SSRs in the *Scutellaria* chloroplast genomes.

Appendix 7. Statistics of A/T repeat sequences location in *Scutellaria* chloroplast genomes.

Appendix 8. The result of nucleotide diversity analysis of 50 *Scutellaria* chloroplast genomes.

Appendix 9. A super-matrix tree based on Bayesian inference analysis.

Appendix 10. The phylogenetic trees for 52 taxa using complete chloroplast genome sequences, each region and subset with high nucleotide diversity, and the previous three genes.

Appendix 3. The list of taxa included in this chapter with voucher information.

Taxa	Country	Collector and collector number (Herbariumcode)
Holmskiolida sanguinea Retz.	Korea	Y. Lee 2019-0001 (SWU)
Scutellaria albida subsp. *condensata* (Rech.f.) J.R.Edm.	Iraq	s. n. (K)
Scutellaria altamaha Small	USA	R. Kral 51180 (MO)
Scutellaria angustifolia Pursh	s. n.	R. Olmstead 558 (NYBG)
Scutellaria bolanderi A.Gray	USA	R. Olmstead 395 (WTU)
Scutellaria costaricana H. Wendl.	USA	R. Olmstead s.n. (WTU)
Scutellaria dependens Maxim.	Japan	K. Yonekura 12880 (KYO)
Scutellaria discolor Colebr.	Myanmar	J. Murata & al. 25593 (KYO)
Scutellaria drummondii Benth.	USA	J. Bacon 2117 (WTU)
Scutellaria forrestii Diels	China	Liu Zhenwen & Zhoujing 1041 (KH)
Scutellaria franchetiana H.Lev.	China	D. Z. Ji 431 (PE)
Scutellaria gardoquioides (Benth.) Benth.	Peru	E. M. Ortiz V. 1165 (MO)
Scutellaria gaumeri Leonard	Mexico	B. Mogensen 1130 (MO)
Scutellaria guatemalensis Leonard	Mexico	Demetrio Alvarez 11028 (NYBG)
Scutellaria hastifolia L.	Czech Republic	M.Dvorakova, M.Smejkel & J.Vichar 1552 (K)
Scutellaria heterophylla Montbr. & Auch. ex Benth.	Turkey	B.Balansa 498 (K)
Scutellaria heydei Hook. f.	Afghanistan	C. Grey-Wilson & T.F.Hewer 1625 (K)
Scutellaria hintoniana Epling	Mexico	A. de Avila 662 (MO)
Scutellaria hissarica B. Fedtsch.	Tajikistan	L.P.Markova et al. s.n. (K)
Scutellaria immaculata Nevski ex Juz.	Uzbekistan	Vatolkina 4757 (K)
Scutellaria incana Spreng.	USA	R. Olmstead 90-10 (WTU)
Scutellaria indica L.	Taiwan	B.J.Conn 4109 (K)
Scutellaria indica var. *coccinea* S. Kim & S. Lee	Korea	S. Kim 2015-0298 (SWU)
Scutellaria leptosiphon Epling	Tadjikistan	R. Markova et al s.n. (K)
Scutellaria linearis Benth.	India	R.W.Parker 3396 (K)
Scutellaria longifolia Benth.	s. n.	J. J. Castillo & J. M. Vargas 2722 (NYBG)
Scutellaria mexicana (Torr.) A.Paton	USA	W. Messinger 257 (WTU)
Scutellaria mollis R. Br.	Australia	R.G.Coveny 9162 (KEW)
Scutellaria moniliorhiza Kom.	China	J. Kim & Y. Lee 2019-002 (SWU)
Scutellaria multicaulis Boiss.	Afghanistan	Rechinger 19308 (K)
Scutellaria nana A. Gray	USA	M. Bourell 6012 (PE)
Scutellaria nummulariifolia Hook. f.	Argentina	T. M. Pedersen 14491 (MO)
Scutellaria orientalis L.	Georgia	P. Davis 33797 (K)
Scutellaria orthocalyx Hand. -Mazz.	Korea	Y. Lee 2019-003 (SWU)
Scutellaria ovata var. *clacarea* (Epling) Gleason	USA	R. C. Clark 22239 (PE)
Scutellaria parvula Michx.	Lousiana	R. Dale Thomas & Cecil Slaughter 103.883 (NYBG)
Scutellaria pekinensis var. *maxima* S. Kim & S. Lee	Korea	S. Kim 2015-0108 (SWU)
Scutellaria pinnatifida subsp. *mucida* (Stapf) Rech.f.	Persia	K. H. Rechinger 43494 (K)
Scutellaria pinnatifida subsp. *pichleri* (Stapf) Rech.f.	Iraq	A.Rawi 12254 (K)
Scutellaria platensis Speg.	Argentina	B. Sparre 341 (K)
Scutellaria rubicunda Hornem.	Turkey	B.Balansa 499 (K)
Scutellaria scutellarioides (Kunth) Harley	Ecuador	J. A. Hart 1616 (MO)
Scutellaria seleriana Loes.	Mexico	N. M. Melendez 1126 (MO)
Scutellaria siphocampyloides Vatke	USA	R. Olmstead 794 (WTU)
Scutellaria splendens Link, Klotzsch & Otto	Mexico	F. J. Rendon Sandoval 571 (MO)
Scutellaria strigillosa Hemsl.	Korea	S. Kim 20140705-1 (SWU)
Scutellaria suffrutescens S. Watson	Mexico	James L. Reveal 3390 (NYBG)
Scutellaria tomentosa Bertol.	Iran	Rechinger 46863 (K)
Tinnea aethiopica Kotschy ex Hook.f.	South Africa	s. n. (WTU)

Appendix 4. A summary of the sequencing results for the newly completed 49 chloroplast genome in this chapter.

Name	No. of trimed reads	Read length (bp)	GC(%)	No. of reads mapped on cp	% of reads mapped on cp
Holmskioldia sanguinea	32,307,172	100	36.79	3,010,936	9.3
S. albida subsp. *condensata*	14,422,636	100	44.72	48,507	0.3
S. altamaha	5,004,495	150	34.00	86,563	1.7
S. angustifolia	4,302,420	150	31.00	255,241	5.9
S. bolanderi	50,807,914	100	31.79	1,792,198	3.5
S. costaricana	50,329,532	100	36.69	904,084	1.8
S. dependens	3,450,322	150	37.00	56,827	1.6
S. discolor	29,831,260	100	45.69	528,354	1.8
S. drummondii	15,300,750	100	42.50	74,038	0.5
S. forrestii	40,433,460	100	33.39	866,575	2.1
S. franchetiana	46,873,084	100	38.33	1,139,961	2.4
S. gardoquioides	31,081,324	100	49.90	424,048	1.4
S. gaumeri	25,070,028	100	45.89	71,074	0.3
S. guatemalensis	4,375,716	150	34.00	186,767	4.3
S. hastifolia	24,840,820	100	32.02	1,310,687	5.3
S. heterophylla	17,278,266	100	62.35	158,498	0.9
S. heydei	37,024,946	100	45.08	332,320	0.9
S. hintoniana	43,914,160	100	34.63	3,555,126	8.1
S. hissarica	36,661,490	100	35.76	3,557,969	9.7
S. immaculata	41,749,692	100	34.04	1,565,698	3.8
S. incana	30,017,284	100	35.22	169,587	0.6
S. indica	39,800,504	100	34.06	1,519,921	3.8
S. indica var. *coccinea*	34,772,388	100	35.84	1,580,323	4.5
S. leptosiphon	54,878,596	100	32.72	1,592,449	2.9
S. linearis	22,201,132	100	38.02	1,497,428	6.7
S. longifolia	3,917,749	150	34.00	214,615	5.5
S. mexicana	16,826,426	100	45.29	123,076	0.7
S. mollis	19,489,250	100	36.92	365,441	1.9
S. moniliorhiza	4,910,615	150	36.50	120,846	2.5
S. multicaulis	31,178,356	100	33.43	867,050	2.8
S. nana	4,841,433	150	32.00	181,694	3.8
S. nummulariifolia	41,291,772	100	32.70	3,384,704	8.2
S. orientalis	49,132,400	100	34.20	2,785,671	5.7
S. orthocalyx	4,276,624	150	44.00	81,397	1.9
S. ovata	5,699,904	150	33.50	37,034	0.6
S. parvula	4,338,005	150	33.00	262,131	6.0
S. pekinensis var. *maxima*	3,439,253	150	37.00	32,248	0.9
S. pinnatifida subsp. *mucida*	3,269,386	150	33.00	72,718	2.2
S. pinnatifida subsp. *pichleri*	5,129,405	150	34.00	403,475	7.9
S. platensis	3,679,519	150	33.50	72,350	2.0
S. rubicunda	32,164,034	100	35.39	4,665,203	14.5
S. scutellarioides	28,599,490	100	39.01	6,004,992	21.0
S. seleriana	45,752,354	100	35.17	1,996,284	4.4
S. siphocampyloides	16,479,792	100	39.44	142,498	0.9
S. splendens	4,890,687	150	35.00	430,630	8.8
S. strigillosa	2,563,825	150	36.50	130,270	5.1
S. suffrutescens	6,536,592	150	37.50	151,450	2.3
S. tomentosa	45,804,286	100	33.48	2,344,793	5.1
Tinnea aethiopica	3,392,319	150	34.00	250,958	7.4

Appendix 5. Comparison of the repeat structure and repeat length for long repeats in the *Scutellaria* chloroplast genomes.

	Repeat Structure					Repeat Length (bp)																
	Forward	Palindromic	Reverse	Complement	Total	30	31	32	33	34	35	36	37	38	39	40	41	42	44	45	46	48	49
S. albida subsp. *condensata*	9	10	2	0	21	7	2	0	1	2	0	0	0	0	2	0	1	0	0	1	0	2	1
S. altamaha	9	12	0	0	21	8	0	1	1	2	0	0	0	1	2	0	0	1	1	1	0	0	1
S. angustifolia	8	11	0	0	19	6	0	1	1	1	0	1	0	0	3	0	1	0	1	1	0	0	1
S. baicalensis	18	10	0	0	28	8	0	1	0	2	1	0	0	0	4	0	1	0	0	0	0	0	1
S. bolanderi	8	11	0	0	19	6	0	1	1	1	0	1	1	0	2	0	1	0	1	1	0	0	1
S. costaricana	10	9	0	0	19	7	1	2	1	1	0	0	0	0	2	0	1	0	0	0	0	0	1
S. dependens	12	14	1	0	27	7	1	2	3	1	0	0	0	0	2	0	1	0	1	1	0	0	1
S. discolor	8	9	0	0	17	7	0	1	1	1	1	0	0	0	2	0	1	0	1	1	0	0	0
S. drummondii	10	12	0	0	22	7	0	1	2	1	0	1	0	0	2	1	1	0	0	1	2	0	1
S. forrestii	9	10	0	0	19	9	0	1	0	2	1	0	0	0	2	0	1	0	1	0	0	0	1
S. franchetiana	11	10	0	0	21	7	0	1	1	1	0	0	0	0	3	0	1	0	1	1	0	0	0
S. gardoquioides	8	13	0	0	21	8	0	1	1	1	0	0	0	1	2	1	1	0	2	0	0	0	1
S. gaumeri	8	9	0	0	17	6	0	1	1	1	0	1	0	0	2	0	1	0	1	0	0	0	1
S. guatemalensis	8	9	0	0	17	6	0	1	1	1	0	1	0	0	2	0	1	0	1	0	0	0	1
S. hastifolia	8	10	0	0	18	6	0	1	1	1	0	0	0	0	2	0	1	0	1	1	0	0	1
S. heterophylla	13	11	0	1	25	9	2	1	1	2	0	0	0	0	3	0	0	1	0	1	0	0	1
S. heydei	11	10	0	0	21	9	0	1	0	2	1	0	1	0	2	0	1	0	0	1	0	0	1
S. hintoniana	9	9	1	0	19	7	0	2	0	1	0	1	1	0	2	0	1	0	1	0	0	0	0
S. hissarica	8	9	0	0	17	4	0	1	1	2	0	0	0	0	2	0	1	0	0	1	0	1	1
S. immaculata	14	10	0	0	24	9	0	1	0	2	1	0	1	0	3	0	1	0	0	1	0	0	1
S. incana	9	11	0	0	20	7	0	1	1	2	0	0	0	1	2	0	0	1	1	1	0	0	1
S. indica	8	9	1	0	18	6	0	2	1	1	1	0	0	0	2	0	1	0	1	1	0	0	0
S. indica var. *coccinea*	9	10	0	0	19	7	0	1	1	1	1	1	0	0	2	0	1	0	1	1	0	0	0
S. insignis	9	9	0	0	18	5	0	1	1	1	0	0	0	0	2	1	1	0	1	1	0	0	0
S. lateriflora	9	14	0	0	23	8	0	2	2	1	0	0	0	1	2	0	1	0	2	1	0	0	1
S. leptosiphon	9	8	0	0	17	7	0	1	0	2	0	0	0	0	2	0	1	0	0	1	0	0	1
S. linearis	12	8	0	0	20	8	0	2	0	1	0	0	0	0	3	0	2	0	0	0	0	0	1
S. longifolia	8	10	0	0	18	8	0	1	0	1	0	1	0	0	3	0	1	0	0	0	0	0	1
S. mexicana	9	11	0	0	20	6	0	1	1	1	0	1	0	0	2	0	1	0	1	1	0	0	1
S. mollis	12	9	0	0	21	8	0	1	1	1	0	1	0	0	2	0	1	0	1	1	0	0	1
S. moniliorhiza	11	11	0	0	22	8	0	1	1	2	0	0	0	0	2	0	1	0	0	1	0	0	1
S. multicaulis	14	10	0	0	24	12	0	1	0	2	1	0	0	0	2	0	1	0	0	1	0	0	1
S. nana	8	11	0	0	19	12	0	1	1	1	0	1	0	0	3	0	1	0	1	1	0	0	1
S. nummulariifolia	12	11	0	0	23	8	1	1	1	1	0	1	0	0	3	0	1	0	1	1	0	0	1
S. orientalis	13	9	0	0	22	8	0	1	0	2	1	0	0	0	3	0	1	0	0	0	1	0	1
S. orthocalyx	13	10	1	0	24	9	0	1	0	2	1	0	0	0	4	0	1	0	0	0	0	0	1
S. ovata	9	11	1	0	21	6	1	1	1	2	0	1	0	1	2	0	0	1	1	1	0	0	1
S. parvula	8	13	0	0	21	8	0	1	1	0	0	1	0	0	3	0	2	0	1	1	0	0	1
S. pekinensis var. *maxima*	8	10	0	0	18	6	0	1	1	1	0	0	0	0	2	1	1	0	1	1	0	0	0
S. pinnatifida subsp. *mucida*	10	9	1	0	20	8	1	1	0	2	0	0	1	0	2	0	2	0	0	0	0	0	1
S. pinnatifida subsp. *pichleri*	9	8	1	0	18	7	1	1	0	2	0	0	0	0	2	0	2	0	0	0	0	0	1
S. platensis	8	10	0	0	18	8	0	1	1	1	0	0	0	0	2	0	1	0	0	0	0	0	2
S. rubicunda	9	11	0	0	20	6	1	0	1	3	0	1	0	0	2	0	1	0	0	1	0	1	1
S. scutellarioides	15	12	0	0	27	7	2	1	1	1	0	0	0	0	2	0	1	0	2	0	0	0	1
S. seleriana	9	9	0	0	18	8	0	1	1	1	0	1	0	0	3	0	1	0	0	0	0	0	0
S. siphocampyloides	11	11	0	0	22	7	0	1	1	1	0	0	1	0	2	0	2	0	1	1	0	0	1
S. splendens	8	10	0	0	18	6	0	1	1	1	1	0	1	0	2	0	1	0	0	0	0	0	1
S. strigillosa	9	11	0	0	20	8	0	1	1	2	0	0	0	0	2	0	1	0	0	1	0	0	1
S. suffrutescens	8	12	0	0	20	7	0	1	1	1	1	0	0	0	3	0	1	0	0	1	1	0	1
S. tomentosa	10	9	1	0	20	9	0	1	0	2	0	0	1	1	2	0	2	0	0	0	0	0	1

Appendix 5. Continued.

50	52	53	54	56	59	60	61	63	65	67	70	72	73	76	78	80	86	88	91	92	94	97	99	101	104	109	110	113	115	117	119	120	124	125	127	128	129	130	215
0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	1	0	0	1	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0	0	1	0
0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
0	1	0	1	0	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0
0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
0	1	0	1	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Appendix 6. Comparison of the repeat unit types, sequences, and length for SSRs in the *S.* chloroplast genomes.

Taxa	Types of repeat unit						Sequence and length of repeats																														
	Mono	Di	Tri	Tetra	Penta	Hexa	(A)ˆ1(	(A)ˆ11	(A)ˆ12	(A)ˆ1(	(A)ˆ14	(A)ˆ15	(A)ˆ17	(A)ˆ1ᵹ	(AAAAG)ˆ3	(AAAAG)ˆ4	(AAATAG)ˆ3	(AAG)ˆ4	(AAGA)ˆ3	(AAT)ˆ4	(AATA)ˆ3	(AATGGA)ˆ3	(ACCA)ˆ3	(AT)ˆ5	(AT)ˆ6	(AT)ˆ7	(AT)ˆ8	(AT)ˆ4	(ATA)ˆ5	(ATA)ˆ3	(ATAA)ˆ3	(ATCTT)ˆ3	(ATTT)(C)ˆ10	(C)ˆ11		
S. albida subsp. *condensata*	22	5	3	8	0	0	4	2	1	0	1	1	0	0	0	0	0	1	0	0	2	0	0	2	0	0	0	0	1	0	0	0	1	0		
S. altamaha	23	4	4	5	0	0	4	2	2	1	0	0	0	0	0	0	0	1	0	0	1	0	1	1	1	0	0	2	0	0	0	0	0	0		
S. angustifolia	20	2	3	6	0	0	2	2	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	2	0	0	0	1	0	0	0	0	0	0		
S. baicalensis	23	3	3	5	0	1	5	1	0	0	0	0	0	0	0	1	0	0	1	1	0	1	1	1	2	0	0	1	0	0	0	0	0	0		
S. bolanderi	17	5	3	6	0	2	2	0	0	1	0	0	0	0	0	0	0	1	0	0	1	1	1	2	0	0	0	1	0	0	0	0	0	0		
S. costaricana	24	3	4	7	2	0	4	3	0	0	0	0	0	0	1	0	1	0	1	1	0	1	1	0	0	0	1	0	0	0	0	1	0			
S. dependens	20	5	3	5	0	0	6	1	0	0	0	0	0	0	0	0	1	0	1	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0		
S. discolor	20	4	3	6	0	0	5	0	1	1	0	0	0	0	0	0	1	0	0	1	0	1	1	1	0	0	1	0	0	0	0	0	0	0		
S. drummondii	26	3	4	6	0	0	7	1	1	0	0	0	0	0	0	0	1	0	0	1	0	1	1	1	0	0	2	0	1	0	0	0	0	0		
S. forrestii	23	5	4	5	0	0	1	3	0	0	0	0	0	0	0	0	1	0	1	1	0	1	2	0	0	1	1	0	0	0	0	0	0	0		
S. franchetiana	23	4	3	5	0	0	4	4	0	0	1	0	0	0	0	0	1	0	0	1	0	1	1	1	0	0	1	0	0	0	0	0	0	1		
S. gardoquioides	29	3	4	5	0	0	7	4	1	2	0	0	0	0	0	0	1	0	1	1	0	0	1	0	0	0	1	0	0	0	0	0	1	0		
S. gaumeri	20	3	3	5	1	0	4	0	0	1	0	0	0	0	1	0	0	1	0	0	0	0	1	2	0	0	0	1	0	0	0	0	0	0		
S. guatemalensis	23	2	4	5	2	0	5	1	0	1	1	0	0	0	1	0	0	1	0	1	0	0	1	1	1	0	0	1	0	0	0	0	0	2		
S. hastifolia	19	4	3	5	0	0	6	1	1	0	1	0	0	0	0	0	1	0	0	1	0	0	1	1	0	0	1	0	0	0	0	0	0	0		
S. heterophylla	22	5	4	5	0	0	3	3	0	0	0	0	0	0	0	0	1	0	1	1	0	0	3	0	0	0	1	0	0	0	1	0	0	0		
S. heydei	20	4	4	6	0	0	0	1	0	0	0	0	0	0	0	0	1	0	1	1	0	1	1	0	1	0	1	0	0	0	0	0	2	1		
S. hintoniana	28	5	4	4	0	0	8	0	0	1	0	0	0	0	0	0	1	0	1	1	0	0	2	0	0	0	1	0	0	0	0	0	1	0		
S. hissarica	29	5	3	5	0	0	7	1	1	1	1	1	0	1	0	0	1	0	0	1	0	0	0	2	1	0	0	0	1	0	0	0	0	0		
S. immaculata	17	4	4	5	0	0	2	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	1	0	1	0	1	0	0	0	0	0	1	1		
S. incana	24	4	4	5	0	0	7	2	1	1	0	0	0	0	0	0	1	0	0	1	0	1	1	1	0	0	2	0	0	0	0	0	0	0		
S. indica	22	4	3	6	0	0	5	1	0	0	0	1	0	0	0	0	1	0	0	1	0	1	2	0	0	0	1	0	0	0	0	0	1	0		
S. indica var. *coccinea*	25	4	3	6	0	0	9	0	0	0	2	0	0	0	0	0	1	0	0	1	0	1	1	1	0	0	1	0	0	0	0	0	1	0		
S. insignis	22	5	3	6	0	0	4	1	0	1	1	1	0	0	0	0	1	0	0	1	0	1	2	1	0	0	1	0	0	0	0	0	1	0		
S. lateriflora	22	3	3	5	0	0	7	1	0	1	0	0	1	0	0	0	1	0	0	1	1	0	0	1	1	0	0	0	0	0	0	0	0	1		
S. leptosiphon	14	3	4	5	0	0	0	1	0	0	0	0	0	0	0	0	1	0	1	1	0	1	1	0	1	0	1	0	0	0	0	0	0	0		
S. linearis	16	3	4	6	0	0	2	1	0	0	0	0	0	0	0	0	1	0	1	1	0	1	1	1	0	0	1	0	0	0	0	0	1	0		
S. longifolia	25	5	4	6	2	0	6	1	0	0	0	0	0	0	1	0	0	1	0	1	1	0	1	1	0	1	0	1	0	0	0	0	1	0		
S. mexicana	26	3	3	6	0	0	6	3	1	0	0	0	0	0	0	0	1	0	1	0	1	0	1	1	0	0	1	0	0	0	0	0	1	0		
S. mollis	16	2	2	7	1	0	2	0	1	0	0	0	0	0	0	0	1	1	0	1	0	1	1	0	0	0	1	0	0	1	0	0	0	0		
S. moniliorhiza	20	4	3	5	1	0	5	3	0	0	0	0	0	0	0	0	1	0	0	1	0	0	1	1	0	0	1	0	0	0	0	0	1	1		
S. multicaulis	16	3	4	6	1	0	1	0	1	0	0	0	0	0	0	0	1	0	1	1	0	1	1	1	0	0	1	0	0	0	0	0	1	0		
S. nana	21	4	3	6	0	0	3	1	1	0	0	0	0	0	0	0	1	0	0	1	0	1	2	0	0	0	1	0	0	0	0	0	1	0		
S. nummulariifolia	19	2	1	6	0	0	5	1	0	0	0	0	0	0	0	0	1	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	1	0		
S. orientalis	17	2	4	8	0	0	3	0	0	0	0	0	0	0	0	0	1	0	1	1	0	1	1	0	0	0	1	0	0	0	0	0	0	1		
S. orthocalyx	24	3	2	5	0	0	6	0	0	0	0	1	0	0	0	0	0	0	1	1	0	1	2	0	1	0	1	0	0	0	0	0	0	0		
S. ovata	23	4	4	5	0	0	7	2	1	1	0	0	0	0	0	0	1	0	0	1	0	1	1	1	0	0	2	0	0	0	0	0	0	0		
S. parvula	22	3	3	6	0	0	3	2	0	1	0	0	0	0	0	0	1	0	0	1	0	1	2	0	0	0	1	0	0	0	0	0	0	0		
S. pekinensis var. *maxima*	19	6	4	5	0	0	3	1	2	0	0	0	0	0	0	0	1	0	0	1	0	1	2	1	0	0	1	0	0	0	0	0	1	1		
S. pinnatifida subsp. *mucida*	19	3	4	6	0	0	4	0	0	0	0	0	0	0	0	0	1	0	1	1	0	1	1	0	1	0	1	0	0	0	0	0	1	2		
S. pinnatifida subsp. *pichleri*	18	4	4	6	0	0	3	0	0	0	0	0	0	0	0	0	1	0	1	1	0	1	2	0	0	0	1	0	0	0	0	0	1	0		
S. platensis	23	3	2	5	0	0	2	3	1	0	0	0	0	0	0	0	1	0	0	1	0	1	2	1	0	0	1	0	0	0	0	0	0	1		
S. rubicunda	26	4	3	7	0	0	5	4	0	0	1	1	0	0	0	0	1	0	0	2	0	0	2	0	0	0	0	1	0	0	0	0	0	0		
S. scutellarioides	22	4	4	5	0	0	7	0	1	0	0	0	0	0	0	0	1	0	1	1	0	0	2	0	0	0	1	0	0	0	0	0	1	0		
S. seleriana	22	4	4	6	0	0	4	0	0	0	0	0	0	0	0	0	1	0	1	1	0	1	1	0	0	0	1	0	0	0	0	0	0	0		
S. siphocampyloides	17	3	3	6	0	2	1	0	0	1	0	0	0	0	0	0	1	0	0	1	1	1	2	0	0	0	1	0	0	0	0	0	1	0		
S. splendens	27	3	4	5	2	0	8	1	0	0	0	0	0	1	0	0	1	0	1	1	0	1	1	0	0	0	1	0	0	0	0	1	0			
S. strigillosa	19	4	3	5	1	0	5	1	1	1	0	0	0	0	0	0	1	0	0	1	0	0	1	1	0	0	1	0	0	0	0	0	0	2		
S. suffrutescens	25	3	5	5	0	0	6	0	1	0	0	0	0	0	0	0	1	0	0	1	0	1	1	1	0	0	3	0	1	0	0	1	0			
S. tomentosa	17	4	4	6	0	0	3	0	0	0	0	0	0	0	0	0	1	0	1	1	0	1	2	0	0	0	1	0	0	0	0	0	1			

Appendix 6. Continued.

(C)^12	(C)^13	(C)^14	(C)^15	(CTTAT)^3	(CTTTT)^3	(CTTTT)^4	(G)^10	(G)^11	(G)^12	(G)^13	(G)^14	(G)^15	(G)^16	(GTCT)^3	(GTT)^4	(T)^10	(T)^11	(T)^12	(T)^13	(T)^14	(T)^15	(T)^16	(T)^17	(T)^18	(TA)^5	(TA)^6	(TA)^7	(TA)^8	(TAAA)^3	(TCTA)^3	(TTAA)^3	(TTC)^4	(TTCCT)^3	(TTCT)^3	(TTTA)^3	(TTTC)^3	(TTTC)^4	(TTTCCA)^3
0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	6	1	0	1	1	1	0	0	1	1	1	0	1	1	1	1	0	0	1	0	1	0
1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	7	2	2	1	0	0	0	0	0	1	0	0	1	0	1	0	1	0	0	0	0	1	0
0	0	0	1	0	0	0	1	0	0	0	0	1	0	1	0	8	4	0	1	0	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	1	0
1	1	1	0	0	0	0	0	0	0	1	1	0	0	1	0	7	2	2	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0
0	1	0	0	0	0	0	0	0	0	1	0	0	0	1	0	9	2	0	1	0	0	0	0	0	3	0	0	0	1	1	0	1	0	0	0	0	1	1
0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	7	6	1	1	1	0	0	0	0	1	1	0	0	1	1	1	1	0	0	0	0	1	0
1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	5	2	1	1	2	0	0	0	0	3	0	0	0	1	1	0	1	0	0	0	0	1	0
2	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0	6	1	1	1	0	0	0	0	0	1	1	0	0	1	1	0	1	0	0	0	0	1	0
1	0	1	0	0	0	0	0	0	1	0	0	0	0	1	0	9	2	2	1	0	0	0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	1	0
2	1	0	0	0	0	0	1	0	2	0	0	0	0	1	0	5	1	3	2	2	0	0	0	0	1	0	0	1	0	0	1	1	0	0	0	1	0	0
0	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	5	2	0	1	2	1	0	0	0	0	2	0	0	0	1	0	1	0	0	0	0	1	0
0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	6	2	2	0	0	2	0	0	1	1	1	0	0	0	1	1	1	0	1	1	0	0	0
0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	9	1	1	1	1	1	0	0	0	1	0	0	0	0	1	1	1	0	0	0	0	1	0
0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	5	2	2	2	0	1	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	1	0
0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	6	1	1	0	0	1	0	0	0	2	0	0	0	1	1	0	1	0	0	0	0	1	0
0	0	1	0	0	0	0	0	0	1	0	0	0	0	1	0	8	4	0	2	0	0	0	0	0	1	1	0	0	0	0	1	1	0	0	0	1	0	0
0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	6	7	2	0	0	0	0	0	0	1	1	0	0	0	0	2	1	0	0	0	1	0	0
0	0	1	0	0	0	0	2	0	0	0	1	0	0	1	0	8	4	0	1	1	0	0	0	0	1	2	0	0	0	0	1	1	0	0	0	0	1	0
0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	10	3	0	1	1	0	0	0	0	1	1	0	0	1	1	0	1	0	0	0	0	1	0
1	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	6	4	0	0	0	0	1	0	0	1	1	0	0	0	0	2	1	0	0	0	1	0	0
0	0	1	0	0	0	0	0	0	0	0	1	0	0	1	0	8	2	0	1	0	0	0	0	0	1	0	0	1	0	1	0	1	0	0	0	0	1	0
0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	6	4	2	0	0	0	0	0	0	1	0	1	0	1	1	0	1	0	0	0	1	0	0
0	1	0	0	0	0	0	0	0	0	1	0	0	0	1	0	6	2	2	1	0	0	0	0	0	1	0	1	0	1	1	0	1	0	0	0	1	0	0
0	0	1	0	0	0	0	0	0	0	0	1	0	0	1	0	5	2	0	2	0	0	1	1	0	1	1	0	0	1	1	0	1	0	0	0	0	1	0
0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	4	1	2	1	2	0	0	0	0	1	0	0	0	1	1	0	1	0	0	0	0	1	0
0	0	1	0	0	0	0	1	0	0	0	1	0	0	1	0	7	1	1	1	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	1	0	0
0	1	0	0	0	0	0	0	0	1	0	0	0	0	1	0	8	1	1	0	0	0	0	0	1	0	0	0	0	0	2	1	0	0	0	0	1	0	0
0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	8	6	2	0	1	0	0	0	0	2	0	1	0	0	1	1	1	0	0	0	0	1	0
1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	7	3	1	1	0	0	0	1	0	2	0	0	0	1	0	1	1	0	0	0	0	1	0
0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	6	2	4	0	0	0	0	0	0	0	1	0	0	1	1	0	0	0	0	0	0	1	0
0	0	0	0	1	0	0	0	1	0	0	0	0	0	1	0	4	3	1	1	0	0	0	0	0	2	0	0	0	1	1	0	1	0	0	0	0	1	0
0	0	0	0	0	0	0	1	0	0	0	0	1	0	1	0	7	1	1	1	1	0	0	0	0	1	0	0	0	0	0	2	1	1	0	0	1	0	0
0	0	1	0	0	0	0	0	0	0	0	1	0	0	1	0	5	6	2	0	0	0	0	0	0	2	0	0	0	1	1	0	1	0	0	0	1	0	0
1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	5	1	1	1	1	1	0	0	0	1	0	0	0	1	1	0	0	0	0	0	1	0	0
1	0	0	0	0	0	0	1	0	1	0	0	0	0	1	0	7	2	0	0	1	0	0	0	0	1	0	0	0	1	0	2	1	0	0	0	2	0	0
1	0	0	1	0	0	0	0	0	0	1	0	1	0	1	0	3	9	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0
1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	4	5	0	1	0	0	0	0	0	1	0	0	0	1	0	1	1	0	0	0	0	1	0
1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	6	5	0	2	1	0	0	0	0	0	0	1	0	1	1	0	1	0	0	0	0	1	0
0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	6	1	0	0	0	2	1	0	0	2	1	0	0	1	1	0	1	0	0	0	0	1	0
1	0	0	1	0	0	0	0	0	1	0	0	0	0	1	0	5	2	0	1	1	0	0	0	0	0	1	0	0	0	0	2	1	0	0	0	1	0	0
2	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	8	2	0	0	0	0	0	0	0	2	0	0	0	0	2	1	0	0	0	1	0	0
0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	8	4	1	0	2	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	7	2	1	1	2	1	0	0	0	2	0	0	0	0	1	1	1	0	0	1	0	1	0
0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	6	4	1	1	0	0	0	0	0	1	1	0	0	1	0	1	1	0	1	1	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	10	7	1	0	0	0	0	0	0	2	0	1	0	0	1	1	1	0	0	0	0	1	0
1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	8	3	0	0	0	1	0	0	0	1	0	0	0	1	1	0	1	0	0	0	0	1	1
0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	7	8	0	1	0	1	0	0	0	1	1	0	0	0	1	1	1	0	0	0	0	1	0
0	0	0	0	1	0	0	0	1	0	0	0	0	0	1	0	4	1	2	1	0	0	0	0	0	2	0	0	0	1	1	0	1	0	0	0	0	1	0
1	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	6	3	2	1	1	1	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	1	0
0	0	0	0	0	0	0	1	0	0	0	1	0	0	1	0	7	3	0	1	0	0	0	0	0	1	0	1	0	0	0	2	1	0	0	0	1	0	0

Appendix 7. Summary of A/T repeat sequences location in *Scutellaria* chloroplast genomes.

ID	SSR	start	end	length	region	Location
S. orientalis	(A)10	8609	8618	10	non-coding	LSC
S. orientalis	(T)14	13429	13442	14	non-coding	LSC
S. orientalis	(T)11	19010	19020	11	rpoC2	LSC
S. orientalis	(T)10	32163	32172	10	non-coding	LSC
S. orientalis	(T)11	43862	43872	11	ycf3	LSC
S. orientalis	(A)10	45071	45080	10	non-coding	LSC
S. orientalis	(T)10	64438	64447	10	non-coding	LSC
S. orientalis	(T)10	70095	70104	10	non-coding	LSC
S. orientalis	(T)10	75690	75699	10	petB	LSC
S. orientalis	(T)10	76094	76103	10	petB	LSC
S. orientalis	(T)10	80981	80990	10	non-coding	LSC
S. orientalis	(T)10	102546	102555	10	trnI-GAU	IRB
S. orientalis	(A)10	112510	112519	10	non-coding	SSC
S. orientalis	(A)10	134002	134011	10	trnI-GAU	IRA
S. hastifolia	(A)10	5101	5110	10	non-coding	LSC
S. hastifolia	(A)10	8553	8562	10	non-coding	LSC
S. hastifolia	(T)10	13238	13247	10	non-coding	LSC
S. hastifolia	(T)11	18842	18852	11	rpoC2	LSC
S. hastifolia	(A)12	27589	27600	12	non-coding	LSC
S. hastifolia	(T)12	29595	29606	12	non-coding	LSC
S. hastifolia	(T)10	35771	35780	10	non-coding	LSC
S. hastifolia	(A)14	42584	42597	14	non-coding	LSC
S. hastifolia	(A)11	44846	44856	11	non-coding	LSC
S. hastifolia	(A)10	47410	47419	10	non-coding	LSC
S. hastifolia	(T)10	64192	64201	10	non-coding	LSC
S. hastifolia	(T)15	69540	69554	15	non-coding	LSC
S. hastifolia	(A)10	71632	71641	10	clpP	LSC
S. hastifolia	(T)10	75541	75550	10	petB	LSC
S. hastifolia	(T)10	78454	78463	10	rpoA	LSC
S. hastifolia	(T)10	80350	80359	10	non-coding	LSC
S. hastifolia	(A)10	112614	112623	10	non-coding	SSC
S. hastifolia	(A)10	112966	112975	10	non-coding	SSC
S. pinnatifida subsp. *pichleri*	(A)10	4472	4481	10	non-coding	LSC
S. pinnatifida subsp. *pichleri*	(T)10	5943	5952	10	rps16	LSC
S. pinnatifida subsp. *pichleri*	(A)10	8221	8230	10	non-coding	LSC
S. pinnatifida subsp. *pichleri*	(T)10	12404	12413	10	atpF	LSC
S. pinnatifida subsp. *pichleri*	(A)10	16390	16399	10	non-coding	LSC
S. pinnatifida subsp. *pichleri*	(T)11	18635	18645	11	rpoC2	LSC
S. pinnatifida subsp. *pichleri*	(T)10	31736	31745	10	non-coding	LSC
S. pinnatifida subsp. *pichleri*	(T)10	63508	63517	10	non-coding	LSC
S. pinnatifida subsp. *pichleri*	(T)10	75093	75102	10	petB	LSC
S. pinnatifida subsp. *pichleri*	(T)10	75501	75510	10	petB	LSC
S. pinnatifida subsp. *pichleri*	(T)11	77983	77993	11	rpoA	LSC
S. pinnatifida subsp. *pichleri*	(T)10	80395	80404	10	non-coding	LSC
S. pinnatifida subsp. *pichleri*	(T)10	101951	101960	10	trnI-GAU	IRB
S. pinnatifida subsp. *pichleri*	(A)10	133386	133395	10	trnI-GAU	IRA
S. multicaulis	(T)10	5553	5562	10	rps16	LSC
S. multicaulis	(T)10	5942	5951	10	rps16	LSC
S. multicaulis	(T)10	12792	12801	10	atpF	LSC
S. multicaulis	(T)13	13442	13454	13	non-coding	LSC
S. multicaulis	(T)11	19028	19038	11	rpoC2	LSC
S. multicaulis	(A)12	27773	27784	12	non-coding	LSC
S. multicaulis	(T)10	32205	32214	10	non-coding	LSC
S. multicaulis	(T)12	64440	64451	12	non-coding	LSC
S. multicaulis	(T)10	75702	75711	10	petB	LSC
S. multicaulis	(T)14	76110	76123	14	petB	LSC
S. multicaulis	(T)10	78597	78606	10	rpoA	LSC
S. multicaulis	(T)10	102570	102579	10	trnI-GAU	IRB
S. multicaulis	(A)10	112538	112547	10	non-coding	SSC
S. multicaulis	(A)10	134061	134070	10	trnI-GAU	IRA
S. albida subsp. *condensata*	(A)11	4426	4436	11	non-coding	LSC
S. albida subsp. *condensata*	(A)10	5023	5032	10	non-coding	LSC
S. albida subsp. *condensata*	(T)11	8423	8433	11	non-coding	LSC
S. albida subsp. *condensata*	(A)11	12091	12101	11	non-coding	LSC
S. albida subsp. *condensata*	(T)16	13456	13471	16	non-coding	LSC
S. albida subsp. *condensata*	(T)11	19057	19067	11	rpoC2	LSC
S. albida subsp. *condensata*	(A)12	42826	42837	12	non-coding	LSC
S. albida subsp. *condensata*	(T)11	43761	43771	11	ycf3	LSC
S. albida subsp. *condensata*	(T)15	43890	43904	15	ycf3	LSC
S. albida subsp. *condensata*	(A)14	44842	44855	14	ycf3	LSC
S. albida subsp. *condensata*	(A)10	45100	45109	10	non-coding	LSC
S. albida subsp. *condensata*	(T)11	45975	45985	11	non-coding	LSC
S. albida subsp. *condensata*	(A)10	47634	47643	10	non-coding	LSC

S. albida subsp. *condensata*	(T)11	69219	69229	11	non-coding	LSC
S. albida subsp. *condensata*	(A)10	74152	74161	10	non-coding	LSC
S. albida subsp. *condensata*	(T)10	78100	78109	10	rpoA	LSC
S. albida subsp. *condensata*	(T)14	80517	80530	14	non-coding	LSC
S. albida subsp. *condensata*	(T)11	83977	83987	11	non-coding	IRB
S. albida subsp. *condensata*	(T)12	112595	112606	12	non-coding	SSC
S. albida subsp. *condensata*	(A)15	115406	115420	15	non-coding	SSC
S. albida subsp. *condensata*	(A)11	151844	151854	11	non-coding	IRA
S. heterophylla	(A)11	5585	5595	11	rps16	LSC
S. heterophylla	(T)10	5960	5969	10	rps16	LSC
S. heterophylla	(T)10	8383	8392	10	non-coding	LSC
S. heterophylla	(A)11	8618	8628	11	non-coding	LSC
S. heterophylla	(T)10	8982	8991	10	non-coding	LSC
S. heterophylla	(T)10	12773	12782	10	atpF	LSC
S. heterophylla	(T)13	13423	13435	13	non-coding	LSC
S. heterophylla	(A)10	16762	16771	10	non-coding	LSC
S. heterophylla	(T)10	17065	17074	10	rpoC2	LSC
S. heterophylla	(T)11	18995	19005	11	rpoC2	LSC
S. heterophylla	(T)10	29735	29744	10	non-coding	LSC
S. heterophylla	(T)13	43819	43831	13	ycf3	LSC
S. heterophylla	(A)10	45022	45031	10	non-coding	LSC
S. heterophylla	(T)11	64396	64406	11	non-coding	LSC
S. heterophylla	(A)10	69035	69044	10	non-coding	LSC
S. heterophylla	(A)11	71301	71311	11	clpP	LSC
S. heterophylla	(T)11	75673	75683	11	petB	LSC
S. heterophylla	(T)11	78576	78586	11	rpoA	LSC
S. heterophylla	(T)10	80985	80994	10	non-coding	LSC
S. heterophylla	(T)10	102554	102563	10	trnI-GAU	IRB
S. heterophylla	(A)10	134220	134229	10	trnI-GAU	IRA
S. rubicunda	(T)10	106	115	10	non-coding	LSC
S. rubicunda	(A)11	4427	4437	11	non-coding	LSC
S. rubicunda	(A)10	5024	5033	10	non-coding	LSC
S. rubicunda	(A)11	12109	12119	11	non-coding	LSC
S. rubicunda	(T)15	13474	13488	15	non-coding	LSC
S. rubicunda	(T)11	19074	19084	11	rpoC2	LSC
S. rubicunda	(T)10	29838	29847	10	non-coding	LSC
S. rubicunda	(T)10	32211	32220	10	non-coding	LSC
S. rubicunda	(A)10	37880	37889	10	non-coding	LSC
S. rubicunda	(A)14	42846	42859	14	non-coding	LSC
S. rubicunda	(T)11	43804	43814	11	ycf3	LSC
S. rubicunda	(T)14	43933	43946	14	ycf3	LSC
S. rubicunda	(A)10	44885	44894	10	ycf3	LSC
S. rubicunda	(A)11	45139	45149	11	non-coding	LSC
S. rubicunda	(T)10	46015	46024	10	non-coding	LSC
S. rubicunda	(A)10	47677	47686	10	non-coding	LSC
S. rubicunda	(A)11	63943	63953	11	non-coding	LSC
S. rubicunda	(T)12	69262	69273	12	non-coding	LSC
S. rubicunda	(A)10	70874	70883	10	clpP	LSC
S. rubicunda	(T)10	78144	78153	10	rpoA	LSC
S. rubicunda	(T)13	80561	80573	13	non-coding	LSC
S. rubicunda	(T)10	84011	84020	10	non-coding	IRB
S. rubicunda	(T)10	111923	111932	10	non-coding	SSC
S. rubicunda	(T)14	112697	112710	14	non-coding	SSC
S. rubicunda	(A)15	115510	115524	15	non-coding	SSC
S. rubicunda	(A)10	151953	151962	10	non-coding	IRA
S. leptosiphon	(T)10	13435	13444	10	non-coding	LSC
S. leptosiphon	(A)11	16774	16784	11	non-coding	LSC
S. leptosiphon	(T)11	19026	19036	11	rpoC2	LSC
S. leptosiphon	(T)10	29754	29763	10	non-coding	LSC
S. leptosiphon	(T)10	35933	35942	10	non-coding	LSC
S. leptosiphon	(T)10	64368	64377	10	non-coding	LSC
S. leptosiphon	(T)10	67814	67823	10	non-coding	LSC
S. leptosiphon	(T)12	75633	75644	12	petB	LSC
S. leptosiphon	(T)13	76043	76055	13	petB	LSC
S. leptosiphon	(T)10	78536	78545	10	rpoA	LSC
S. leptosiphon	(T)10	102509	102518	10	trnI-GAU	IRB
S. leptosiphon	(A)10	133973	133982	10	trnI-GAU	IRA
S. heydei	(T)11	5940	5950	11	rps16	LSC
S. heydei	(A)11	8612	8622	11	non-coding	LSC
S. heydei	(T)10	8981	8990	10	non-coding	LSC
S. heydei	(T)12	12814	12825	12	atpF	LSC
S. heydei	(T)12	13466	13477	12	non-coding	LSC
S. heydei	(T)11	19050	19060	11	rpoC2	LSC
S. heydei	(T)11	32176	32186	11	non-coding	LSC
S. heydei	(T)11	43872	43882	11	ycf3	LSC
S. heydei	(T)10	61519	61528	10	non-coding	LSC
S. heydei	(T)11	64423	64433	11	non-coding	LSC
S. heydei	(T)10	66484	66493	10	non-coding	LSC
S. heydei	(T)10	70081	70090	10	non-coding	LSC
S. heydei	(T)10	75605	75614	10	petB	LSC

S. heydei	(T)11	76013	76023	11	petB	LSC
S. heydei	(T)11	78500	78510	11	rpoA	LSC
S. heydei	(T)10	102475	102484	10	trnI-GAU	IRB
S. heydei	(A)10	133936	133945	10	trnI-GAU	IRA
S. hissarica	(A)10	4409	4418	10	non-coding	LSC
S. hissarica	(A)11	8618	8628	11	non-coding	LSC
S. hissarica	(A)12	9087	9098	12	non-coding	LSC
S. hissarica	(T)10	12780	12789	10	atpF	LSC
S. hissarica	(T)14	13431	13444	14	non-coding	LSC
S. hissarica	(T)11	19032	19042	11	rpoC2	LSC
S. hissarica	(T)10	32144	32153	10	non-coding	LSC
S. hissarica	(T)11	32526	32536	11	non-coding	LSC
S. hissarica	(T)10	35967	35976	10	non-coding	LSC
S. hissarica	(A)15	42769	42783	15	non-coding	LSC
S. hissarica	(T)10	43728	43737	10	ycf3	LSC
S. hissarica	(A)10	44811	44820	10	ycf3	LSC
S. hissarica	(A)13	45065	45077	13	non-coding	LSC
S. hissarica	(T)10	45948	45957	10	non-coding	LSC
S. hissarica	(A)10	47628	47637	10	non-coding	LSC
S. hissarica	(A)10	61518	61527	10	non-coding	LSC
S. hissarica	(A)10	63938	63947	10	non-coding	LSC
S. hissarica	(T)10	66016	66025	10	non-coding	LSC
S. hissarica	(T)10	69253	69262	10	non-coding	LSC
S. hissarica	(A)10	70858	70867	10	clpP	LSC
S. hissarica	(T)13	78142	78154	13	rpoA	LSC
S. hissarica	(T)11	80562	80572	11	non-coding	LSC
S. hissarica	(T)10	84015	84024	10	non-coding	IRB
S. hissarica	(T)10	102127	102136	10	trnI-GAU	IRB
S. hissarica	(A)10	111503	111512	10	non-coding	SSC
S. hissarica	(A)14	111834	111847	14	non-coding	SSC
S. hissarica	(A)18	115540	115557	18	non-coding	SSC
S. hissarica	(T)10	119294	119303	10	ndhA	SSC
S. hissarica	(A)10	133867	133876	10	trnI-GAU	IRA
S. hissarica	(A)10	151979	151988	10	non-coding	IRA
S. tomentosa	(A)10	4472	4481	10	non-coding	LSC
S. tomentosa	(T)10	7990	7999	10	non-coding	LSC
S. tomentosa	(T)10	12408	12417	10	atpA	LSC
S. tomentosa	(T)13	13058	13070	13	non-coding	LSC
S. tomentosa	(A)10	16399	16408	10	non-coding	LSC
S. tomentosa	(T)11	18644	18654	11	rpoC2	LSC
S. tomentosa	(T)10	31755	31764	10	non-coding	LSC
S. tomentosa	(T)10	43466	43475	10	ycf3	LSC
S. tomentosa	(T)10	75276	75285	10	petB	LSC
S. tomentosa	(T)11	75684	75694	11	petB	LSC
S. tomentosa	(T)11	78167	78177	11	rpoA	LSC
S. tomentosa	(T)10	80579	80588	10	non-coding	LSC
S. tomentosa	(T)10	102140	102149	10	trnI-GAU	IRB
S. tomentosa	(A)10	112106	112115	10	non-coding	SSC
S. tomentosa	(A)10	133612	133621	10	trnI-GAU	IRA
S. immaculata	(T)10	5940	5949	10	rps16	LSC
S. immaculata	(T)10	12636	12645	10	atpF	LSC
S. immaculata	(T)16	13286	13301	16	non-coding	LSC
S. immaculata	(A)10	16631	16640	10	non-coding	LSC
S. immaculata	(T)11	18876	18886	11	rpoC2	LSC
S. immaculata	(T)10	32031	32040	10	non-coding	LSC
S. immaculata	(T)11	35840	35850	11	non-coding	LSC
S. immaculata	(T)11	43750	43760	11	ycf3	LSC
S. immaculata	(T)10	64316	64325	10	non-coding	LSC
S. immaculata	(T)10	67765	67774	10	non-coding	LSC
S. immaculata	(A)10	68942	68951	10	non-coding	LSC
S. immaculata	(T)11	75582	75592	11	petB	LSC
S. immaculata	(T)10	102435	102444	10	trnI-GAU	IRB
S. immaculata	(A)10	133843	133852	10	trnI-GAU	IRA
S. linearis	(T)10	5939	5948	10	rps16	LSC
S. linearis	(A)10	8224	8233	10	non-coding	LSC
S. linearis	(A)11	8841	8851	11	non-coding	LSC
S. linearis	(T)10	12408	12417	10	atpF	LSC
S. linearis	(T)12	13058	13069	12	non-coding	LSC
S. linearis	(T)11	18642	18652	11	rpoC2	LSC
S. linearis	(T)10	35574	35583	10	non-coding	LSC
S. linearis	(T)10	63998	64007	10	non-coding	LSC
S. linearis	(T)10	67445	67454	10	non-coding	LSC
S. linearis	(A)10	68622	68631	10	non-coding	LSC
S. linearis	(T)10	75258	75267	10	petB	LSC
S. linearis	(T)10	80557	80566	10	non-coding	LSC
S. linearis	(T)10	102106	102115	10	trnI-GAU	IRB
S. linearis	(A)10	133580	133589	10	trnI-GAU	IRA
S. indica	(T)11	8265	8275	11	non-coding	LSC
S. indica	(A)10	8491	8500	10	non-coding	LSC
S. indica	(A)10	9013	9022	10	non-coding	LSC
S. indica	(T)11	13314	13324	11	non-coding	LSC

S. indica	(T)11	18908	18918	11	rpoC2	LSC
S. indica	(T)10	32243	32252	10	non-coding	LSC
S. indica	(A)15	42502	42516	15	non-coding	LSC
S. indica	(A)10	44529	44538	10	ycf3	LSC
S. indica	(A)10	47348	47357	10	non-coding	LSC
S. indica	(T)10	63841	63850	10	non-coding	LSC
S. indica	(T)12	69200	69211	12	non-coding	LSC
S. indica	(T)10	69568	69577	10	non-coding	LSC
S. indica	(A)10	71289	71298	10	non-coding	LSC
S. indica	(T)10	75190	75199	10	petB	LSC
S. indica	(T)10	75599	75608	10	petB	LSC
S. indica	(T)10	78104	78113	10	rpoA	LSC
S. indica	(T)11	80000	80010	11	non-coding	LSC
S. indica	(T)12	80519	80530	12	non-coding	LSC
S. indica	(A)11	119476	119486	11	ndhA	SSC
S. mollis	(A)10	5100	5109	10	non-coding	LSC
S. mollis	(T)10	8387	8396	10	non-coding	LSC
S. mollis	(T)12	13429	13440	12	non-coding	LSC
S. mollis	(T)11	19017	19027	11	rpoC2	LSC
S. mollis	(T)10	32600	32609	10	non-coding	LSC
S. mollis	(A)10	44868	44877	10	ycf3	LSC
S. mollis	(T)12	69868	69879	12	non-coding	LSC
S. mollis	(T)10	70236	70245	10	non-coding	LSC
S. mollis	(T)10	75877	75886	10	petB	LSC
S. mollis	(T)12	76286	76297	12	petB	LSC
S. mollis	(T)10	78793	78802	10	rpoA	LSC
S. mollis	(T)11	80689	80699	11	non-coding	LSC
S. mollis	(T)12	81209	81220	12	non-coding	LSC
S. mollis	(A)12	112954	112965	12	non-coding	SSC
S. mollis	(T)10	113382	113391	10	non-coding	SSC
S. pinnatifida subsp. *mucida*	(A)10	4463	4472	10	non-coding	LSC
S. pinnatifida subsp. *mucida*	(T)11	5933	5943	11	rps16	LSC
S. pinnatifida subsp. *mucida*	(A)10	8221	8230	10	non-coding	LSC
S. pinnatifida subsp. *mucida*	(T)10	12420	12429	10	atpF	LSC
S. pinnatifida subsp. *mucida*	(T)13	13070	13082	13	non-coding	LSC
S. pinnatifida subsp. *mucida*	(A)10	16411	16420	10	non-coding	LSC
S. pinnatifida subsp. *mucida*	(T)11	18656	18666	11	rpoC2	LSC
S. pinnatifida subsp. *mucida*	(T)10	28443	28452	10	non-coding	LSC
S. pinnatifida subsp. *mucida*	(T)10	31767	31776	10	non-coding	LSC
S. pinnatifida subsp. *mucida*	(T)14	75712	75725	14	petB	LSC
S. pinnatifida subsp. *mucida*	(T)10	80610	80619	10	non-coding	LSC
S. pinnatifida subsp. *mucida*	(T)10	102166	102175	10	trnI-GAU	IRB
S. pinnatifida subsp. *mucida*	(A)10	133648	133657	10	trnI-GAU	IRA
S. platensis	(A)11	8635	8645	11	non-coding	LSC
S. platensis	(T)14	13471	13484	14	non-coding	LSC
S. platensis	(T)11	19028	19038	11	rpoC2	LSC
S. platensis	(T)11	29793	29803	11	non-coding	LSC
S. platensis	(T)10	32223	32232	10	non-coding	LSC
S. platensis	(T)14	36005	36018	14	non-coding	LSC
S. platensis	(T)10	43912	43921	10	ycf3	LSC
S. platensis	(A)12	44861	44872	12	ycf3	LSC
S. platensis	(A)11	45117	45127	11	non-coding	LSC
S. platensis	(A)10	47668	47677	10	non-coding	LSC
S. platensis	(T)10	51662	51671	10	non-coding	LSC
S. platensis	(T)10	66589	66598	10	non-coding	LSC
S. platensis	(T)12	69818	69829	12	non-coding	LSC
S. platensis	(T)10	71960	71969	10	clpP	LSC
S. platensis	(T)11	75826	75836	11	petB	LSC
S. platensis	(T)10	78733	78742	10	rpoA	LSC
S. platensis	(T)10	80628	80637	10	non-coding	LSC
S. platensis	(T)10	81146	81155	10	non-coding	LSC
S. platensis	(A)11	81664	81674	11	non-coding	LSC
S. platensis	(A)10	112865	112874	10	non-coding	SSC
S. platensis	(T)11	113290	113300	11	non-coding	SSC
S. forrestii	(A)11	4380	4390	11	non-coding	LSC
S. forrestii	(T)14	13307	13320	14	non-coding	LSC
S. forrestii	(T)14	18893	18906	14	rpoC2	LSC
S. forrestii	(T)10	29197	29206	10	non-coding	LSC
S. forrestii	(T)12	35412	35423	12	non-coding	LSC
S. forrestii	(T)12	43185	43196	12	ycf3	LSC
S. forrestii	(A)11	44522	44532	11	non-coding	LSC
S. forrestii	(T)10	51092	51101	10	non-coding	LSC
S. forrestii	(T)13	63929	63941	13	non-coding	LSC
S. forrestii	(T)10	69215	69224	10	non-coding	LSC
S. forrestii	(A)11	74179	74189	11	non-coding	LSC
S. forrestii	(T)12	75203	75214	12	petB	LSC
S. forrestii	(T)11	78114	78124	11	rpoA	LSC
S. forrestii	(T)10	80518	80527	10	non-coding	LSC
S. forrestii	(T)10	102077	102086	10	trnI-GAU	IRB
S. forrestii	(T)13	112512	112524	13	non-coding	SSC

S. forrestii	(A)10	133571	133580	10	trnI-GAU	IRA
S. discolor	(T)10	8250	8259	10	non-coding	LSC
S. discolor	(A)10	8484	8493	10	non-coding	LSC
S. discolor	(T)10	13297	13306	10	non-coding	LSC
S. discolor	(T)11	18889	18899	11	rpoC2	LSC
S. discolor	(A)13	42592	42604	13	non-coding	LSC
S. discolor	(A)10	44600	44609	10	ycf3	LSC
S. discolor	(A)10	44854	44863	10	non-coding	LSC
S. discolor	(A)10	47404	47413	10	non-coding	LSC
S. discolor	(T)10	63936	63945	10	non-coding	LSC
S. discolor	(T)13	69288	69300	13	non-coding	LSC
S. discolor	(A)10	71378	71387	10	clpP	LSC
S. discolor	(T)10	75257	75266	10	petB	LSC
S. discolor	(T)10	78159	78168	10	rpoA	LSC
S. discolor	(T)10	80055	80064	10	non-coding	LSC
S. discolor	(T)12	80573	80584	12	non-coding	LSC
S. discolor	(A)12	119540	119551	12	ndhA	LSC
S. dependens	(T)10	8391	8400	10	non-coding	LSC
S. dependens	(T)13	13439	13451	13	non-coding	LSC
S. dependens	(T)11	19048	19058	11	rpoC2	LSC
S. dependens	(T)11	29796	29806	11	non-coding	LSC
S. dependens	(T)12	32544	32555	12	non-coding	LSC
S. dependens	(A)10	42782	42791	10	non-coding	LSC
S. dependens	(A)10	44787	44796	10	ycf3	LSC
S. dependens	(A)10	45041	45050	10	non-coding	LSC
S. dependens	(A)11	47517	47527	11	non-coding	LSC
S. dependens	(A)10	48183	48192	10	trnI-GAU	LSC
S. dependens	(T)14	64329	64342	14	non-coding	LSC
S. dependens	(A)10	69022	69031	10	non-coding	LSC
S. dependens	(T)14	69685	69698	14	non-coding	LSC
S. dependens	(T)10	70055	70064	10	non-coding	LSC
S. dependens	(A)10	71777	71786	10	clpP	LSC
S. dependens	(T)10	75672	75681	10	petB	LSC
S. dependens	(T)10	78578	78587	10	rpo	LSC
S. dependens	(T)10	80474	80483	10	non-coding	LSC
S. pekinensis var. *maxima*	(T)11	18999	19009	11	rpoC2	LSC
S. pekinensis var. *maxima*	(T)15	29751	29765	15	non-coding	LSC
S. pekinensis var. *maxima*	(A)12	42766	42777	12	non-coding	LSC
S. pekinensis var. *maxima*	(A)11	44773	44783	11	ycf3	LSC
S. pekinensis var. *maxima*	(A)12	45028	45039	12	non-coding	LSC
S. pekinensis var. *maxima*	(A)10	47590	47599	10	non-coding	LSC
S. pekinensis var. *maxima*	(A)10	48254	48263	10	trnL-UAA	LSC
S. pekinensis var. *maxima*	(T)10	64381	64390	10	non-coding	LSC
S. pekinensis var. *maxima*	(T)15	69746	69760	15	non-coding	LSC
S. pekinensis var. *maxima*	(T)10	70117	70126	10	non-coding	LSC
S. pekinensis var. *maxima*	(T)10	75747	75756	10	petB	LSC
S. pekinensis var. *maxima*	(T)10	78660	78669	10	rpoA	LSC
S. pekinensis var. *maxima*	(T)10	80556	80565	10	non-coding	LSC
S. pekinensis var. *maxima*	(T)16	81074	81089	16	non-coding	LSC
S. pekinensis var. *maxima*	(T)10	84555	84564	10	non-coding	IRB
S. pekinensis var. *maxima*	(A)10	152467	152476	10	non-coding	IRB
S. moniliorhiza	(T)10	8397	8406	10	non-coding	LSC
S. moniliorhiza	(T)13	13442	13454	13	non-coding	LSC
S. moniliorhiza	(T)11	19049	19059	11	rpoC2	LSC
S. moniliorhiza	(A)10	27797	27806	10	non-coding	LSC
S. moniliorhiza	(T)11	29808	29818	11	non-coding	LSC
S. moniliorhiza	(A)11	42816	42826	11	non-coding	LSC
S. moniliorhiza	(A)10	44823	44832	10	ycf3	LSC
S. moniliorhiza	(A)11	45077	45087	11	non-coding	LSC
S. moniliorhiza	(A)10	47639	47648	10	non-coding	LSC
S. moniliorhiza	(A)10	48304	48313	10	trnL-UAA	LSC
S. moniliorhiza	(T)10	64450	64459	10	non-coding	LSC
S. moniliorhiza	(T)12	69806	69817	12	non-coding	LSC
S. moniliorhiza	(T)11	70175	70185	11	non-coding	LSC
S. moniliorhiza	(A)10	71899	71908	10	clpP	LSC
S. moniliorhiza	(T)10	78719	78728	10	rpoA	LSC
S. moniliorhiza	(T)10	80615	80624	10	non-coding	LSC
S. moniliorhiza	(A)11	112874	112884	11	non-coding	SSC
S. orthocalyx	(T)11	9116	9126	11	non-coding	LSC
S. orthocalyx	(A)10	11993	12002	10	non-coding	LSC
S. orthocalyx	(T)14	13372	13385	14	non-coding	LSC
S. orthocalyx	(A)10	13928	13937	10	non-coding	LSC
S. orthocalyx	(T)11	18947	18957	11	rpoC2	LSC
S. orthocalyx	(T)11	29279	29289	11	non-coding	LSC
S. orthocalyx	(T)10	36677	36686	10	non-coding	LSC
S. orthocalyx	(A)10	44348	44357	10	ycf3	LSC
S. orthocalyx	(A)15	44602	44616	15	non-coding	LSC
S. orthocalyx	(T)11	46404	46414	11	non-coding	LSC
S. orthocalyx	(T)11	57385	57395	11	non-coding	LSC
S. orthocalyx	(T)10	69262	69271	10	non-coding	LSC

S. orthocalyx	(A)10	71368	71377	10	clpP	LSC
S. orthocalyx	(T)11	75239	75249	11	petB	LSC
S. orthocalyx	(T)11	75647	75657	11	petB	LSC
S. orthocalyx	(T)11	78150	78160	11	rpoA	LSC
S. orthocalyx	(T)10	84023	84032	10	non-coding	IRB
S. orthocalyx	(T)11	112572	112582	11	non-coding	SSC
S. orthocalyx	(A)10	115378	115387	10	non-coding	SSC
S. orthocalyx	(A)10	151749	151758	10	non-coding	IRA
S. indica var. *coccinea*	(T)12	8252	8263	12	non-coding	LSC
S. indica var. *coccinea*	(A)10	8488	8497	10	non-coding	LSC
S. indica var. *coccinea*	(A)10	9010	9019	10	non-coding	LSC
S. indica var. *coccinea*	(T)11	13304	13314	11	non-coding	LSC
S. indica var. *coccinea*	(T)11	18898	18908	11	rpoC2	LSC
S. indica var. *coccinea*	(T)10	32258	32267	10	non-coding	LSC
S. indica var. *coccinea*	(A)14	42525	42538	14	non-coding	LSC
S. indica var. *coccinea*	(A)10	44533	44542	10	ycf3	LSC
S. indica var. *coccinea*	(A)10	44787	44796	10	non-coding	LSC
S. indica var. *coccinea*	(A)10	47353	47362	10	non-coding	LSC
S. indica var. *coccinea*	(A)10	48011	48020	10	trnL-UAA	LSC
S. indica var. *coccinea*	(A)10	61488	61497	10	non-coding	LSC
S. indica var. *coccinea*	(T)10	63870	63879	10	non-coding	LSC
S. indica var. *coccinea*	(T)12	69229	69240	12	non-coding	LSC
S. indica var. *coccinea*	(T)10	69597	69606	10	non-coding	LSC
S. indica var. *coccinea*	(A)10	71318	71327	10	clpP	LSC
S. indica var. *coccinea*	(T)10	75219	75228	10	petB	LSC
S. indica var. *coccinea*	(T)10	78132	78141	10	rpoA	LSC
S. indica var. *coccinea*	(T)10	80028	80037	10	non-coding	LSC
S. indica var. *coccinea*	(T)13	80546	80558	13	non-coding	LSC
S. indica var. *coccinea*	(A)10	112275	112284	10	non-coding	SSC
S. indica var. *coccinea*	(A)14	119514	119527	14	ndhA	SSC
S. strigillosa	(T)13	13451	13463	13	non-coding	LSC
S. strigillosa	(T)11	19052	19062	11	rpoC2	LSC
S. strigillosa	(A)10	27800	27809	10	non-coding	LSC
S. strigillosa	(T)12	29811	29822	12	non-coding	LSC
S. strigillosa	(A)13	42822	42834	13	non-coding	LSC
S. strigillosa	(A)10	44831	44840	10	ycf3	LSC
S. strigillosa	(A)12	45085	45096	12	non-coding	LSC
S. strigillosa	(A)10	47648	47657	10	non-coding	LSC
S. strigillosa	(A)10	48313	48322	10	trnL-UAA	LSC
S. strigillosa	(T)10	64459	64468	10	non-coding	LSC
S. strigillosa	(T)12	69816	69827	12	non-coding	LSC
S. strigillosa	(T)10	70185	70194	10	non-coding	LSC
S. strigillosa	(A)10	71908	71917	10	clpP	LSC
S. strigillosa	(T)10	78728	78737	10	rpoA	LSC
S. strigillosa	(T)10	80624	80633	10	non-coding	LSC
S. strigillosa	(A)11	112871	112881	11	non-coding	LSC
S. baicalensis	(A)11	4377	4387	11	non-coding	LSC
S. baicalensis	(T)10	9120	9129	10	non-coding	LSC
S. baicalensis	(A)10	11996	12005	10	non-coding	LSC
S. baicalensis	(T)13	13365	13377	13	non-coding	LSC
S. baicalensis	(T)11	18937	18947	11	rpoC2	LSC
S. baicalensis	(T)10	36712	36721	10	non-coding	LSC
S. baicalensis	(T)10	43429	43438	10	ycf3	LSC
S. baicalensis	(A)10	44384	44393	10	ycf3	LSC
S. baicalensis	(T)10	51154	51163	10	non-coding	LSC
S. baicalensis	(T)10	57432	57441	10	non-coding	LSC
S. baicalensis	(T)12	63996	64007	12	non-coding	LSC
S. baicalensis	(T)11	69282	69292	11	non-coding	LSC
S. baicalensis	(A)10	71388	71397	10	clpP	LSC
S. baicalensis	(A)10	74231	74240	10	non-coding	LSC
S. baicalensis	(T)10	75668	75677	10	petB	LSC
S. baicalensis	(T)10	84044	84053	10	non-coding	IRB
S. baicalensis	(T)12	112582	112593	12	non-coding	SSC
S. baicalensis	(A)10	151746	151755	10	non-coding	IRA
S. altamaha	(A)10	6870	6879	10	non-coding	LSC
S. altamaha	(A)11	8081	8091	11	non-coding	LSC
S. altamaha	(T)10	8365	8374	10	non-coding	LSC
S. altamaha	(A)10	8590	8599	10	non-coding	LSC
S. altamaha	(A)10	9929	9938	10	trnG-UCC	LSC
S. altamaha	(T)11	18979	18989	11	rpoC2	LSC
S. altamaha	(T)10	27819	27828	10	non-coding	LSC
S. altamaha	(T)12	32115	32126	12	non-coding	LSC
S. altamaha	(T)10	32450	32459	10	non-coding	LSC
S. altamaha	(T)10	35872	35881	10	non-coding	LSC
S. altamaha	(A)12	42688	42699	12	non-coding	LSC
S. altamaha	(T)13	46800	46812	13	non-coding	LSC
S. altamaha	(A)13	46929	46941	13	non-coding	LSC
S. altamaha	(A)12	47586	47597	12	trnL-UAA	LSC
S. altamaha	(A)10	57332	57341	10	non-coding	LSC
S. altamaha	(T)12	69070	69081	12	non-coding	LSC

S. altamaha	(T)10	77959	77968	10	rpoA	LSC
S. altamaha	(T)10	79856	79865	10	non-coding	LSC
S. altamaha	(T)10	83473	83482	10	rpl22	LSC
S. altamaha	(T)11	83847	83857	11	non-coding	IRB
S. altamaha	(A)11	151456	151466	11	non-coding	IRA
S. gardoquioides	(A)10	4959	4968	10	non-coding	LSC
S. gardoquioides	(A)10	7951	7960	10	non-coding	LSC
S. gardoquioides	(T)10	8246	8255	10	non-coding	LSC
S. gardoquioides	(A)10	8471	8480	10	non-coding	LSC
S. gardoquioides	(A)13	8950	8962	13	non-coding	LSC
S. gardoquioides	(A)10	11948	11957	10	non-coding	LSC
S. gardoquioides	(T)15	13321	13335	15	non-coding	LSC
S. gardoquioides	(A)11	15780	15790	11	non-coding	LSC
S. gardoquioides	(T)11	18918	18928	11	rpoC2	LSC
S. gardoquioides	(A)11	28950	28960	11	non-coding	LSC
S. gardoquioides	(T)11	31994	32004	11	non-coding	LSC
S. gardoquioides	(A)12	42621	42632	12	non-coding	LSC
S. gardoquioides	(A)13	44901	44913	13	non-coding	LSC
S. gardoquioides	(T)10	45796	45805	10	non-coding	LSC
S. gardoquioides	(A)11	48126	48136	11	trnL-UAA	LSC
S. gardoquioides	(T)10	48886	48895	10	non-coding	LSC
S. gardoquioides	(A)10	59559	59568	10	non-coding	LSC
S. gardoquioides	(T)12	67239	67250	12	non-coding	LSC
S. gardoquioides	(T)12	69095	69106	12	non-coding	LSC
S. gardoquioides	(T)10	69463	69472	10	non-coding	LSC
S. gardoquioides	(A)10	70709	70718	10	clpP	LSC
S. gardoquioides	(A)10	74044	74053	10	non-coding	LSC
S. gardoquioides	(T)10	75078	75087	10	petB	LSC
S. gardoquioides	(T)18	77995	78012	18	rpoA	LSC
S. gardoquioides	(T)15	80411	80425	15	non-coding	LSC
S. gardoquioides	(A)11	112127	112137	11	non-coding	SSC
S. gardoquioides	(T)10	119073	119082	10	ndhA	SSC
S. gaumeri	(A)10	4365	4374	10	non-coding	LSC
S. gaumeri	(T)14	13418	13431	14	non-coding	LSC
S. gaumeri	(T)11	18989	18999	11	rpoC2	LSC
S. gaumeri	(T)12	32166	32177	12	non-coding	LSC
S. gaumeri	(A)13	44768	44780	13	ycf3	LSC
S. gaumeri	(T)10	45894	45903	10	non-coding	LSC
S. gaumeri	(A)10	47549	47558	10	non-coding	LSC
S. gaumeri	(A)10	48429	48438	10	non-coding	LSC
S. gaumeri	(T)10	51541	51550	10	non-coding	LSC
S. gaumeri	(T)15	69633	69647	15	non-coding	LSC
S. gaumeri	(A)10	71252	71261	10	clpP	LSC
S. gaumeri	(T)10	75612	75621	10	petB	LSC
S. gaumeri	(T)10	76021	76030	10	petB	LSC
S. gaumeri	(T)10	78536	78545	10	rpoA	LSC
S. gaumeri	(T)13	80949	80961	13	non-coding	LSC
S. gaumeri	(T)10	84419	84428	10	non-coding	IRB
S. gaumeri	(T)10	111954	111963	10	non-coding	SSC
S. gaumeri	(T)10	113139	113148	10	non-coding	SSC
S. gaumeri	(T)10	126307	126316	10	ycf1	SSC
S. gaumeri	(A)10	152276	152285	10	non-coding	IRA
S. hintoniana	(A)10	4290	4299	10	non-coding	LSC
S. hintoniana	(A)10	4982	4991	10	non-coding	LSC
S. hintoniana	(A)10	8170	8179	10	non-coding	LSC
S. hintoniana	(A)13	11657	11669	13	non-coding	LSC
S. hintoniana	(T)10	12365	12374	10	atpF	LSC
S. hintoniana	(T)11	18625	18635	11	rpoC2	LSC
S. hintoniana	(T)11	29361	29371	11	non-coding	LSC
S. hintoniana	(T)10	35513	35522	10	non-coding	LSC
S. hintoniana	(A)10	42320	42329	10	non-coding	LSC
S. hintoniana	(A)10	44605	44614	10	non-coding	LSC
S. hintoniana	(T)10	46420	46429	10	non-coding	LSC
S. hintoniana	(T)10	48568	48577	10	non-coding	LSC
S. hintoniana	(A)10	63361	63370	10	non-coding	LSC
S. hintoniana	(A)10	67612	67621	10	non-coding	LSC
S. hintoniana	(T)14	68279	68292	14	non-coding	LSC
S. hintoniana	(T)10	68649	68658	10	non-coding	LSC
S. hintoniana	(T)10	74250	74259	10	petB	LSC
S. hintoniana	(T)11	77179	77189	11	rpoA	LSC
S. hintoniana	(T)10	79082	79091	10	non-coding	LSC
S. hintoniana	(T)10	83062	83071	10	non-coding	IRB
S. hintoniana	(T)11	123709	123719	11	ycf1	SSC
S. hintoniana	(T)13	124209	124221	13	ycf1	SSC
S. hintoniana	(A)10	150966	150975	10	non-coding	IRA
S. nummulariifolia	(T)10	8369	8378	10	non-coding	LSC
S. nummulariifolia	(A)10	8603	8612	10	non-coding	LSC
S. nummulariifolia	(T)14	13432	13445	14	non-coding	LSC
S. nummulariifolia	(T)11	19053	19063	11	rpoC2	LSC
S. nummulariifolia	(A)10	42869	42878	10	non-coding	LSC
S. nummulariifolia	(A)11	44874	44884	11	ycf3	LSC

S. nummulariifolia	(A)10	45129	45138	10	non-coding	LSC
S. nummulariifolia	(A)10	48342	48351	10	trnL-UAA	LSC
S. nummulariifolia	(T)13	69854	69866	13	non-coding	LSC
S. nummulariifolia	(T)10	70218	70227	10	non-coding	LSC
S. nummulariifolia	(T)12	75853	75864	12	petB	LSC
S. nummulariifolia	(T)10	76264	76273	10	petB	LSC
S. nummulariifolia	(T)10	78763	78772	10	rpoA	LSC
S. nummulariifolia	(T)10	80659	80668	10	non-coding	LSC
S. nummulariifolia	(T)15	81177	81191	15	non-coding	LSC
S. nummulariifolia	(A)10	112919	112928	10	non-coding	SSC
S. scutellarioides	(A)10	4965	4974	10	non-coding	LSC
S. scutellarioides	(T)10	8142	8151	10	non-coding	LSC
S. scutellarioides	(A)10	8377	8386	10	non-coding	LSC
S. scutellarioides	(T)13	13223	13235	13	non-coding	LSC
S. scutellarioides	(A)10	15678	15687	10	non-coding	LSC
S. scutellarioides	(T)11	18815	18825	11	rpoC2	LSC
S. scutellarioides	(A)10	28869	28878	10	non-coding	LSC
S. scutellarioides	(T)10	29534	29543	10	non-coding	LSC
S. scutellarioides	(T)12	43577	43588	12	ycf3	LSC
S. scutellarioides	(A)10	44530	44539	10	ycf3	LSC
S. scutellarioides	(A)12	44784	44795	12	non-coding	LSC
S. scutellarioides	(T)11	48449	48459	11	non-coding	LSC
S. scutellarioides	(A)10	63345	63354	10	non-coding	LSC
S. scutellarioides	(T)11	68655	68665	11	non-coding	LSC
S. scutellarioides	(T)10	69030	69039	10	non-coding	LSC
S. scutellarioides	(T)10	77551	77560	10	rpoA	LSC
S. scutellarioides	(T)11	80078	80088	11	non-coding	LSC
S. scutellarioides	(T)10	83548	83557	10	non-coding	IRB
S. scutellarioides	(T)10	118846	118855	10	ndhA	SSC
S. scutellarioides	(A)10	151480	151489	10	non-coding	IRA
S. seleriana	(T)11	107	117	11	non-coding	LSC
S. seleriana	(A)10	4403	4412	10	non-coding	LSC
S. seleriana	(T)10	6597	6606	10	non-coding	LSC
S. seleriana	(A)10	6875	6884	10	non-coding	LSC
S. seleriana	(T)10	8362	8371	10	non-coding	LSC
S. seleriana	(A)10	8588	8597	10	non-coding	LSC
S. seleriana	(T)10	12760	12769	10	atpF	LSC
S. seleriana	(T)10	13411	13420	10	non-coding	LSC
S. seleriana	(T)11	18997	19007	11	rpoC2	LSC
S. seleriana	(T)10	29717	29726	10	non-coding	LSC
S. seleriana	(T)11	32142	32152	11	non-coding	LSC
S. seleriana	(T)10	35929	35938	10	non-coding	LSC
S. seleriana	(T)10	45878	45887	10	non-coding	LSC
S. seleriana	(T)10	51502	51511	10	non-coding	LSC
S. seleriana	(T)11	69181	69191	11	non-coding	LSC
S. seleriana	(A)10	72021	72030	10	non-coding	LSC
S. seleriana	(T)12	75167	75178	12	petB	LSC
S. seleriana	(T)11	78085	78095	11	rpoA	LSC
S. seleriana	(T)10	79970	79979	10	non-coding	LSC
S. seleriana	(T)11	80488	80498	11	non-coding	LSC
S. seleriana	(T)11	112626	112636	11	non-coding	SSC
S. seleriana	(T)10	119157	119166	10	ndhA	SSC
S. splendens	(A)10	4434	4443	10	non-coding	LSC
S. splendens	(T)11	6621	6631	11	non-coding	LSC
S. splendens	(A)11	6900	6910	11	non-coding	LSC
S. splendens	(T)10	8403	8412	10	non-coding	LSC
S. splendens	(A)10	8638	8647	10	non-coding	LSC
S. splendens	(A)10	12085	12094	10	non-coding	LSC
S. splendens	(T)11	12799	12809	11	atpF	LSC
S. splendens	(T)13	13450	13462	13	non-coding	LSC
S. splendens	(T)11	19032	19042	11	rpoC2	LSC
S. splendens	(A)10	29088	29097	10	non-coding	LSC
S. splendens	(T)10	29758	29767	10	non-coding	LSC
S. splendens	(T)10	32183	32192	10	non-coding	LSC
S. splendens	(T)10	32515	32524	10	non-coding	LSC
S. splendens	(T)11	35964	35974	11	non-coding	LSC
S. splendens	(A)10	37808	37817	10	non-coding	LSC
S. splendens	(A)10	42776	42785	10	non-coding	LSC
S. splendens	(A)10	44806	44815	10	non-coding	LSC
S. splendens	(T)11	45952	45962	11	non-coding	LSC
S. splendens	(A)10	48282	48291	10	non-coding	LSC
S. splendens	(T)10	51593	51602	10	non-coding	LSC
S. splendens	(T)11	69250	69260	11	non-coding	LSC
S. splendens	(T)10	75224	75233	10	petB	LSC
S. splendens	(T)15	78152	78166	15	rpoA	LSC
S. splendens	(T)11	80041	80051	11	non-coding	LSC
S. splendens	(T)11	80552	80562	11	non-coding	LSC
S. splendens	(T)10	112710	112719	10	non-coding	SSC
S. insignis	(T)11	2213	2223	11	non-coding	LSC
S. insignis	(T)13	13161	13173	13	non-coding	LSC

S. insignis	(T)11	18756	18766	11	rpoC2	LSC
S. insignis	(T)10	35709	35718	10	non-coding	LSC
S. insignis	(A)14	42512	42525	14	non-coding	LSC
S. insignis	(A)10	44525	44534	10	non-coding	LSC
S. insignis	(A)10	44779	44788	10	non-coding	LSC
S. insignis	(A)11	47326	47336	11	non-coding	LSC
S. insignis	(A)10	61431	61440	10	non-coding	LSC
S. insignis	(T)13	69191	69203	13	non-coding	LSC
S. insignis	(T)10	69560	69569	10	non-coding	LSC
S. insignis	(A)10	71272	71281	10	non-coding	LSC
S. insignis	(T)10	75173	75182	10	non-coding	LSC
S. insignis	(T)10	75582	75591	10	non-coding	LSC
S. insignis	(T)17	78087	78103	17	non-coding	LSC
S. insignis	(T)10	79983	79992	10	non-coding	LSC
S. insignis	(T)16	80501	80516	16	non-coding	LSC
S. insignis	(A)15	112231	112245	15	non-coding	SSC
S. insignis	(A)13	115458	115470	13	non-coding	SSC
S. lateriflora	(A)10	5087	5096	10	non-coding	LSC
S. lateriflora	(T)10	8309	8318	10	non-coding	LSC
S. lateriflora	(A)10	8542	8551	10	non-coding	LSC
S. lateriflora	(T)13	13300	13312	13	non-coding	LSC
S. lateriflora	(T)11	18900	18910	11	rpoC2	LSC
S. lateriflora	(A)10	27647	27656	10	non-coding	LSC
S. lateriflora	(T)12	29658	29669	12	non-coding	LSC
S. lateriflora	(A)13	42644	42656	13	non-coding	LSC
S. lateriflora	(A)10	44652	44661	10	non-coding	LSC
S. lateriflora	(A)10	44906	44915	10	non-coding	LSC
S. lateriflora	(A)11	48143	48153	11	non-coding	LSC
S. lateriflora	(T)10	64257	64266	10	non-coding	LSC
S. lateriflora	(T)14	69612	69625	14	non-coding	LSC
S. lateriflora	(A)10	71685	71694	10	non-coding	LSC
S. lateriflora	(T)10	75590	75599	10	non-coding	LSC
S. lateriflora	(T)10	78503	78512	10	non-coding	LSC
S. lateriflora	(T)12	80398	80409	12	non-coding	LSC
S. lateriflora	(T)14	80914	80927	14	non-coding	LSC
S. lateriflora	(A)17	112648	112664	17	non-coding	SSC
S. lateriflora	(A)10	120404	120413	10	non-coding	SSC
S. angustifolia	(T)11	13406	13416	11	non-coding	LSC
S. angustifolia	(T)11	19012	19022	11	rpoC2	LSC
S. angustifolia	(A)10	27687	27696	10	non-coding	LSC
S. angustifolia	(T)10	29757	29766	10	non-coding	LSC
S. angustifolia	(T)10	35942	35951	10	non-coding	LSC
S. angustifolia	(A)11	45017	45027	11	non-coding	LSC
S. angustifolia	(T)11	47403	47413	11	non-coding	LSC
S. angustifolia	(A)10	48237	48246	10	trnL-UAA	LSC
S. angustifolia	(T)11	51385	51395	11	non-coding	LSC
S. angustifolia	(T)10	51582	51591	10	non-coding	LSC
S. angustifolia	(A)11	69101	69111	11	non-coding	LSC
S. angustifolia	(T)10	69763	69772	10	non-coding	LSC
S. angustifolia	(T)10	75772	75781	10	petB	LSC
S. angustifolia	(T)10	78685	78694	10	rpoA	LSC
S. angustifolia	(T)13	81099	81111	13	non-coding	LSC
S. angustifolia	(T)10	112099	112108	10	non-coding	SSC
S. angustifolia	(T)10	119806	119815	10	ndhA	SSC
S. parvula	(T)11	8391	8401	11	non-coding	LSC
S. parvula	(A)10	8626	8635	10	non-coding	LSC
S. parvula	(T)13	13440	13452	13	non-coding	LSC
S. parvula	(T)10	16015	16024	10	non-coding	LSC
S. parvula	(T)11	19052	19062	11	rpoC2	LSC
S. parvula	(A)11	27735	27745	11	non-coding	LSC
S. parvula	(T)10	29781	29790	10	non-coding	LSC
S. parvula	(A)13	42835	42847	13	non-coding	LSC
S. parvula	(A)11	45092	45102	11	non-coding	LSC
S. parvula	(T)11	47487	47497	11	non-coding	LSC
S. parvula	(A)10	47658	47667	10	non-coding	LSC
S. parvula	(T)10	51635	51644	10	non-coding	LSC
S. parvula	(A)10	69163	69172	10	non-coding	LSC
S. parvula	(T)11	75806	75816	11	petB	LSC
S. parvula	(T)10	76216	76225	10	petB	LSC
S. parvula	(T)10	78721	78730	10	rpoA	LSC
S. parvula	(T)11	80617	80627	11	non-coding	LSC
S. parvula	(T)14	81136	81149	14	non-coding	LSC
S. parvula	(T)13	113296	113308	13	non-coding	SSC
S. parvula	(T)10	119849	119858	10	ndhA	SSC
S. guatemalensis	(A)11	4364	4374	11	non-coding	LSC
S. guatemalensis	(A)10	5061	5070	10	non-coding	LSC
S. guatemalensis	(T)10	12778	12787	10	atpF	LSC
S. guatemalensis	(T)13	13429	13441	13	non-coding	LSC
S. guatemalensis	(T)11	18999	19009	11	rpoC2	LSC
S. guatemalensis	(T)12	32159	32170	12	non-coding	LSC

S. guatemalensis	(A)13	44763	44775	13	non-coding	LSC
S. guatemalensis	(T)10	45889	45898	10	non-coding	LSC
S. guatemalensis	(A)14	47545	47558	14	non-coding	LSC
S. guatemalensis	(A)10	48214	48223	10	trnL-UAA	LSC
S. guatemalensis	(A)10	48430	48439	10	non-coding	LSC
S. guatemalensis	(A)10	64335	64344	10	non-coding	LSC
S. guatemalensis	(T)15	69635	69649	15	non-coding	LSC
S. guatemalensis	(T)11	76025	76035	11	petB	LSC
S. guatemalensis	(T)10	78541	78550	10	rpoA	LSC
S. guatemalensis	(T)13	80954	80966	13	non-coding	LSC
S. guatemalensis	(T)10	84424	84433	10	non-coding	IRB
S. guatemalensis	(T)12	111952	111963	12	non-coding	SSC
S. guatemalensis	(T)10	126305	126314	10	ycf1	SSC
S. guatemalensis	(A)10	152267	152276	10	non-coding	IRA
S. longifolia	(T)10	6594	6603	10	non-coding	LSC
S. longifolia	(T)10	8380	8389	10	non-coding	LSC
S. longifolia	(A)10	8606	8615	10	non-coding	LSC
S. longifolia	(T)10	12772	12781	10	atpF	LSC
S. longifolia	(T)11	13422	13432	11	non-coding	LSC
S. longifolia	(T)10	15984	15993	10	non-coding	LSC
S. longifolia	(T)11	19009	19019	11	rpoC2	LSC
S. longifolia	(T)10	29703	29712	10	non-coding	LSC
S. longifolia	(T)10	32107	32116	10	non-coding	LSC
S. longifolia	(T)11	35886	35896	11	non-coding	LSC
S. longifolia	(T)10	45865	45874	10	non-coding	LSC
S. longifolia	(A)10	59620	59629	10	non-coding	LSC
S. longifolia	(A)10	61418	61427	10	non-coding	LSC
S. longifolia	(A)10	63829	63838	10	non-coding	LSC
S. longifolia	(A)10	68492	68501	10	non-coding	LSC
S. longifolia	(T)11	69154	69164	11	non-coding	LSC
S. longifolia	(A)11	71994	72004	11	non-coding	LSC
S. longifolia	(T)12	75141	75152	12	petB	LSC
S. longifolia	(T)14	78053	78066	14	rpoA	LSC
S. longifolia	(T)11	79941	79951	11	non-coding	LSC
S. longifolia	(T)12	80460	80471	12	non-coding	LSC
S. longifolia	(A)10	112143	112152	10	non-coding	SSC
S. longifolia	(T)10	112587	112596	10	non-coding	SSC
S. longifolia	(T)11	119117	119127	11	ndhA	SSC
S. suffrutescens	(A)10	4391	4400	10	non-coding	LSC
S. suffrutescens	(T)10	4525	4534	10	non-coding	LSC
S. suffrutescens	(A)10	8532	8541	10	non-coding	LSC
S. suffrutescens	(A)10	9004	9013	10	non-coding	LSC
S. suffrutescens	(T)12	11978	11989	12	non-coding	LSC
S. suffrutescens	(A)10	12619	12628	10	atpF	LSC
S. suffrutescens	(T)12	13359	13370	12	non-coding	LSC
S. suffrutescens	(A)12	13904	13915	12	non-coding	LSC
S. suffrutescens	(T)11	18931	18941	11	rpoC2	LSC
S. suffrutescens	(T)10	27812	27821	10	non-coding	LSC
S. suffrutescens	(T)10	32087	32096	10	non-coding	LSC
S. suffrutescens	(T)11	35854	35864	11	non-coding	LSC
S. suffrutescens	(A)10	44979	44988	10	non-coding	LSC
S. suffrutescens	(T)10	45864	45873	10	non-coding	LSC
S. suffrutescens	(T)13	47354	47366	13	non-coding	LSC
S. suffrutescens	(T)10	64243	64252	10	non-coding	LSC
S. suffrutescens	(T)14	69581	69594	14	non-coding	LSC
S. suffrutescens	(T)15	75566	75580	15	petB	LSC
S. suffrutescens	(T)16	75980	75995	16	petB	LSC
S. suffrutescens	(T)10	78500	78509	10	rpoA	LSC
S. suffrutescens	(A)10	112090	112099	10	non-coding	SSC
S. suffrutescens	(T)11	124852	124862	11	ycf1	SSC
S. franchetiana	(T)11	8263	8273	11	non-coding	LSC
S. franchetiana	(A)10	8498	8507	10	non-coding	LSC
S. franchetiana	(T)10	13303	13312	10	non-coding	LSC
S. franchetiana	(T)11	18896	18906	11	rpoC2	LSC
S. franchetiana	(T)14	32404	32417	14	non-coding	LSC
S. franchetiana	(A)14	42655	42668	14	non-coding	LSC
S. franchetiana	(A)11	44665	44675	11	ycf3	LSC
S. franchetiana	(A)10	44920	44929	10	non-coding	LSC
S. franchetiana	(A)11	47470	47480	11	non-coding	LSC
S. franchetiana	(A)10	48129	48138	10	trnL-UAA	LSC
S. franchetiana	(T)10	63993	64002	10	non-coding	LSC
S. franchetiana	(T)13	69352	69364	13	non-coding	LSC
S. franchetiana	(T)10	69721	69730	10	non-coding	LSC
S. franchetiana	(A)10	71442	71451	10	clpP	LSC
S. franchetiana	(T)10	75343	75352	10	petB	LSC
S. franchetiana	(T)15	78256	78270	15	rpoA	LSC
S. franchetiana	(T)10	80150	80159	10	non-coding	LSC
S. franchetiana	(T)14	80663	80676	14	non-coding	LSC
S. franchetiana	(A)11	81213	81223	11	non-coding	LSC
S. franchetiana	(A)11	119641	119651	11	ndhA	SSC

S. nana	(A)10	5087	5096	10	non-coding	LSC
S. nana	(T)11	13421	13431	11	non-coding	LSC
S. nana	(T)11	19021	19031	11	rpoC2	LSC
S. nana	(A)11	27702	27712	11	non-coding	LSC
S. nana	(T)11	29774	29784	11	non-coding	LSC
S. nana	(T)11	35953	35963	11	non-coding	LSC
S. nana	(A)12	45049	45060	12	non-coding	LSC
S. nana	(A)10	47592	47601	10	non-coding	LSC
S. nana	(A)10	48256	48265	10	trnL-UAA	LSC
S. nana	(T)10	51584	51593	10	non-coding	LSC
S. nana	(T)11	69773	69783	11	non-coding	LSC
S. nana	(T)12	75767	75778	12	petB	LSC
S. nana	(T)10	78681	78690	10	rpoA	LSC
S. nana	(T)10	80577	80586	10	non-coding	LSC
S. nana	(T)12	81095	81106	12	non-coding	LSC
S. nana	(T)11	112075	112085	11	non-coding	SSC
S. nana	(T)10	112993	113002	10	non-coding	SSC
S. nana	(T)10	119767	119776	10	ndhA	SSC
S. ovata	(A)10	6865	6874	10	non-coding	LSC
S. ovata	(A)10	8076	8085	10	non-coding	LSC
S. ovata	(A)10	8583	8592	10	non-coding	LSC
S. ovata	(A)10	9922	9931	10	trnG-UCC	LSC
S. ovata	(T)11	18981	18991	11	rpoC2	LSC
S. ovata	(T)10	27821	27830	10	non-coding	LSC
S. ovata	(T)10	32116	32125	10	non-coding	LSC
S. ovata	(T)11	32449	32459	11	non-coding	LSC
S. ovata	(T)11	35872	35882	11	non-coding	LSC
S. ovata	(A)10	36010	36019	10	non-coding	LSC
S. ovata	(A)12	42690	42701	12	non-coding	LSC
S. ovata	(T)13	46801	46813	13	non-coding	LSC
S. ovata	(A)13	46930	46942	13	non-coding	LSC
S. ovata	(A)11	47588	47598	11	trnL-UAA	LSC
S. ovata	(A)10	57320	57329	10	non-coding	LSC
S. ovata	(T)11	69064	69074	11	non-coding	LSC
S. ovata	(T)10	77952	77961	10	rpoA	LSC
S. ovata	(T)10	79849	79858	10	non-coding	LSC
S. ovata	(T)11	83831	83841	11	non-coding	IRB
S. ovata	(A)10	112069	112078	10	non-coding	SSC
S. ovata	(A)11	151572	151582	11	non-coding	IRA
S. bolanderi	(T)10	13401	13410	10	non-coding	LSC
S. bolanderi	(T)11	18998	19008	11	rpoC2	LSC
S. bolanderi	(A)10	27680	27689	10	non-coding	LSC
S. bolanderi	(T)10	35937	35946	10	non-coding	LSC
S. bolanderi	(A)10	42754	42763	10	non-coding	LSC
S. bolanderi	(A)13	45003	45015	13	non-coding	LSC
S. bolanderi	(T)10	47388	47397	10	non-coding	LSC
S. bolanderi	(T)11	69691	69701	11	non-coding	LSC
S. bolanderi	(T)10	75684	75693	10	petB	LSC
S. bolanderi	(T)10	78598	78607	10	rpoA	LSC
S. bolanderi	(T)10	80494	80503	10	non-coding	LSC
S. bolanderi	(T)13	81012	81024	13	non-coding	LSC
S. bolanderi	(T)10	112017	112026	10	non-coding	SSC
S. bolanderi	(T)10	112934	112943	10	non-coding	SSC
S. bolanderi	(T)10	119706	119715	10	ndhA	SSC
S. costaricana	(T)10	107	116	10	non-coding	LSC
S. costaricana	(T)11	6593	6603	11	non-coding	LSC
S. costaricana	(T)10	8373	8382	10	non-coding	LSC
S. costaricana	(A)10	9549	9558	10	trnG-UCC	LSC
S. costaricana	(A)10	12060	12069	10	non-coding	LSC
S. costaricana	(T)11	12774	12784	11	atpF	LSC
S. costaricana	(T)11	19009	19019	11	rpoC2	LSC
S. costaricana	(A)10	29069	29078	10	non-coding	LSC
S. costaricana	(T)10	29739	29748	10	non-coding	LSC
S. costaricana	(T)10	32126	32135	10	non-coding	LSC
S. costaricana	(T)10	32505	32514	10	non-coding	LSC
S. costaricana	(T)10	35946	35955	10	non-coding	LSC
S. costaricana	(A)11	42754	42764	11	non-coding	LSC
S. costaricana	(T)11	45922	45932	11	non-coding	LSC
S. costaricana	(A)10	63890	63899	10	non-coding	LSC
S. costaricana	(A)11	68558	68568	11	non-coding	LSC
S. costaricana	(T)11	69221	69231	11	non-coding	LSC
S. costaricana	(T)11	75200	75210	11	petB	LSC
S. costaricana	(T)14	78114	78127	14	rpoA	LSC
S. costaricana	(T)12	80002	80013	12	non-coding	LSC
S. costaricana	(T)13	80522	80534	13	non-coding	LSC
S. costaricana	(A)11	112246	112256	11	non-coding	SSC
S. costaricana	(T)10	112691	112700	10	non-coding	SSC
S. drummondii	(T)11	105	115	11	non-coding	LSC
S. drummondii	(T)10	4523	4532	10	non-coding	LSC
S. drummondii	(T)10	6793	6802	10	non-coding	LSC

S. drummondii	(A)10	9075	9084	10	non-coding	LSC
S. drummondii	(T)10	9729	9738	10	trnG-UCC	LSC
S. drummondii	(A)10	10346	10355	10	non-coding	LSC
S. drummondii	(T)12	12083	12094	12	non-coding	LSC
S. drummondii	(A)10	14010	14019	10	non-coding	LSC
S. drummondii	(T)11	19045	19055	11	rpoC2	LSC
S. drummondii	(T)10	27928	27937	10	non-coding	LSC
S. drummondii	(T)10	35990	35999	10	non-coding	LSC
S. drummondii	(A)10	44862	44871	10	ycf3	LSC
S. drummondii	(A)10	45116	45125	10	non-coding	LSC
S. drummondii	(T)10	46004	46013	10	non-coding	LSC
S. drummondii	(A)11	47390	47400	11	non-coding	LSC
S. drummondii	(A)12	47634	47645	12	non-coding	LSC
S. drummondii	(A)10	64411	64420	10	non-coding	LSC
S. drummondii	(T)12	69722	69733	12	non-coding	LSC
S. drummondii	(T)13	75692	75704	13	petB	LSC
S. drummondii	(T)10	76104	76113	10	petB	LSC
S. drummondii	(T)10	78608	78617	10	rpoA	LSC
S. drummondii	(A)10	113024	113033	10	non-coding	SSC
S. drummondii	(T)10	119661	119670	10	ndhA	SSC
S. incana	(A)10	6860	6869	10	non-coding	LSC
S. incana	(A)10	8071	8080	10	non-coding	LSC
S. incana	(A)10	8578	8587	10	non-coding	LSC
S. incana	(A)10	9917	9926	10	trnG-UCC	LSC
S. incana	(T)11	18976	18986	11	rpoC2	LSC
S. incana	(T)10	27816	27825	10	non-coding	LSC
S. incana	(T)10	32112	32121	10	non-coding	LSC
S. incana	(T)10	32445	32454	10	non-coding	LSC
S. incana	(T)10	35867	35876	10	non-coding	LSC
S. incana	(A)10	36004	36013	10	non-coding	LSC
S. incana	(A)12	42684	42695	12	non-coding	LSC
S. incana	(T)13	46796	46808	13	non-coding	LSC
S. incana	(A)13	46925	46937	13	non-coding	LSC
S. incana	(A)11	47583	47593	11	trnL-UAA	LSC
S. incana	(A)10	57323	57332	10	non-coding	LSC
S. incana	(T)10	69067	69076	10	non-coding	LSC
S. incana	(A)10	74003	74012	10	non-coding	LSC
S. incana	(T)10	77955	77964	10	rpoA	LSC
S. incana	(T)10	79852	79861	10	non-coding	LSC
S. incana	(T)10	83450	83459	10	rpl22	LSC
S. incana	(T)11	83824	83834	11	non-coding	IRB
S. incana	(A)11	151569	151579	11	non-coding	IRA
S. mexicana	(T)10	8412	8421	10	non-coding	LSC
S. mexicana	(A)10	8646	8655	10	non-coding	LSC
S. mexicana	(T)11	19064	19074	11	rpoC2	LSC
S. mexicana	(T)10	32623	32632	10	non-coding	LSC
S. mexicana	(T)10	36069	36078	10	non-coding	LSC
S. mexicana	(A)12	42890	42901	12	non-coding	LSC
S. mexicana	(A)10	44911	44920	10	ycf3	LSC
S. mexicana	(A)11	45155	45165	11	non-coding	LSC
S. mexicana	(A)11	47721	47731	11	non-coding	LSC
S. mexicana	(A)10	48383	48392	10	trnL-UAA	LSC
S. mexicana	(A)10	64537	64546	10	non-coding	LSC
S. mexicana	(T)13	69860	69872	13	non-coding	LSC
S. mexicana	(A)10	71945	71954	10	clpP	LSC
S. mexicana	(T)10	75852	75861	10	petB	LSC
S. mexicana	(T)11	76260	76270	11	petB	LSC
S. mexicana	(T)10	78767	78776	10	rpoA	LSC
S. mexicana	(T)10	80663	80672	10	non-coding	LSC
S. mexicana	(T)17	81180	81196	17	non-coding	LSC
S. mexicana	(T)10	102775	102784	10	trnI-GAU	IRB
S. mexicana	(A)11	112937	112947	11	non-coding	SSC
S. mexicana	(T)11	113373	113383	11	non-coding	SSC
S. mexicana	(T)12	125441	125452	12	ycf1	SSC
S. mexicana	(A)10	134370	134379	10	trnI-GAU	IRA
S. siphocampyloides	(T)10	13413	13422	10	non-coding	LSC
S. siphocampyloides	(T)11	19018	19028	11	rpoC2	LSC
S. siphocampyloides	(T)10	35955	35964	10	petB	LSC
S. siphocampyloides	(A)10	42772	42781	10	non-coding	LSC
S. siphocampyloides	(T)10	43831	43840	10	ycf3	LSC
S. siphocampyloides	(A)13	45022	45034	13	non-coding	LSC
S. siphocampyloides	(T)10	47446	47455	10	non-coding	LSC
S. siphocampyloides	(T)11	69696	69706	11	non-coding	LSC
S. siphocampyloides	(T)10	75694	75703	10	petB	LSC
S. siphocampyloides	(T)10	78608	78617	10	rpoA	LSC
S. siphocampyloides	(T)10	80505	80514	10	non-coding	LSC
S. siphocampyloides	(T)15	81023	81037	15	non-coding	LSC
S. siphocampyloides	(T)10	112029	112038	10	non-coding	SSC
S. siphocampyloides	(T)11	119726	119736	11	ndhA	SSC

Appendix 8. The result of nucleotide diversity analysis of 50 *Scutellaria* chloroplast genomes.

Window	Midpoint	Pi	Theta	S
1-810	510	0.00962	0.01898	51
411-1010	710	0.00413	0.00819	22
611-1210	910	0.00357	0.00595	16
811-1410	1110	0.00529	0.0067	18
1011-1624	1310	0.00677	0.0134	36
1211-1873	1510	0.01126	0.03051	82
1411-2111	1773	0.01775	0.03944	106
1625-2317	1986	0.01915	0.03833	103
1874-2517	2217	0.0148	0.02493	67
2112-2717	2417	0.01072	0.02195	59
2318-2923	2617	0.01056	0.02084	56
2518-3147	2823	0.0129	0.02381	64
2718-3347	3047	0.01292	0.0227	61
2924-3565	3247	0.01427	0.0253	68
3148-3784	3456	0.01713	0.02791	75
3348-4017	3671	0.01799	0.03014	81
3566-4218	3917	0.01956	0.02902	78
3785-4418	4118	0.01374	0.0227	61
4018-4716	4318	0.02873	0.03944	106
4219-4960	4576	0.03488	0.05098	137
4419-5280	4844	0.04405	0.06177	166
4717-5563	5098	0.02986	0.04912	132
4961-5771	5454	0.02093	0.03386	91
5281-6097	5663	0.01626	0.02791	75
5564-6311	5987	0.01423	0.02121	57
5772-6528	6203	0.01749	0.02865	77
6098-6734	6427	0.01331	0.02381	64
6312-7574	6628	0.0132	0.02642	71
6529-7924	6906	0.01277	0.02977	80
6735-8156	7698	0.01386	0.03126	84
7575-8371	8042	0.01177	0.02716	73
7925-8571	8256	0.00865	0.01674	45
8157-8845	8471	0.01335	0.02344	63
8372-9077	8695	0.0144	0.02493	67
8572-9314	8968	0.01616	0.02791	75
8846-9555	9179	0.01601	0.0253	68
9078-9961	9439	0.02112	0.03535	95
9315-10226	9848	0.01829	0.03423	92
9556-10443	10118	0.01486	0.03126	84
9962-10659	10336	0.01244	0.02419	65
10227-10895	10550	0.01236	0.02121	57
10444-11193	10795	0.01736	0.02567	69
10660-11415	11070	0.01355	0.02009	54
10896-11615	11315	0.01277	0.02009	54
11194-11815	11515	0.0056	0.01079	29
11416-12015	11715	0.00656	0.00967	26
11616-12215	11915	0.00787	0.01005	27
11816-12415	12115	0.00741	0.01005	27
12016-12615	12315	0.00607	0.01153	31
12216-12815	12515	0.00454	0.01005	27
12416-13056	12715	0.00598	0.01228	33
12616-13256	12922	0.00562	0.01228	33
12816-13456	13156	0.00687	0.01563	42
13057-13672	13356	0.00963	0.016	43
13257-13918	13562	0.01353	0.0227	61
13457-14124	13797	0.01276	0.02195	59
13673-14325	14018	0.00959	0.02084	56
13919-14592	14224	0.01016	0.01935	52
14125-14794	14456	0.01031	0.0186	50
14326-15012	14694	0.01161	0.01749	47
14593-15273	14894	0.01231	0.01935	52
14795-15519	15123	0.01479	0.02493	67
15013-15751	15406	0.01867	0.03423	92
15274-15963	15643	0.0172	0.03051	82
15520-16163	15861	0.01457	0.0227	61
15752-16363	16063	0.00838	0.0134	36
15964-16563	16263	0.00431	0.00744	20
16164-16769	16463	0.00508	0.00967	26
16364-16999	16663	0.00838	0.01414	38
16564-17254	16879	0.00948	0.01526	41
16770-17454	17154	0.0112	0.01823	49
17000-17654	17354	0.00764	0.01265	34
17255-17859	17554	0.00659	0.0134	36
17455-18096	17754	0.00721	0.01302	35
17655-18311	17966	0.00904	0.01563	42
17860-18511	18211	0.00918	0.01377	37
18097-18711	18411	0.00633	0.01005	27
18312-18911	18611	0.00642	0.00967	26
18512-19111	18811	0.00929	0.01563	42
18712-19323	19011	0.01104	0.02009	54
18912-19523	19217	0.00987	0.01935	52
19112-19723	19423	0.00843	0.01674	45
19324-19923	19623	0.00572	0.01302	35
19524-20123	19823	0.00605	0.0134	36

19724–20323	20023	0.00551	0.01228	33
19924–20526	20223	0.00664	0.01377	37
20124–20726	20424	0.00857	0.01823	49
20324–20926	20626	0.00826	0.01712	46
20527–21126	20826	0.00787	0.01637	44
20727–21326	21026	0.00583	0.01228	33
20927–21526	21226	0.00594	0.01228	33
21127–21726	21426	0.00597	0.01191	32
21327–21926	21626	0.00577	0.00856	23
21527–22126	21826	0.0065	0.0093	25
21727–22327	22026	0.00666	0.01116	30
21927–22537	22226	0.01015	0.01749	47
22127–22743	22437	0.00915	0.0186	50
22328–22943	22637	0.00863	0.01637	44
22538–23143	22843	0.00481	0.01302	35
22744–23343	23043	0.00451	0.01042	28
22944–23543	23243	0.00363	0.00781	21
23144–23743	23443	0.00232	0.00484	13
23344–23943	23643	0.00422	0.00744	20
23544–24194	23843	0.00544	0.00893	24
23744–24474	24066	0.00892	0.01712	46
23944–24684	24325	0.00896	0.01972	53
24195–24891	24574	0.00998	0.0227	61
24475–25091	24791	0.00657	0.01526	41
24685–25291	24991	0.00341	0.00893	24
24892–25491	25191	0.00375	0.00819	22
25092–25694	25391	0.00434	0.00707	19
25292–25993	25591	0.00521	0.00744	20
25492–26200	25794	0.00444	0.00521	14
25695–26404	26098	0.00431	0.00744	20
25994–26604	26304	0.00498	0.00856	23
26202–26804	26504	0.00443	0.01079	29
26405–27004	26704	0.00443	0.01042	28
26605–27204	26904	0.00392	0.01042	28
26805–27404	27104	0.00556	0.00893	24
27005–27604	27304	0.00646	0.00781	21
27205–27804	27504	0.00758	0.00819	22
27405–28004	27704	0.00532	0.00781	21
27605–28204	27904	0.00455	0.00744	20
27805–28404	28104	0.0027	0.00595	16
28005–28604	28304	0.00306	0.0067	18
28205–28837	28504	0.00601	0.0134	36
28405–29077	28719	0.01441	0.02456	66
28605–29340	28959	0.02109	0.03126	84
28838–29552	29195	0.02379	0.03349	90
29078–29766	29452	0.01759	0.02642	71
29341–30466	29657	0.01192	0.02121	57
29553–30728	29873	0.00856	0.01749	47
29767–30950	30597	0.00853	0.01898	51
30467–31276	30828	0.01882	0.02642	71
30729–31516	31056	0.02598	0.03498	94
30951–32011	31388	0.02861	0.03684	99
31277–32257	31639	0.01981	0.03088	83
31517–32464	32157	0.01356	0.02344	63
32012–32688	32357	0.01397	0.02233	60
32258–32902	32588	0.01541	0.0227	61
32465–33119	32800	0.0143	0.02121	57
32689–33429	33019	0.00968	0.01712	46
32903–33682	33219	0.01141	0.02121	57
33120–33909	33553	0.01259	0.02233	60
33430–34170	33801	0.01664	0.02679	72
33683–34400	34062	0.01522	0.0253	68
33910–34646	34285	0.01985	0.03684	99
34171–34883	34516	0.01845	0.03795	102
34401–35096	34748	0.02078	0.03907	105
34647–35299	34989	0.01418	0.02456	66
34884–35499	35198	0.01157	0.01749	47
35097–35699	35399	0.00632	0.01079	29
35300–35899	35599	0.00511	0.00856	23
35500–36099	35799	0.00416	0.00707	19
35700–36299	35999	0.00158	0.00447	12
35900–36499	36199	0.0013	0.00521	14
36100–36699	36399	0.00269	0.00595	16
36300–36899	36599	0.0031	0.00595	16
36500–37099	36799	0.00425	0.00595	16
36700–37299	36999	0.00471	0.00781	21
36900–37499	37199	0.00584	0.00893	24
37100–37699	37399	0.00501	0.0093	25
37300–37914	37599	0.00488	0.01042	28
37500–38133	37807	0.00581	0.01228	33
37700–38373	38033	0.0101	0.02046	55
37915–38575	38244	0.00926	0.01898	51
38134–38843	38475	0.01376	0.02716	73
38374–39073	38696	0.01244	0.0227	61
38576–39292	38943	0.01676	0.02791	75
38844–39497	39191	0.01178	0.01898	51
39074–39697	39397	0.00856	0.01451	39
39293–39907	39597	0.00389	0.00893	24
39498–40107	39807	0.00257	0.00781	21
39698–40307	40007	0.00192	0.00633	17

39908-40507	40207	0.00213	0.00521	14
40108-40707	40407	0.00194	0.00447	12
40308-40907	40607	0.00412	0.00744	20
40508-41118	40807	0.00509	0.00781	21
40708-41318	41007	0.0065	0.01005	27
40908-41518	41218	0.0057	0.00967	26
41119-41718	41418	0.00434	0.00967	26
41319-41918	41618	0.00322	0.00781	21
41519-42119	41818	0.0019	0.00484	13
41719-42319	42018	0.0024	0.00372	10
41919-42519	42219	0.00233	0.00372	10
42120-42719	42419	0.00341	0.00558	15
42320-42922	42619	0.00337	0.00967	26
42520-43129	42819	0.00706	0.02902	78
42720-43337	43026	0.01006	0.04688	126
42923-43537	43237	0.00993	0.04577	123
43130-43737	43437	0.00602	0.02679	72
43338-43937	43637	0.00302	0.0093	25
43538-44137	43837	0.00422	0.00781	21
43738-44337	44037	0.00432	0.00707	19
43938-44546	44237	0.00443	0.00781	21
44138-44825	44446	0.00782	0.01377	37
44338-45147	44648	0.01098	0.0186	50
44547-45359	44954	0.01042	0.01712	46
44826-45617	45247	0.0099	0.01637	44
45148-45831	45506	0.00698	0.01377	37
45360-46070	45717	0.00714	0.016	43
45618-46270	45951	0.00293	0.00856	23
45832-46477	46170	0.00442	0.00856	23
46071-46678	46372	0.0051	0.00893	24
46271-46888	46577	0.00705	0.01377	37
46478-47098	46788	0.00565	0.01265	34
46679-48094	46998	0.00618	0.01265	34
46889-48317	47988	0.00575	0.01265	34
47099-48548	48209	0.00776	0.01637	44
48095-48748	48448	0.00929	0.01749	47
48318-48948	48648	0.0087	0.01488	40
48549-49174	48848	0.00771	0.01377	37
48749-49480	49062	0.00801	0.016	43
48949-49733	49354	0.01816	0.04316	116
49175-50058	49612	0.02253	0.05098	137
49481-50278	49950	0.02408	0.05023	135
49734-50496	50161	0.01549	0.0253	68
50059-50726	50378	0.01373	0.02009	54
50279-50963	50619	0.01194	0.02046	55
50497-51204	50826	0.01457	0.02456	66
50727-51407	51094	0.0168	0.02605	70
50964-51666	51304	0.01985	0.03237	87
51205-51890	51565	0.01645	0.02977	80
51408-52099	51767	0.01172	0.02567	69
51667-52299	51999	0.00722	0.01488	40
51891-52512	52199	0.00693	0.0134	36
52100-52713	52399	0.00852	0.01377	37
52300-52925	52613	0.00652	0.01228	33
52513-53125	52825	0.00527	0.01005	27
52714-53327	53025	0.00182	0.00595	16
52926-53527	53225	0.00252	0.00744	20
53126-53745	53427	0.00584	0.01042	28
53328-54006	53627	0.01438	0.02158	58
53528-54824	53866	0.01995	0.03498	94
53746-55027	54724	0.01722	0.0346	93
54007-55228	54927	0.01205	0.02902	78
54825-55443	55127	0.00875	0.01712	46
55028-55662	55343	0.00804	0.01488	40
55229-55892	55562	0.00793	0.01526	41
55445-56092	55786	0.00598	0.01451	39
55663-56292	55992	0.00667	0.01377	37
55893-56492	56192	0.00441	0.00856	23
56093-56692	56392	0.00369	0.0067	18
56293-56892	56592	0.00457	0.00707	19
56493-57092	56792	0.00553	0.00893	24
56693-57292	56992	0.00596	0.00893	24
56893-57492	57192	0.00422	0.00819	22
57093-57692	57392	0.00491	0.01042	28
57293-57920	57592	0.00846	0.01563	42
57493-58178	57806	0.01103	0.02084	56
57693-58407	58040	0.01221	0.02419	65
57932-58641	58299	0.00818	0.01972	53
58180-58841	58531	0.0057	0.01414	38
58408-59041	58741	0.00342	0.00744	20
58642-59241	58941	0.0052	0.00744	20
58842-59441	59141	0.00705	0.0093	25
59042-59646	59341	0.00863	0.01079	29
59242-59854	59546	0.0103	0.02233	60
59442-60054	59754	0.00985	0.02381	64
59647-60280	59954	0.00992	0.02567	69
59855-60515	60180	0.01456	0.02121	57
60055-60806	60392	0.02074	0.02679	72
60281-61056	60682	0.02288	0.03014	81
60516-61256	60914	0.01736	0.02642	71

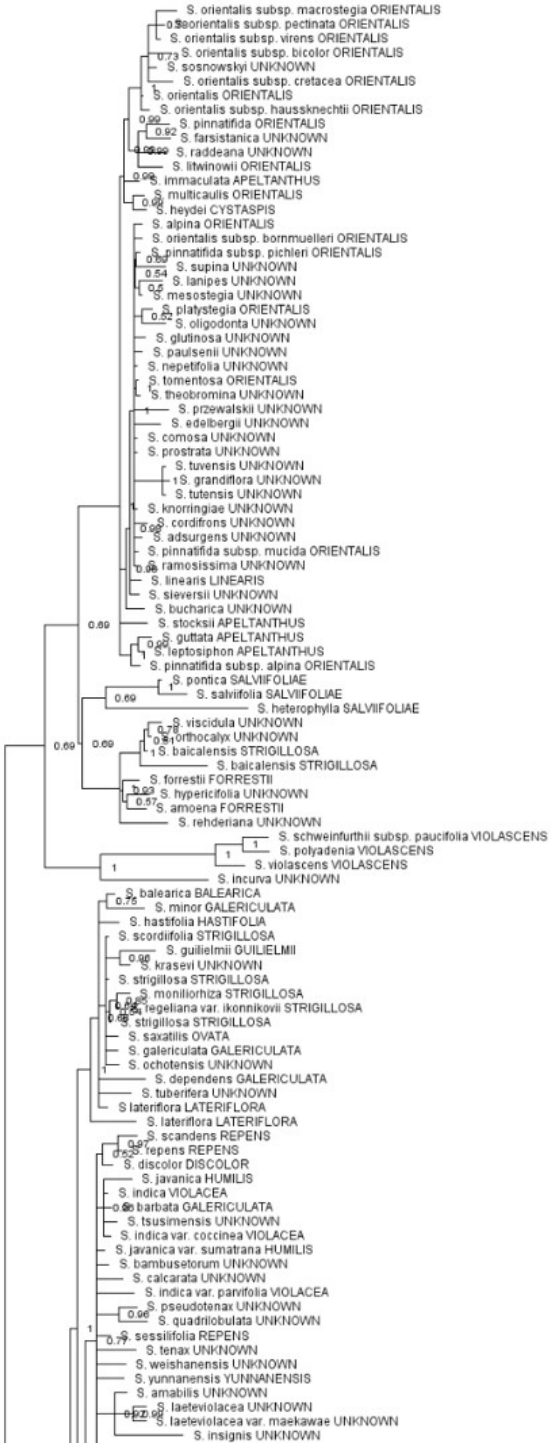
60807–61456	61156	0.01296	0.02195	59
61057–61656	61356	0.01001	0.01674	45
61257–61857	61556	0.00828	0.016	43
61457–62057	61757	0.00669	0.01302	35
61657–62257	61957	0.0051	0.00967	26
61858–62971	62157	0.00714	0.01005	27
62058–63185	62357	0.00924	0.01526	41
62258–63407	63084	0.01241	0.02344	63
62972–63640	63290	0.01346	0.02493	67
63186–63840	63515	0.01114	0.01898	51
63408–64040	63740	0.00836	0.01302	35
63641–64240	63940	0.00815	0.01191	32
63841–64461	64140	0.01372	0.0227	61
64041–64801	64346	0.01697	0.02642	71
64241–65016	64700	0.0178	0.02902	78
64462–65222	64914	0.01242	0.02158	58
64802–65422	65116	0.01149	0.01935	52
65017–65622	65322	0.00749	0.01377	37
65223–65831	65522	0.00736	0.01265	34
65423–66031	65729	0.00701	0.01153	31
65623–66231	65931	0.00665	0.01191	32
65832–66432	66131	0.00624	0.01079	29
66032–66632	66332	0.00419	0.01079	29
66232–66833	66532	0.00631	0.0134	36
66433–67199	66732	0.00977	0.02046	55
66633–67479	66991	0.02646	0.03349	90
66834–67770	67316	0.02867	0.03721	100
67200–67977	67653	0.02574	0.03163	85
67480–68179	67877	0.01037	0.01972	53
67771–68379	68077	0.00711	0.0134	36
67978–68579	68279	0.00577	0.01116	30
68180–68779	68479	0.00396	0.00781	21
68380–68985	68679	0.00735	0.01488	40
68580–69556	68880	0.01283	0.02493	67
68780–69757	69103	0.01458	0.03051	82
68986–69963	69657	0.01173	0.02753	74
69557–70178	69861	0.01005	0.02381	64
69758–70412	70073	0.01147	0.02456	66
69964–70618	70311	0.01293	0.02381	64
70179–70835	70514	0.01035	0.01935	52
70413–71068	70726	0.01643	0.0227	61
70619–71343	70961	0.01814	0.02567	69
70836–71549	71217	0.01754	0.02567	69
71069–71812	71447	0.01032	0.02084	56
71344–72012	71710	0.00559	0.01414	38
71550–72241	71912	0.00884	0.02121	57
71813–72449	72118	0.01043	0.02158	58
72013–72649	72349	0.01185	0.02307	62
72242–72870	72549	0.00954	0.0186	50
72450–73088	72755	0.01123	0.02084	56
72650–73292	72975	0.01243	0.0227	61
72871–73510	73188	0.01426	0.02419	65
73089–73738	73399	0.01051	0.02009	54
73293–73941	73610	0.00887	0.0186	50
73511–74159	73841	0.00716	0.01637	44
73739–74405	74051	0.0085	0.02046	55
73942–74622	74272	0.00866	0.01972	53
74160–74825	74511	0.00683	0.01563	42
74406–75029	74725	0.006	0.01153	31
74623–75268	74925	0.0096	0.016	43
74826–75478	75138	0.0128	0.0227	61
75030–75684	75368	0.01376	0.02493	67
75269–75965	75583	0.01374	0.02419	65
75479–76277	75787	0.01419	0.02344	63
75685–76478	76070	0.01127	0.01786	48
75966–76678	76378	0.00797	0.01265	34
76278–76878	76578	0.00519	0.00781	21
76479–77078	76778	0.00538	0.00893	24
76679–77278	76978	0.00485	0.00967	26
76879–77478	77178	0.00426	0.0093	25
77079–77678	77378	0.00373	0.00893	24
77279–77901	77578	0.00454	0.00967	26
77479–78123	77778	0.00658	0.01191	32
77679–78343	78023	0.00785	0.0134	36
77902–78543	78243	0.0072	0.0134	36
78124–78748	78443	0.00773	0.01414	38
78344–78949	78643	0.00824	0.01712	46
78544–79163	78849	0.00922	0.02121	57
78749–79373	79063	0.0119	0.02456	66
78950–79587	79273	0.0134	0.0253	68
79164–79787	79486	0.01205	0.01972	53
79374–79987	79687	0.00724	0.01302	35
79588–80187	79887	0.00429	0.00781	21
79788–80398	80087	0.00571	0.01265	34
79988–80615	80298	0.01825	0.02046	55
80188–80865	80498	0.01937	0.02307	62
80399–81112	80754	0.02174	0.02456	66
80616–81312	80981	0.0102	0.0186	50
80866–81512	81212	0.00903	0.016	43
81113–81719	81412	0.00617	0.01079	29

81313-81928	81612	0.00861	0.016	43
81513-82168	81826	0.01413	0.03014	81
81720-82368	82068	0.01296	0.0294	79
81929-82568	82268	0.01281	0.02753	74
82169-82768	82468	0.01036	0.02046	55
82369-82968	82668	0.0128	0.0253	68
82569-83168	82868	0.00948	0.02233	60
82769-83481	83068	0.00628	0.01674	45
82969-83688	83268	0.00829	0.01674	45
83169-83891	83582	0.00845	0.01637	44
83482-84101	83791	0.01149	0.01823	49
83689-84306	83991	0.00955	0.01563	42
83892-84506	84206	0.01085	0.01786	48
84102-84753	84406	0.0117	0.02158	58
84307-84953	84638	0.00913	0.0186	50
84507-85181	84853	0.01159	0.01786	48
84754-85383	85053	0.00889	0.01302	35
84954-85592	85283	0.00839	0.01377	37
85182-85904	85483	0.00745	0.016	43
85384-86104	85702	0.01093	0.01972	53
85593-86349	86004	0.01529	0.02753	74
85905-86566	86236	0.01343	0.02642	71
86105-86789	86454	0.01502	0.02679	72
86350-86989	86683	0.0121	0.0186	50
86567-87189	86889	0.01345	0.01935	52
86790-87389	87089	0.01064	0.01786	48
86990-87601	87289	0.01588	0.02567	69
87190-87801	87501	0.01426	0.02307	62
87390-88012	87701	0.01214	0.02233	60
87602-88215	87912	0.00734	0.01823	49
87802-88415	88112	0.00516	0.01377	37
88013-88615	88315	0.00306	0.00819	22
88216-88821	88515	0.00059	0.00223	6
88416-89022	88715	0.00215	0.00335	9
88616-89222	88921	0.00196	0.0026	7
88822-89422	89122	0.00189	0.0026	7
89023-89622	89322	0.00013	0.00074	2
89223-89822	89522	0.0002	0.00112	3
89423-90022	89722	0.00033	0.00149	4
89623-90233	89922	0.00069	0.00186	5
89823-90445	90133	0.00157	0.00186	5
90023-90645	90333	0.00382	0.00335	9
90234-90845	90545	0.00346	0.00298	8
90446-91045	90745	0.0027	0.00298	8
90646-91245	90945	0.00089	0.00223	6
90846-91450	91145	0.00148	0.00484	13
91046-91663	91345	0.00242	0.01042	28
91246-91863	91557	0.00223	0.01005	27
91451-92063	91763	0.00196	0.00856	23
91664-92263	91963	0.00096	0.0026	7
91864-92463	92163	0.00161	0.00223	6
92064-92663	92363	0.00128	0.00112	3
92264-92863	92563	0.00172	0.00186	5
92464-93063	92763	0.00089	0.00186	5
92664-93272	92963	0.00269	0.00372	10
92864-93472	93172	0.00256	0.00298	8
93064-93672	93372	0.0029	0.00335	9
93273-93872	93572	0.00137	0.00223	6
93473-94072	93772	0.00093	0.00223	6
93673-94272	93972	0.00091	0.00298	8
93873-94472	94172	0.00095	0.0026	7
94073-94681	94372	0.00115	0.00372	10
94273-94881	94572	0.00084	0.00298	8
94473-95081	94781	0.00157	0.00335	9
94682-95281	94981	0.00151	0.00298	8
94882-95499	95181	0.00196	0.0026	7
95082-95699	95399	0.00112	0.00298	8
95282-95908	95599	0.00137	0.00335	9
95500-96108	95799	0.00065	0.0026	7
95700-96308	96008	0.00102	0.00298	8
95909-96535	96208	0.00169	0.00633	17
96109-96735	96408	0.00176	0.0067	18
96309-96941	96635	0.00119	0.00521	14
96536-97141	96841	0.00013	0.00074	2
96736-97341	97041	0.00043	0.00074	2
96942-97554	97241	0.00056	0.00149	4
97142-97754	97441	0.00069	0.00223	6
97342-97954	97654	0.00066	0.00298	8
97555-98176	97854	0.00162	0.00372	10
97755-98376	98067	0.00207	0.00521	14
97955-98576	98276	0.00297	0.00521	14
98177-98776	98476	0.00194	0.00409	11
98377-98976	98676	0.00129	0.00149	4
98577-99176	98876	0.00007	0.00037	1
98777-99376	99076	0.00421	0.00298	8
98977-99585	99276	0.0046	0.00447	12
99177-99785	99485	0.00506	0.00484	13
99377-99985	99685	0.00131	0.00223	6
99586-100185	99885	0.00111	0.00112	3
99786-100385	100085	0.00072	0.00112	3

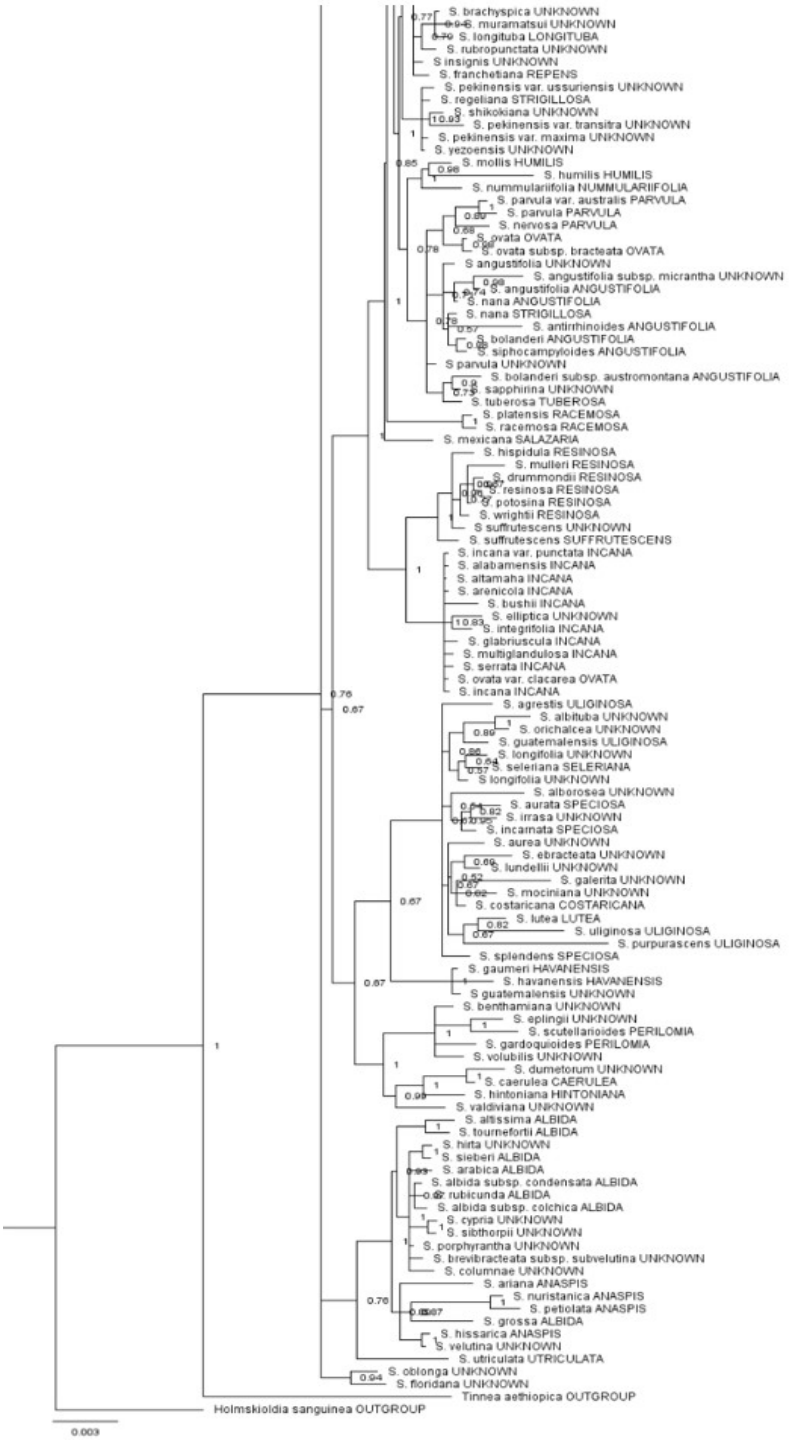
99986–100585	100285	0.00026	0.00074	2
100186–100787	100485	0.00013	0.00074	2
100386–100987	100686	0.00033	0.00149	4
100586–101187	100887	0.00053	0.00223	6
100788–101387	101087	0.00083	0.0026	7
100988–101672	101287	0.00057	0.00149	4
101188–101872	101487	0.00081	0.00149	4
101388–102072	101772	0.00044	0.00074	2
101673–102272	101972	0.0005	0.00112	3
101873–102472	102172	0.0002	0.00074	2
102073–102697	102372	0.00247	0.00372	10
102273–102909	102591	0.00393	0.00521	14
102473–103109	102797	0.00466	0.00558	15
102698–103323	103009	0.00318	0.00372	10
102910–103541	103223	0.0019	0.00223	6
103110–103780	103423	0.0016	0.0026	7
103324–103981	103680	0.00186	0.00223	6
103542–104186	103880	0.0018	0.0026	7
103781–104399	104086	0.00236	0.00558	15
103982–104619	104286	0.00271	0.01116	30
104187–104822	104506	0.00323	0.01265	34
104400–105030	104722	0.00211	0.00856	23
104620–105247	104922	0.00471	0.02233	60
104823–105473	105143	0.00598	0.03051	82
105031–105683	105372	0.00798	0.04093	110
105248–105891	105583	0.00476	0.02381	64
105474–106091	105783	0.00342	0.01377	37
105684–106293	105991	0.00143	0.00335	9
105892–106497	106192	0.00111	0.00186	5
106092–106697	106393	0.00137	0.00298	8
106294–106897	106597	0.00196	0.00484	13
106498–107104	106797	0.00236	0.00409	11
106698–107309	106997	0.00144	0.0026	7
106898–107525	107204	0.00092	0.00112	3
107105–107738	107410	0.00072	0.00074	2
107310–107938	107625	0.00072	0.00074	2
107526–108161	107838	0.00144	0.00149	4
107739–108361	108038	0.00092	0.00149	4
107939–108561	108261	0.00105	0.00223	6
108162–108762	108461	0.00026	0.00112	3
108362–108962	108661	0.00013	0.00074	2
108562–109162	108862	0	0	0
108763–109362	109062	0	0	0
108963–109562	109262	0.00047	0.0026	7
109163–109765	109462	0.00158	0.00484	13
109363–109965	109665	0.00237	0.00856	23
109563–110202	109865	0.00463	0.01972	53
109766–110413	110089	0.00717	0.0346	93
109966–110632	110302	0.00851	0.04279	115
110203–110860	110514	0.00638	0.03237	87
110414–111062	110750	0.00335	0.016	43
110633–111269	110962	0.00361	0.01749	47
110861–111478	111169	0.00641	0.03274	88
111063–111678	111374	0.00671	0.03349	90
111270–111900	111578	0.00463	0.02084	56
111479–112109	111788	0.0056	0.00595	16
111679–112334	112003	0.00702	0.00707	19
111901–112534	112217	0.00734	0.00781	21
112110–112751	112434	0.00547	0.0067	18
112335–112952	112643	0.00414	0.00558	15
112535–113152	112852	0.0037	0.00447	12
112752–113352	113052	0.0028	0.00447	12
112953–113558	113252	0.00693	0.01079	29
113153–113782	113452	0.01996	0.02791	75
113353–113982	113682	0.02837	0.03684	99
113559–114182	113882	0.03004	0.03981	107
113783–114391	114082	0.02511	0.03758	101
113983–114591	114291	0.01959	0.03349	90
114183–114791	114491	0.0161	0.02716	73
114392–114991	114691	0.01348	0.02009	54
114592–115191	114891	0.01284	0.01898	51
114792–115391	115091	0.01446	0.02046	55
114992–115591	115291	0.01066	0.01637	44
115192–115791	115491	0.00973	0.016	43
115392–116298	115691	0.01087	0.02158	58
115592–116515	115971	0.01338	0.02456	66
115792–116920	116415	0.01838	0.03126	84
116299–117129	116627	0.02134	0.032	86
116516–117388	117024	0.02333	0.03498	94
116921–117588	117288	0.0219	0.03274	88
117130–117797	117488	0.01976	0.03014	81
117389–117997	117688	0.01683	0.02679	72
117589–118197	117897	0.01242	0.02009	54
117798–118428	118097	0.02232	0.03423	92
117998–118670	118314	0.02447	0.03572	96
118198–118870	118570	0.02472	0.03907	105
118429–119070	118770	0.00938	0.01898	51
118671–119270	118970	0.00503	0.01563	42
118871–119470	119170	0.00567	0.01637	44
119071–119670	119370	0.00635	0.01637	44

119271–119870	119570	0.00945	0.01749	47
119471–120083	119770	0.01266	0.02084	56
119671–120293	119970	0.01359	0.02233	60
119871–120498	120193	0.01136	0.02158	58
120084–120745	120393	0.01153	0.02233	60
120294–120945	120640	0.00973	0.02084	56
120499–121294	120845	0.01121	0.02381	64
120746–121494	121068	0.00708	0.01786	48
120946–121694	121394	0.00802	0.01972	53
121295–121910	121594	0.01196	0.02046	55
121495–122156	121794	0.02115	0.02902	78
121695–122356	122053	0.02037	0.02642	71
121911–122557	122256	0.01547	0.02307	62
122157–122765	122456	0.00625	0.01563	42
122357–122965	122658	0.00676	0.01823	49
122558–123165	122865	0.00823	0.01749	47
122766–123373	123065	0.00899	0.01712	46
122966–123600	123265	0.01433	0.02419	65
123166–123836	123480	0.0175	0.03088	83
123374–124060	123736	0.01766	0.03014	81
123601–124278	123940	0.01824	0.02865	77
123837–124478	124162	0.01277	0.02195	59
124061–124678	124378	0.01292	0.02307	62
124279–124878	124578	0.00679	0.01674	45
124479–125078	124778	0.00897	0.01674	45
124679–125278	124978	0.01059	0.01712	46
124879–125478	125178	0.01264	0.0186	50
125079–125678	125378	0.01152	0.01972	53
125279–125878	125578	0.00727	0.01414	38
125479–126079	125778	0.0054	0.01265	34
125679–126293	125978	0.00717	0.01414	38
125879–126585	126193	0.01167	0.02158	58
126080–126858	126393	0.01603	0.0294	79
126294–127091	126758	0.01969	0.03498	94
126586–127291	126991	0.02235	0.03758	101
126859–127491	127191	0.02062	0.03163	85
127092–127709	127391	0.02146	0.032	86
127292–127921	127597	0.01887	0.02828	76
127492–128142	127818	0.02493	0.04019	108
127719–128378	128021	0.02363	0.04019	108
127922–128581	128251	0.02496	0.04428	119
128143–128881	128478	0.01892	0.03572	96
128379–129081	128681	0.02076	0.03721	100
128582–129284	128981	0.02299	0.03795	102
128882–129490	129184	0.02562	0.03981	107
129082–129696	129390	0.02647	0.03981	107
129285–129899	129590	0.02176	0.03572	96
129491–130099	129798	0.02095	0.03907	105
129697–130314	129999	0.02324	0.04093	110
129900–130537	130214	0.02548	0.04205	113
130100–130887	130436	0.03277	0.04837	130
130315–131099	130766	0.02962	0.04614	124
130538–131308	130999	0.03308	0.05135	138
130894–131828	131208	0.02823	0.04242	114
131100–132028	131722	0.02029	0.032	86
131309–132228	131928	0.01202	0.01935	52
131829–132428	132128	0.00322	0.0067	18
132029–132636	132328	0.0027	0.00521	14
132229–132846	132529	0.00389	0.00521	14
132429–133053	132746	0.00453	0.0067	18
132637–133277	132946	0.00609	0.00595	16
132847–133493	133163	0.00664	0.00633	17
133054–133703	133381	0.00709	0.00633	17
133278–133907	133603	0.00507	0.00856	23
133494–134112	133803	0.0047	0.02121	57
133704–134319	134012	0.00654	0.03349	90
133908–134531	134219	0.00641	0.03274	88
134113–134767	134420	0.00355	0.01712	46
134320–134979	134645	0.00492	0.02456	66
134532–135190	134867	0.00658	0.03349	90
134768–135416	135079	0.00911	0.04614	124
134980–135616	135312	0.00571	0.02642	71
135191–135819	135516	0.0039	0.01563	42
135417–136019	135719	0.00177	0.00521	14
135617–136219	135919	0.00085	0.00372	10
135820–136419	136119	0.0004	0.00223	6
136020–136620	136319	0	0	0
136220–136820	136519	0	0	0
136420–137020	136720	0.00013	0.00074	2
136621–137241	136920	0.00033	0.00149	4
136821–137441	137120	0.00105	0.00223	6
137021–137654	137341	0.00164	0.00223	6
137242–137869	137541	0.0021	0.00186	5
137442–138075	137754	0.00144	0.00149	4
137655–138282	137970	0.00157	0.00149	4
137870–138482	138175	0.00124	0.00186	5
138076–138682	138382	0.00163	0.00335	9
138283–138886	138582	0.00164	0.00372	10
138483–139087	138782	0.00164	0.00409	11
138683–139288	138986	0.00118	0.00223	6

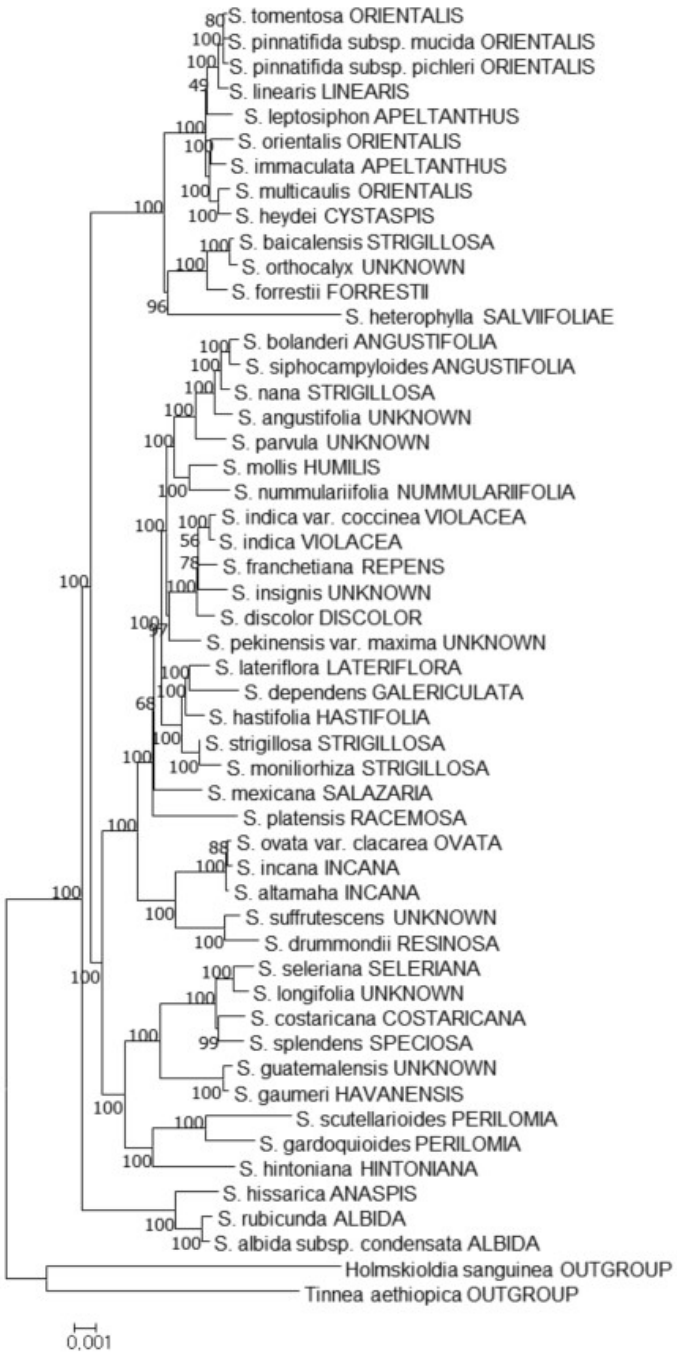
138887–139496	139188	0.00131	0.00335	9
139088–139696	139388	0.00216	0.00744	20
139289–139907	139596	0.00436	0.01898	51
139497–140133	139806	0.00502	0.02456	66
139697–140357	140032	0.00731	0.03684	99
139908–140557	140249	0.00512	0.0253	68
140134–140768	140457	0.00524	0.02493	67
140358–140993	140659	0.00323	0.01265	34
140558–141193	140877	0.00336	0.01302	35
140769–141399	141093	0.00269	0.0067	18
140994–141599	141293	0.00124	0.00149	4
141194–141858	141499	0.00167	0.00223	6
141400–142058	141740	0.00081	0.00149	4
141600–142274	141958	0.0016	0.0026	7
141859–142480	142174	0.00233	0.00298	8
142059–142681	142374	0.00539	0.00633	17
142275–142905	142580	0.00473	0.00558	15
142481–143105	142805	0.00344	0.00409	11
142687–143305	143005	0.0002	0.00074	2
142906–143505	143205	0.0005	0.00112	3
143106–143790	143405	0.0005	0.00112	3
143306–143990	143605	0.00044	0.00074	2
143506–144190	143890	0.00044	0.00112	3
143791–144390	144090	0.00064	0.00186	5
143991–144591	144290	0.0009	0.00298	8
144191–144792	144490	0.00046	0.00186	5
144391–144992	144692	0.00033	0.00149	4
144592–145192	144892	0.00007	0.00037	1
144793–145392	145092	0.00026	0.00074	2
144993–145592	145292	0.00065	0.00074	2
145193–145792	145492	0.00117	0.00149	4
145393–146001	145692	0.00224	0.00335	9
145593–146201	145892	0.00506	0.00484	13
145793–146401	146101	0.00453	0.00409	11
146002–146601	146301	0.00327	0.00186	5
146202–146801	146501	0.00057	0.00074	2
146402–147001	146701	0.00129	0.00149	4
146602–147210	146901	0.00297	0.00521	14
146802–147423	147101	0.00253	0.00521	14
147002–147623	147323	0.00214	0.00558	15
147211–147836	147523	0.00092	0.00372	10
147424–148036	147723	0.00079	0.00298	8
147624–148236	147936	0.00089	0.0026	7
147837–148436	148136	0.00043	0.00074	2
148037–148642	148336	0.00049	0.00112	3
148237–148875	148542	0.00067	0.00372	10
148437–149075	148742	0.00116	0.00447	12
148643–149275	148975	0.0013	0.00521	14
148876–149475	149175	0.0009	0.00298	8
149076–149684	149375	0.00072	0.00298	8
149276–149884	149584	0.00137	0.00335	9
149476–150102	149784	0.00124	0.0026	7
149685–150302	150002	0.00125	0.00298	8
149885–150502	150202	0.00137	0.00223	6
150103–150711	150402	0.00171	0.00409	11
150303–150911	150602	0.00138	0.00298	8
150503–151111	150811	0.00077	0.00298	8
150712–151311	151011	0.00089	0.00223	6
150912–151511	151211	0.00095	0.0026	7
151112–151711	151411	0.00084	0.0026	7
151312–151911	151611	0.00093	0.00223	6
151512–152111	151811	0.0015	0.00298	8
151712–152320	152011	0.00297	0.00372	10
151912–152520	152220	0.00275	0.00335	9
152112–152720	152420	0.00255	0.00298	8
152321–152920	152620	0.00113	0.00186	5
152521–153120	152820	0.00172	0.00186	5
152721–153320	153020	0.00128	0.00112	3
152921–153520	153220	0.00161	0.00223	6
153121–153726	153420	0.0009	0.0026	7
153321–153938	153620	0.00236	0.01079	29
153521–154138	153832	0.00236	0.01079	29
153727–154338	154038	0.00229	0.01005	27
153939–154538	154238	0.00108	0.0026	7
154139–154738	154438	0.00161	0.00186	5
154339–154950	154638	0.0027	0.00298	8
154539–155150	154838	0.00346	0.00298	8
154739–155361	155050	0.00279	0.00298	8
154951–155561	155261	0.00164	0.00223	6
155151–155761	155461	0.00069	0.00186	5
155362–155961	155661	0.00027	0.00149	4
155562–156161	155861	0.00013	0.00074	2
155762–156362	156061	0.00163	0.00186	5
155962–156568	156261	0.00183	0.00223	6
156162–156768	156462	0.00196	0.0026	7
156363–156968	156668	0.00085	0.00298	8
156569–157174	156868	0.00134	0.00372	10
156769–157198	156972	0.00188	0.00547	10



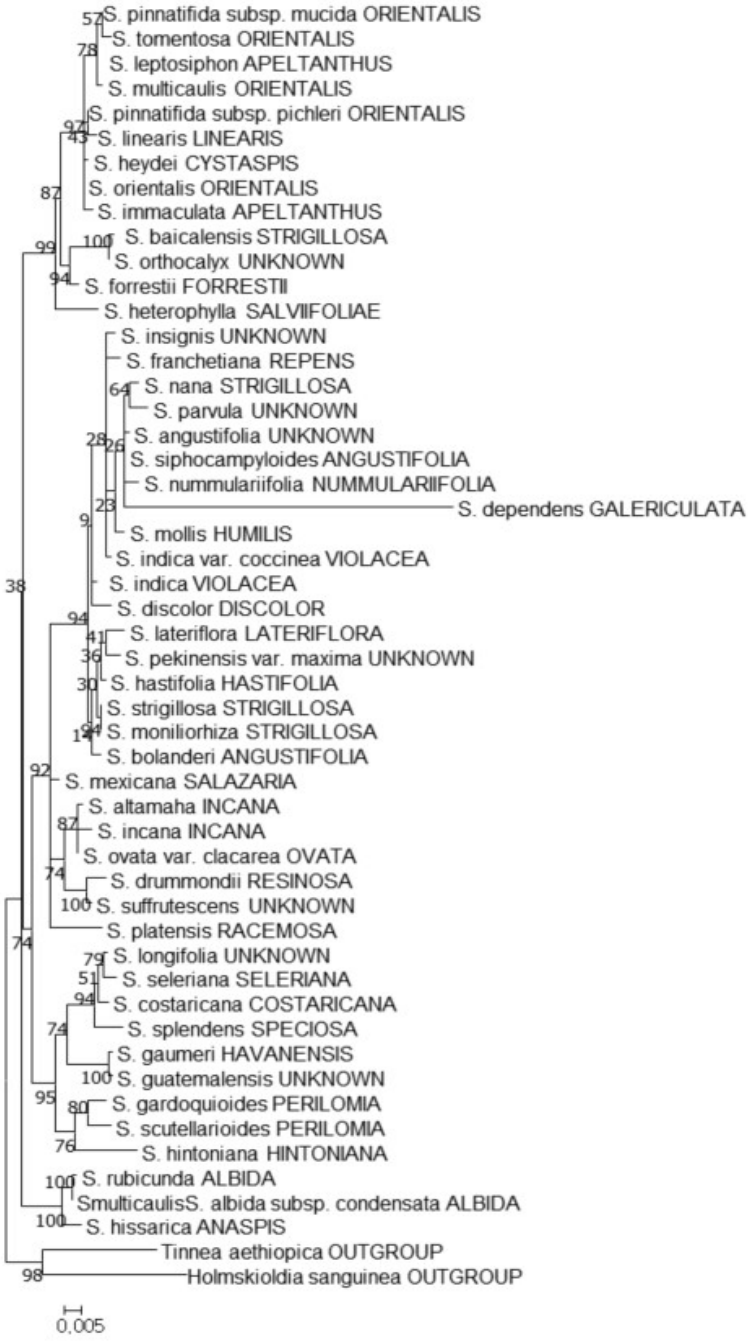
Appendix 9. A super-matrix tree based on Bayesian inference analysis.



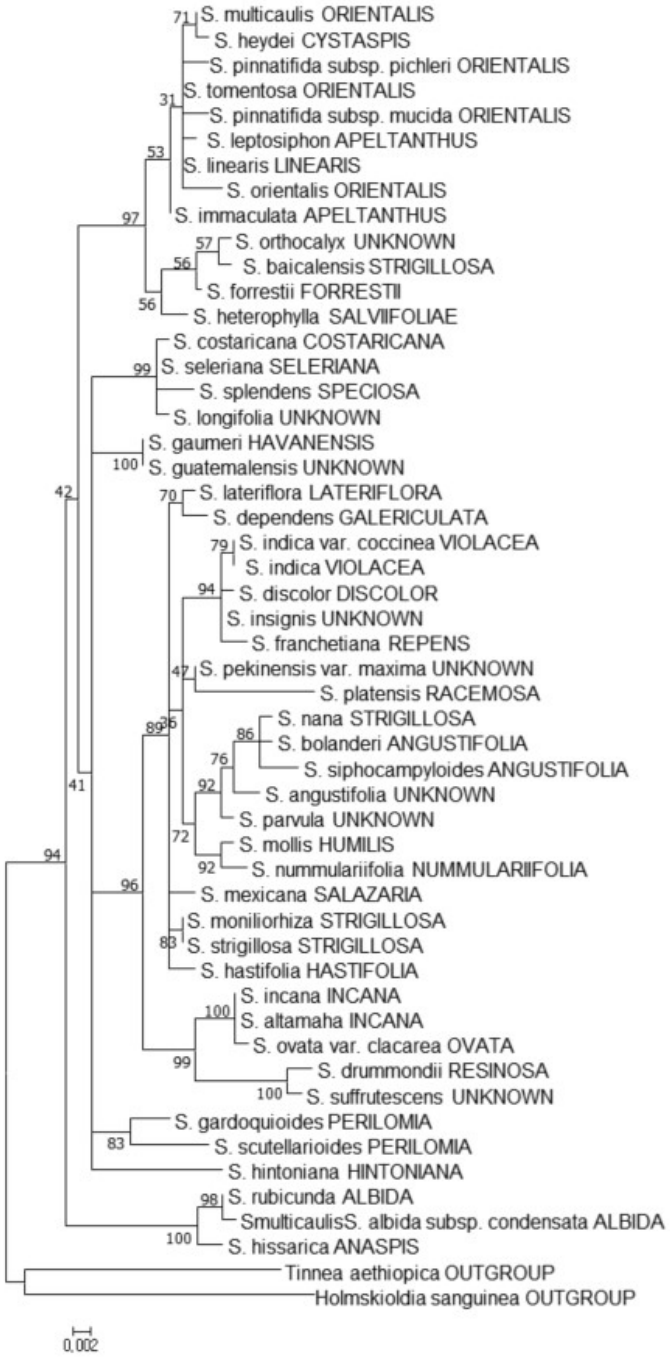
Appendix 9. Continued.



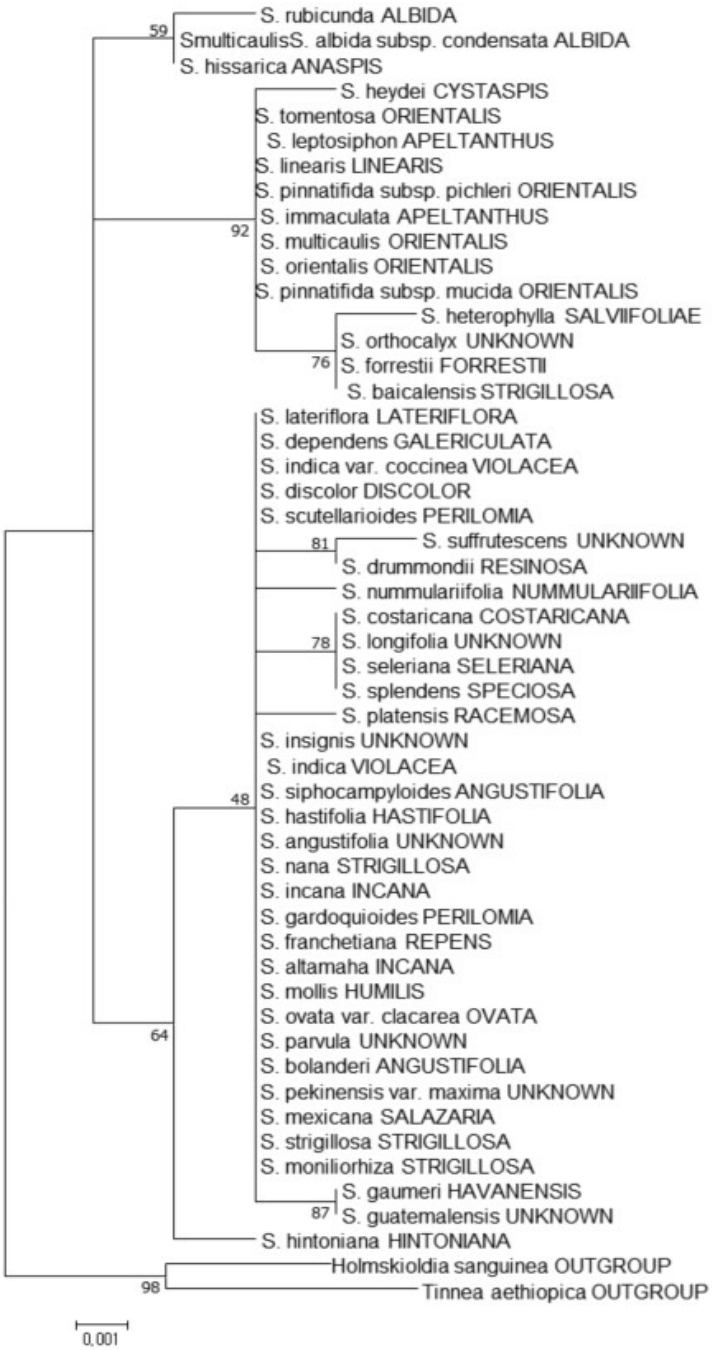
Appendix 10. A phylogenetic tree for 52 taxa used for quartet distance analysis. Based on the complete chloroplast genomes.



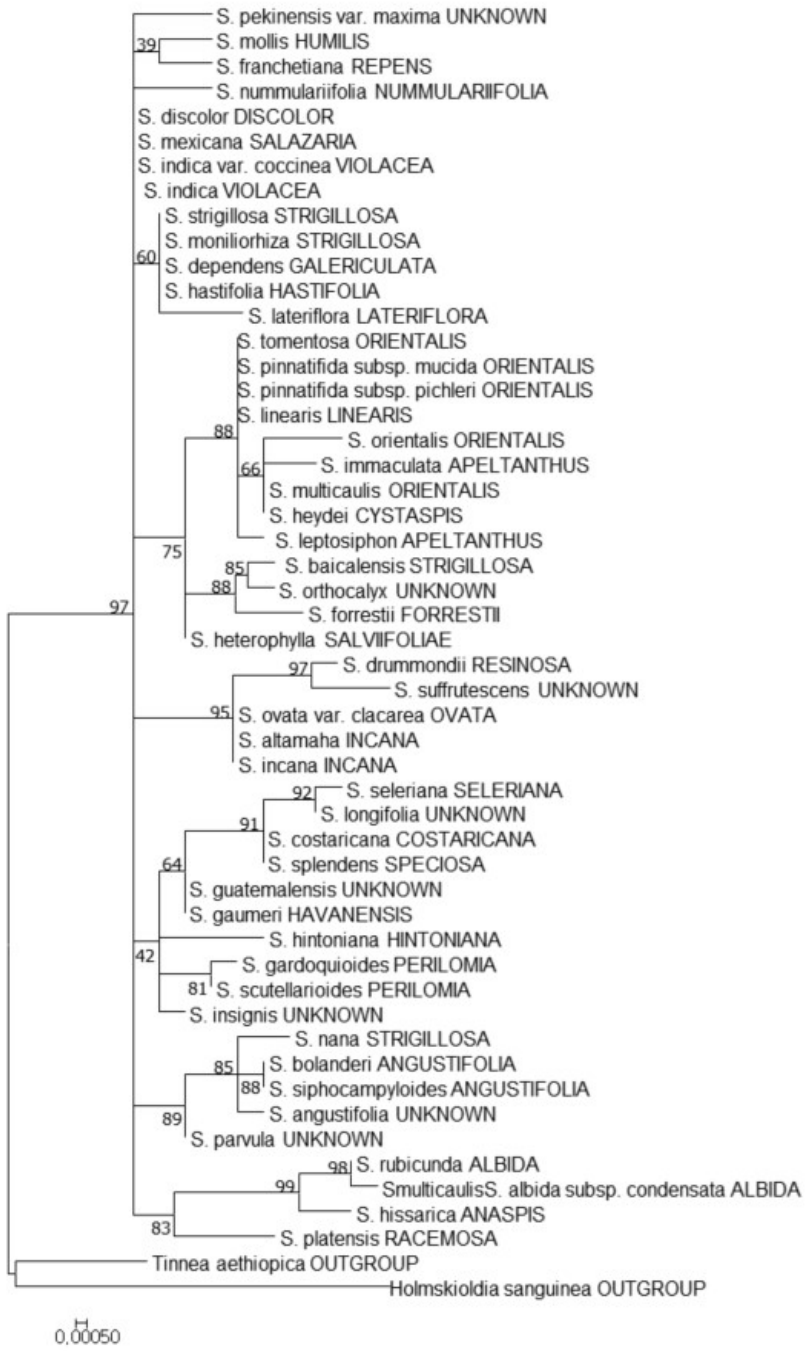
Appendix 10. Continued. Based on *trnK–rps16* among ten regions with high nucleotide diversity.



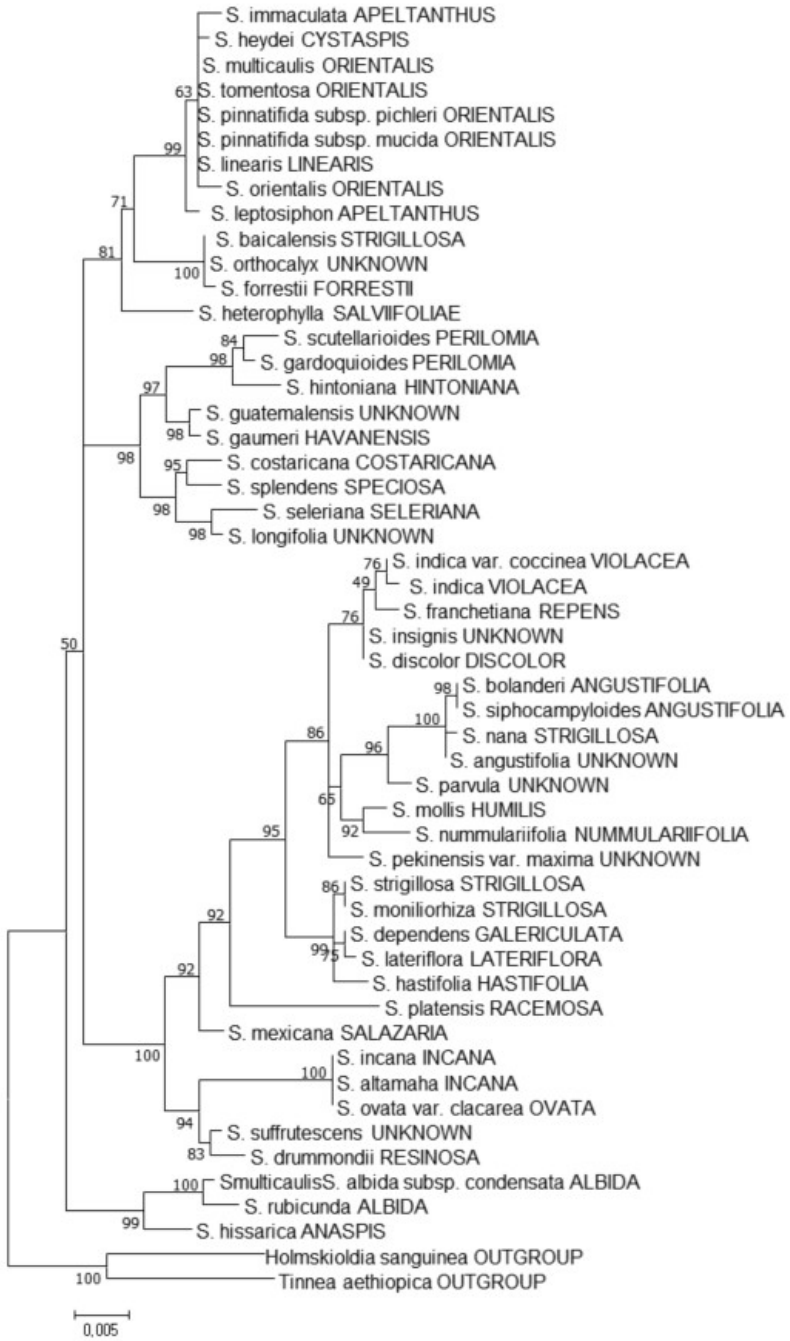
Appendix 10. Continued. Based on *ycf1* among ten regions with high nucleotide diversity.



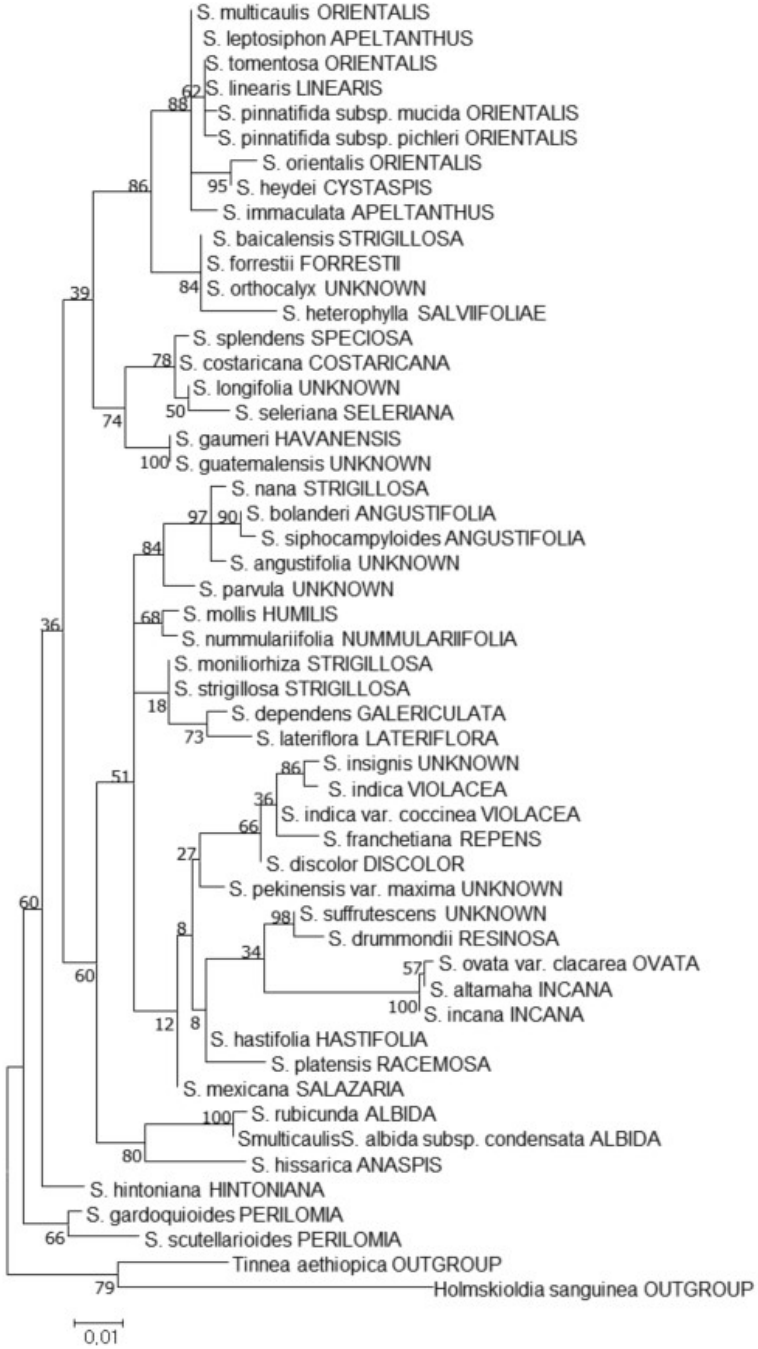
Appendix 10. Continued. Based on *ndhF* among ten regions with high nucleotide diversity.



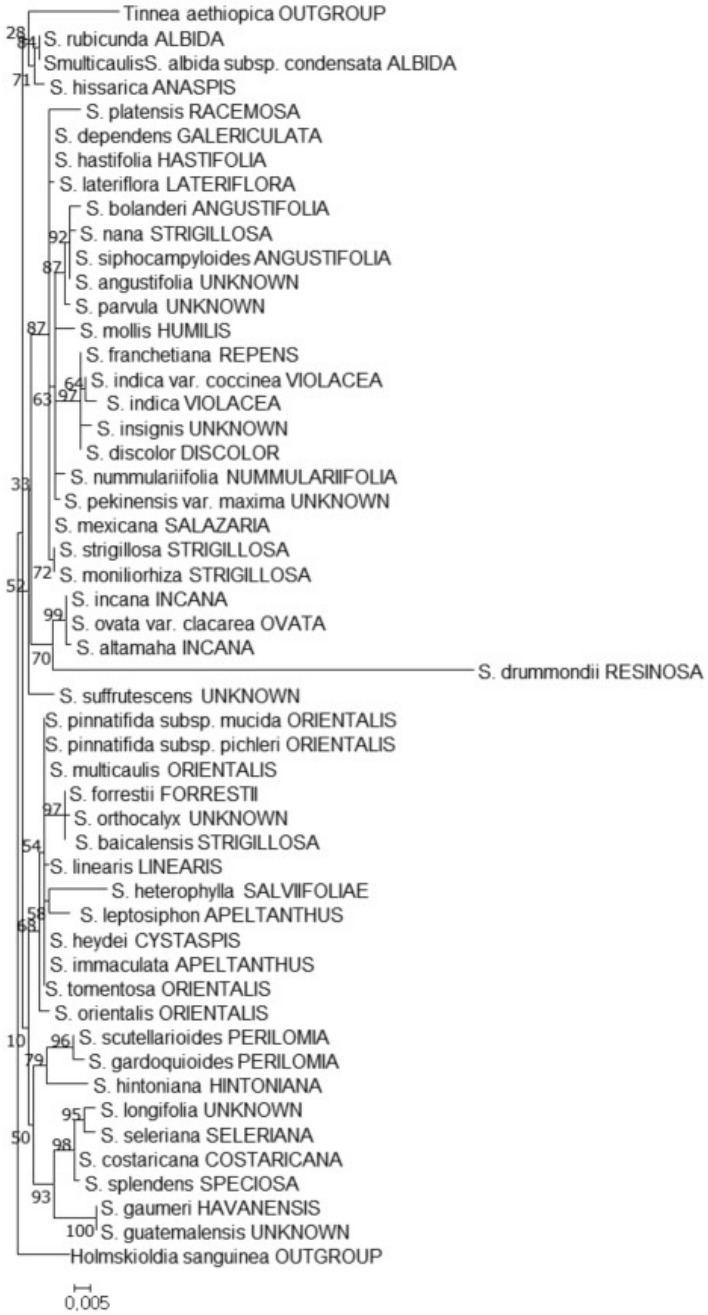
Appendix 10. Continued. Based on *petA−psbJ* among ten regions with high nucleotide diversity.



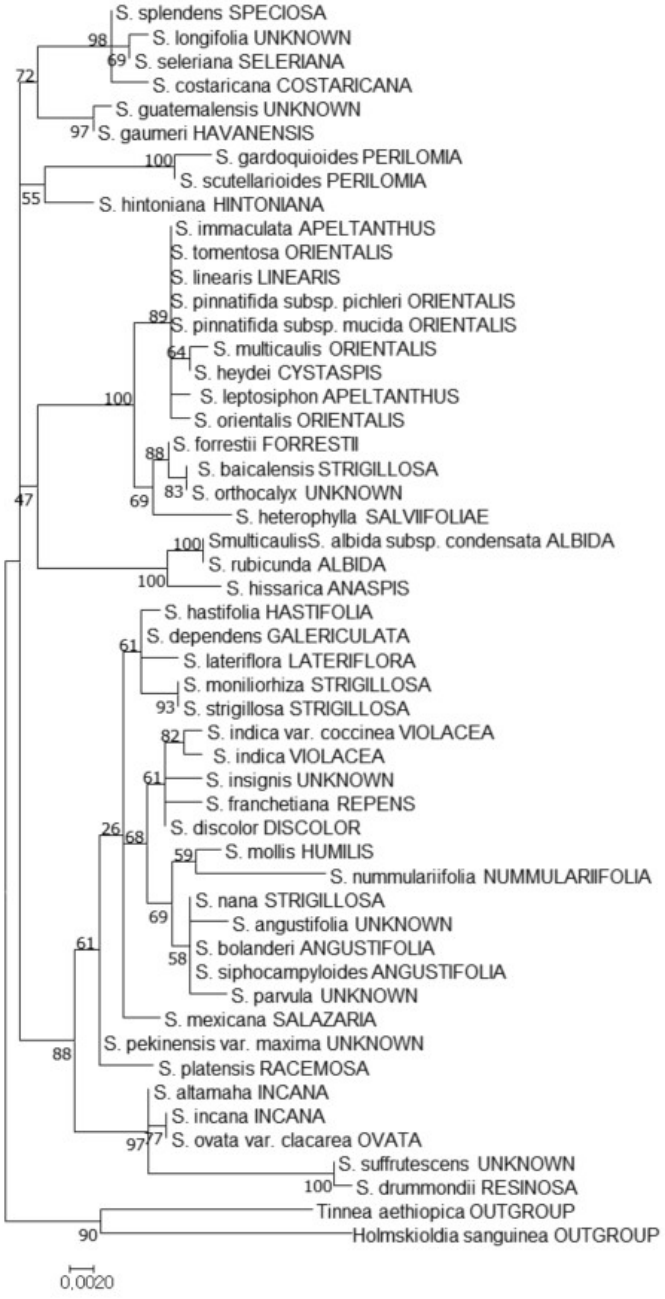
Appendix 10. Continued. Based on *petN–psbM* among ten regions with high nucleotide diversity.



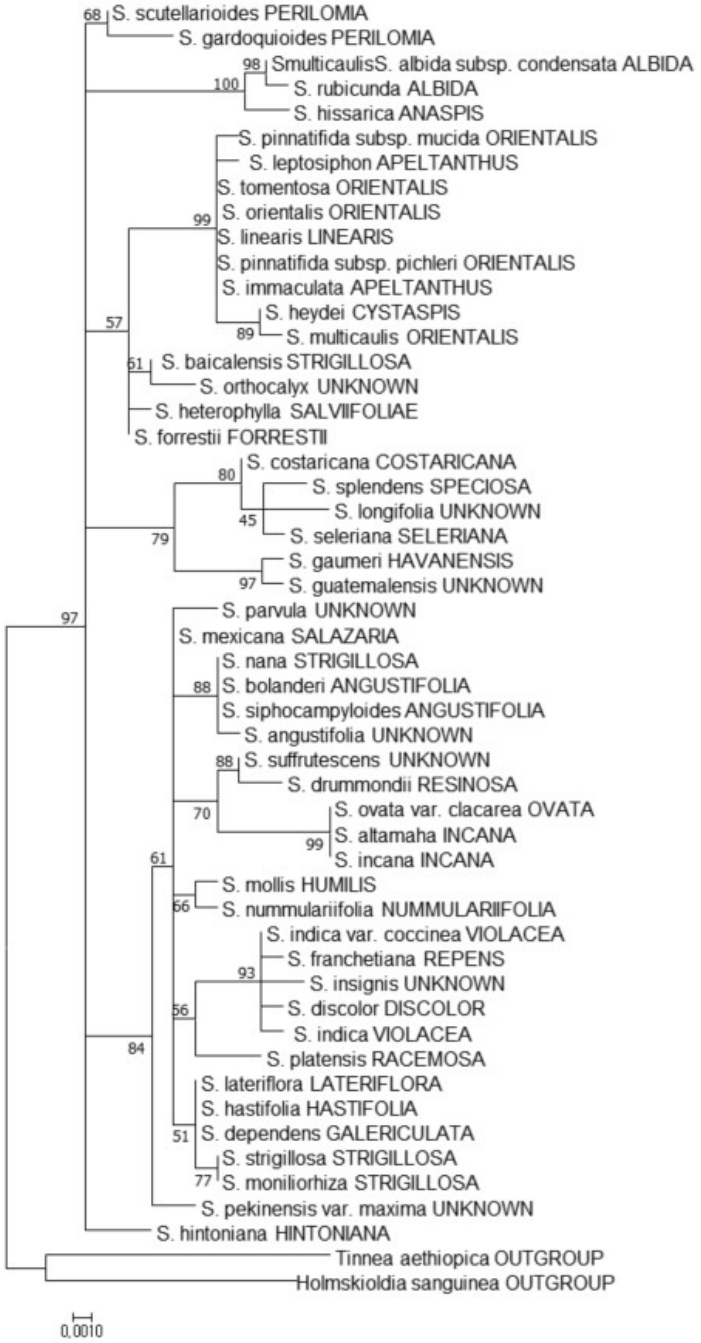
Appendix 10. Continued. Based on *ccsA*−*ndhD* among ten regions with high nucleotide diversity.



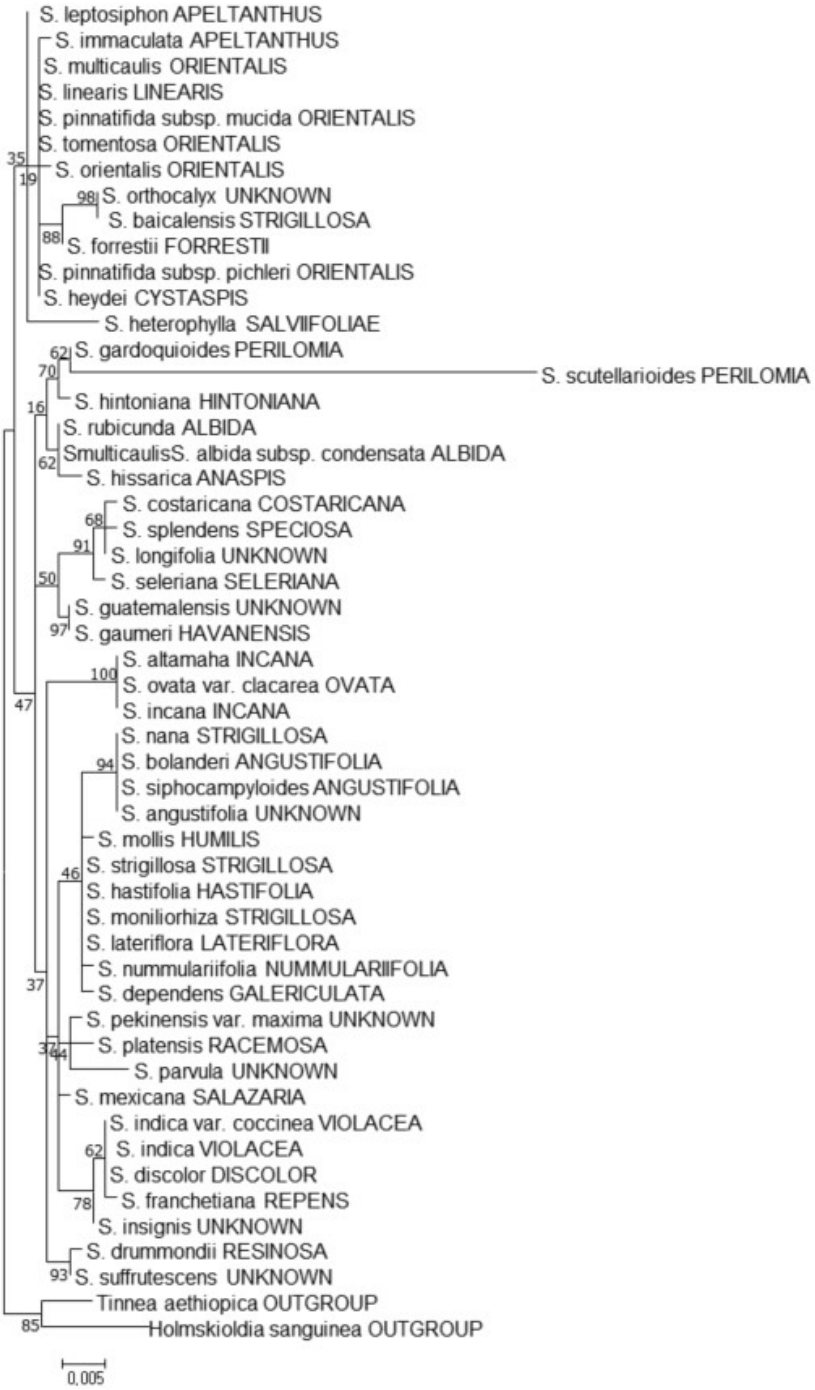
Appendix 10. Continued. Based on *trnT–trnL* among ten regions with high nucleotide diversity.



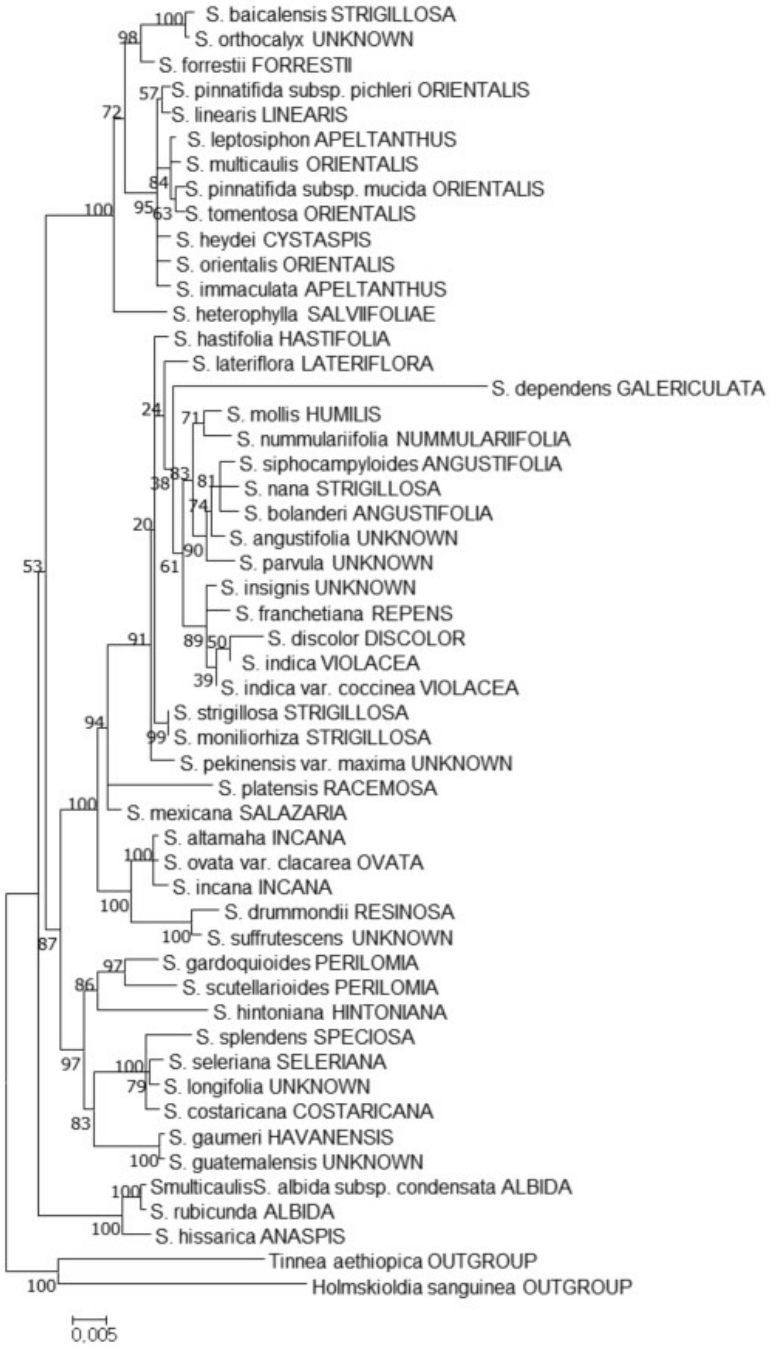
Appendix 10. Continued. Based on *rpoB*–*trnC* among ten regions with high nucleotide diversity.



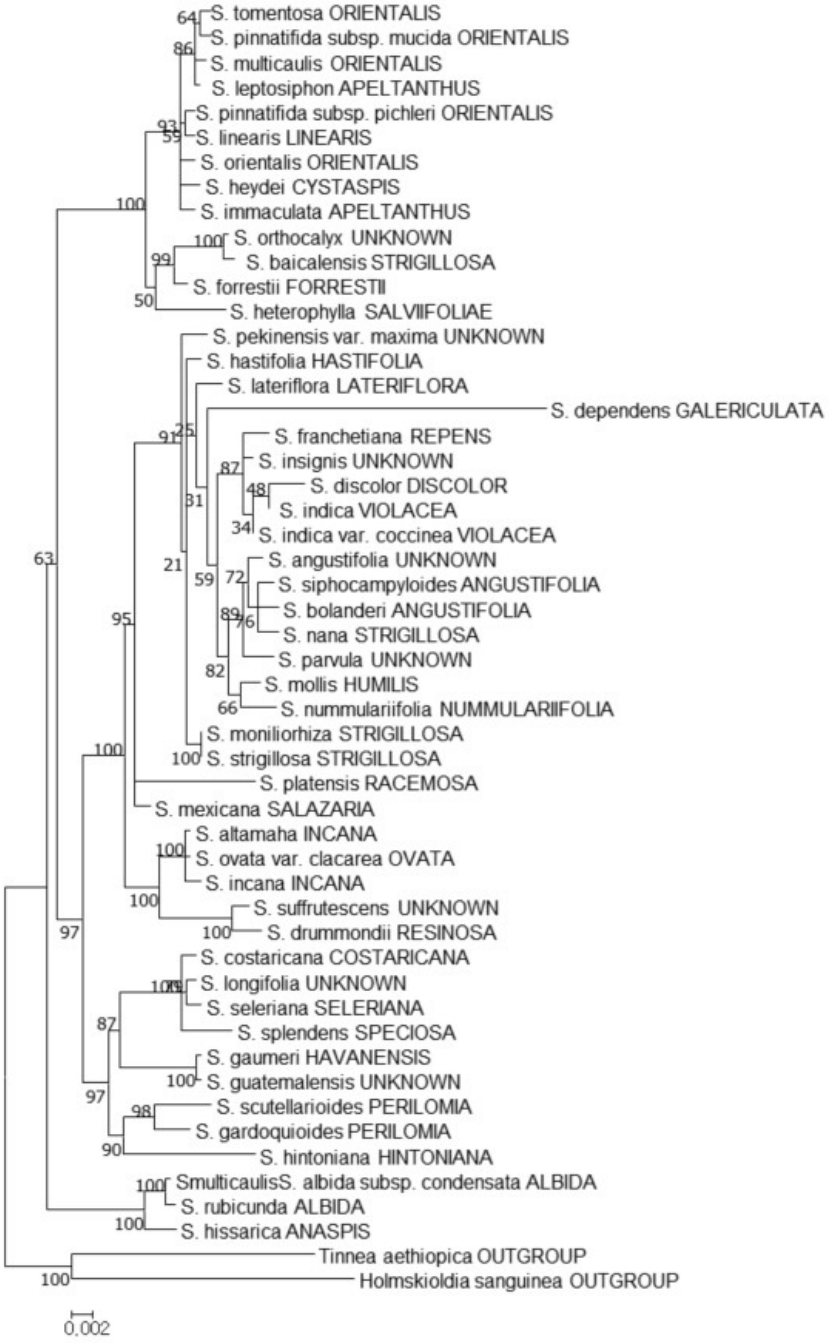
Appendix 10. Continued. Based on *rpl32-ccsA* among ten regions with high nucleotide diversity.



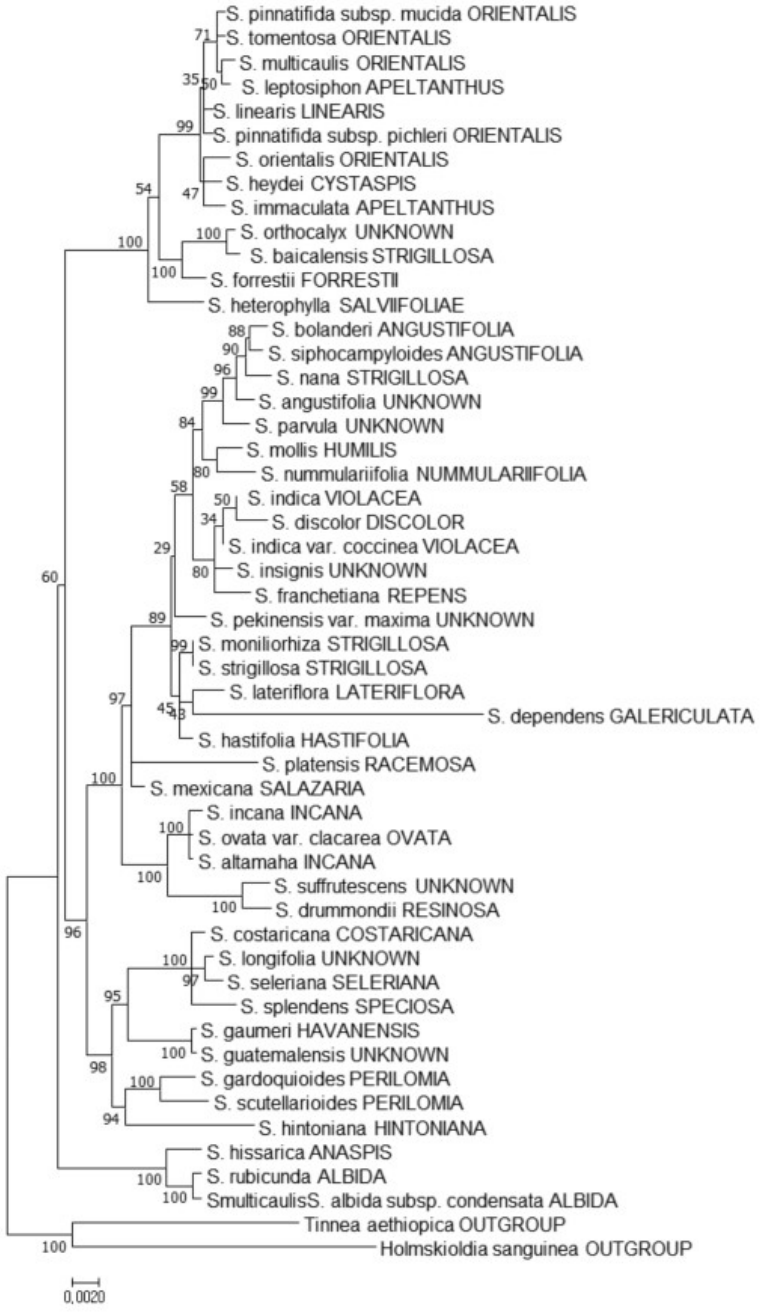
Appendix 10. Continued. Based on *rbcL*−*accD* among ten regions with high nucleotide diversity.



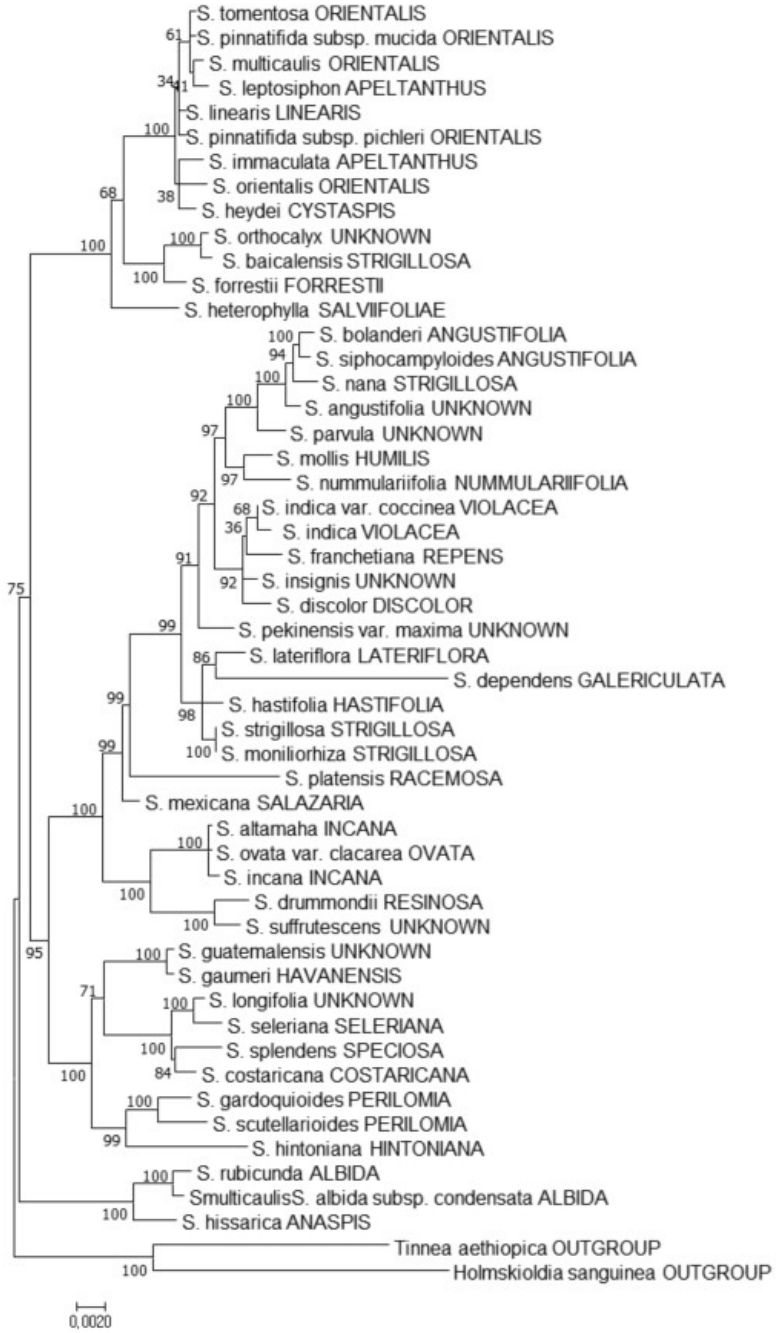
Appendix 10. Continued. Based on the top two (*trnK-rps16* and *ycf1*) regions among ten regions with high nucleotide diversity.



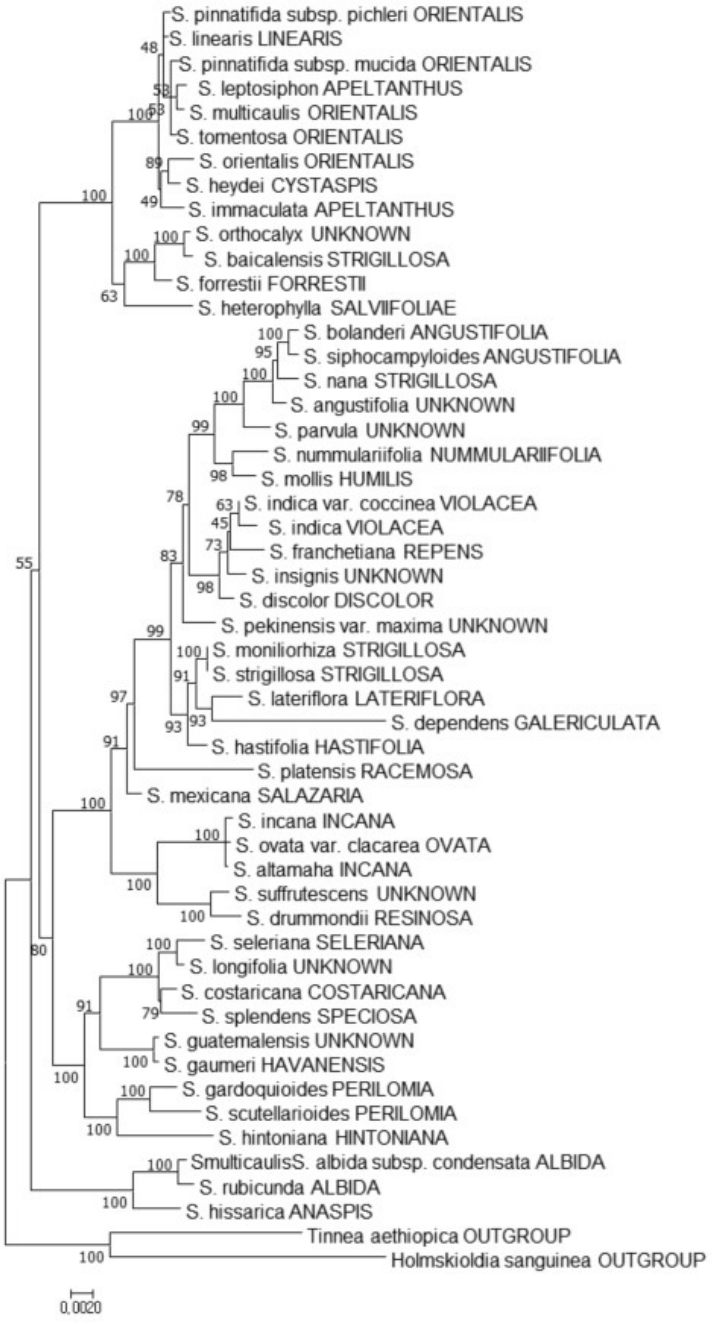
Appendix 10. Continued. Based on the top three regions (*trnK-rps16, ycf1* and *ndhF*) among ten regions with high nucleotide diversity.



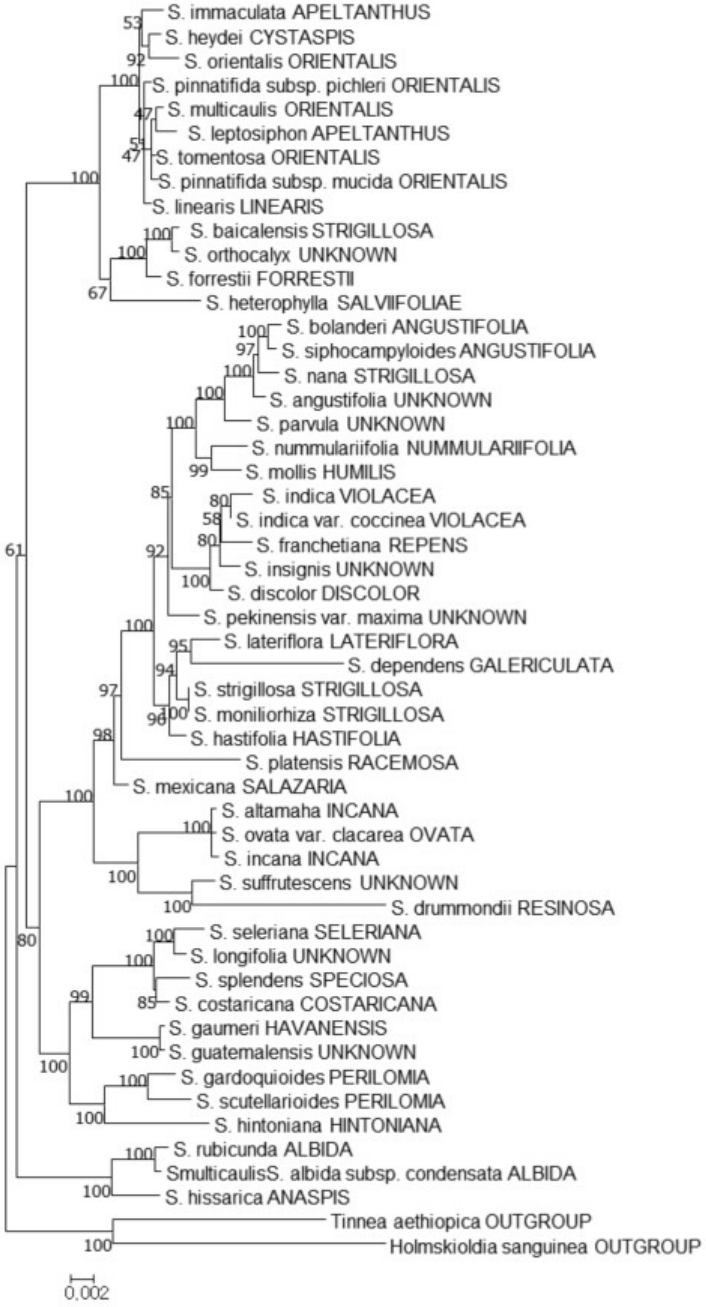
Appendix 10. Continued. Based on the top four regions (*trnK–rps16, ycf1, ndhF,* and *petA–psbJ*) among ten regions with high nucleotide diversity.



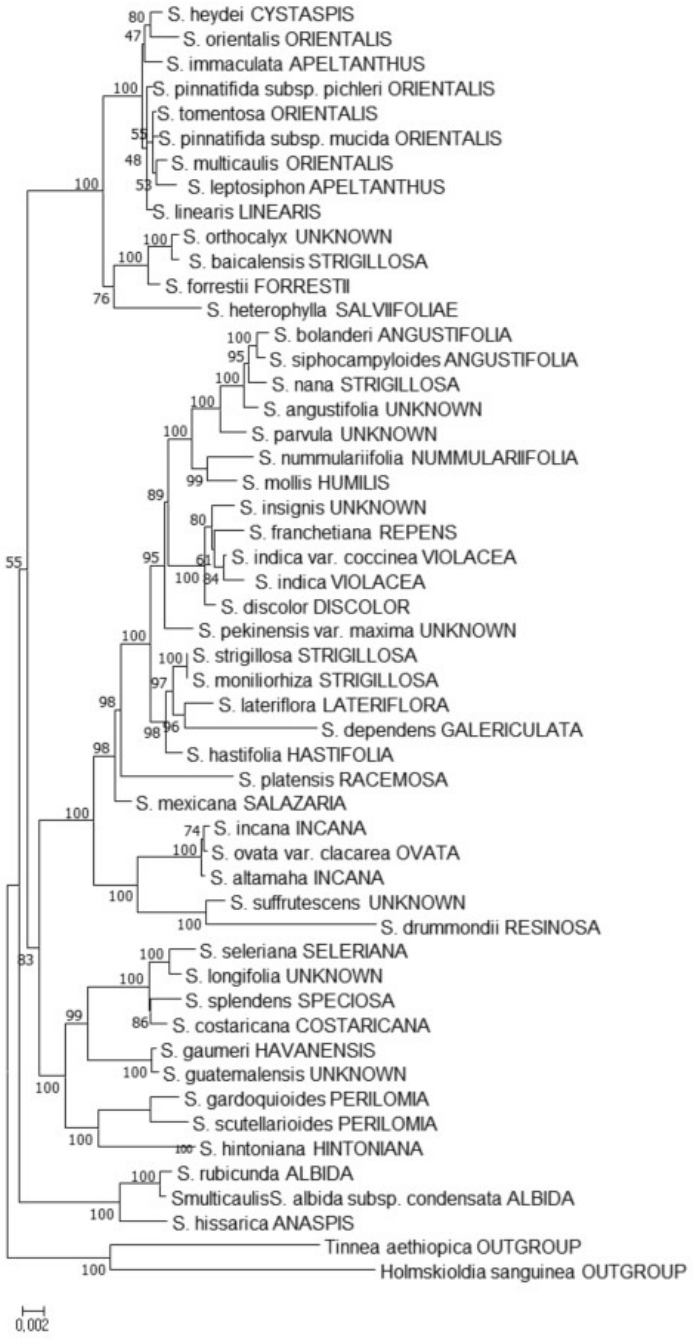
Appendix 10. Continued. Based on the top five regions (*trnK–rps16, ycf1, ndhF, petA–psbJ,* and *petN–psbM*) among ten regions with high nucleotide diversity.



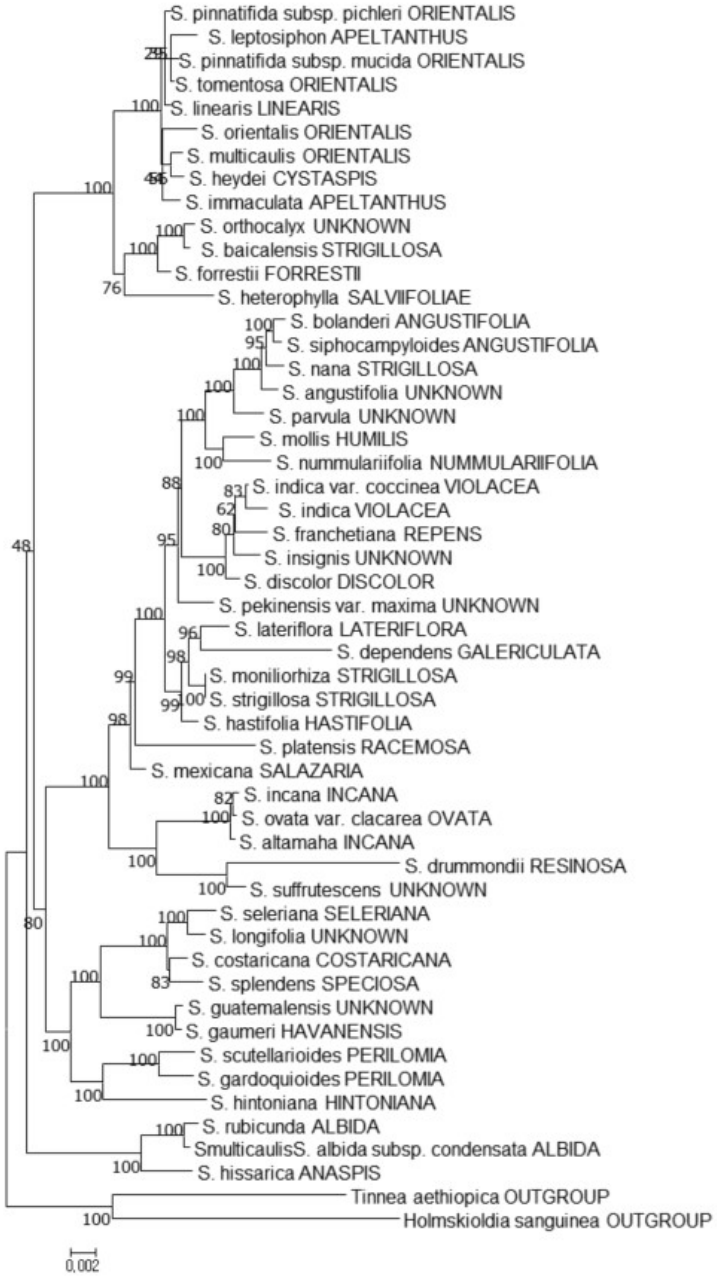
Appendix 10. Continued. Based on the top six regions (*trnK–rps16, ycf1, ndhF, petA–psbJ, petN–psbM,* and *ccsA–ndhD*) among ten regions with high nucleotide diversity.



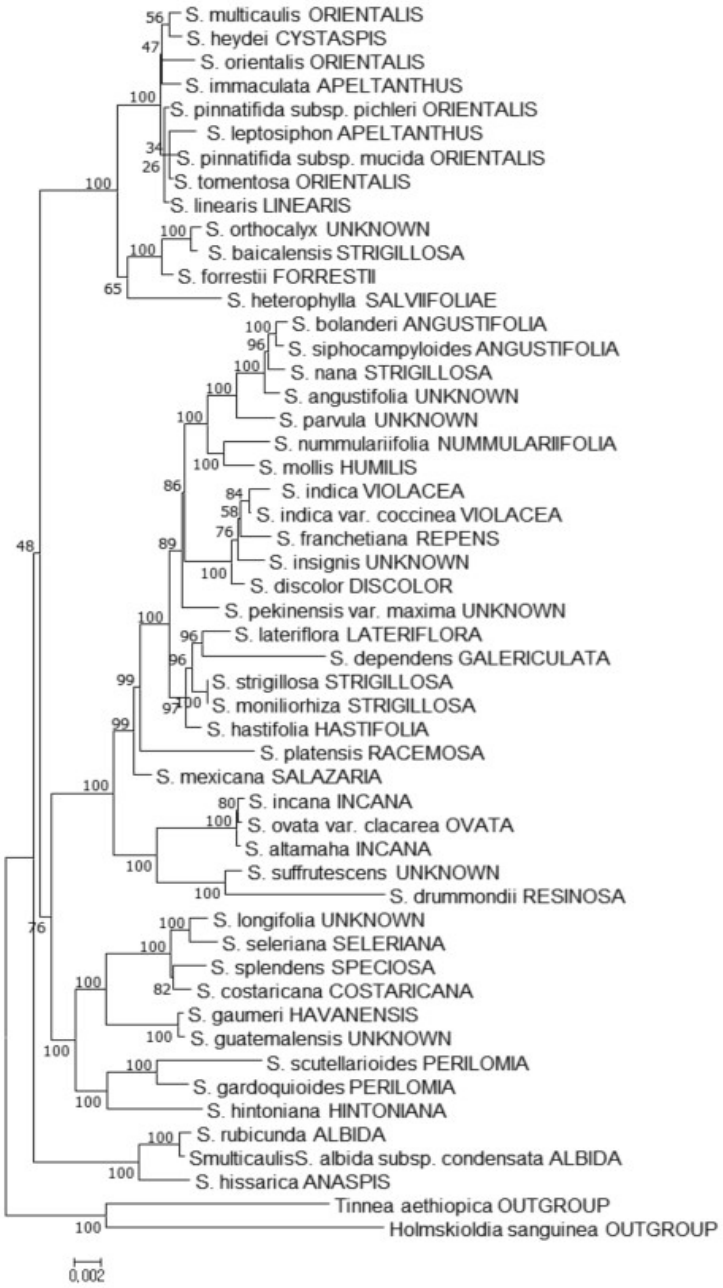
Appendix 10. Continued. Based on the top seven regions (*trnK−rps16, ycf1, ndhF, petA−psbJ, petN−psbM, ccsA−ndhD,* and *trnT−trnL*) among ten regions with high nucleotide diversity.



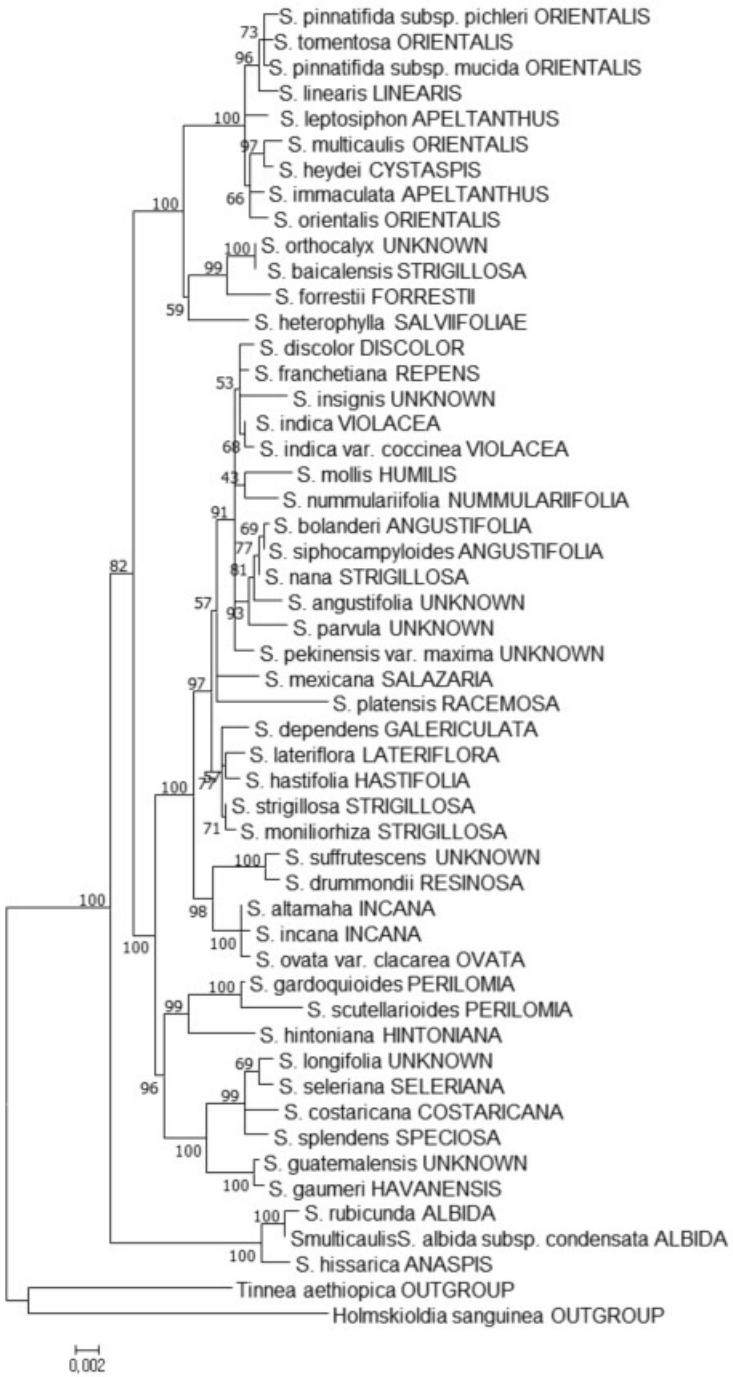
Appendix 10. Continued. Based on the top eight regions (*trnK-rps16, ycf1, ndhF, petA-psbJ, petN-psbM, ccsA-ndhD, trnT-trnL,* and *rpoB-trnC*) among ten regions with high nucleotide diversity.



Appendix 10. Continued. Based on the top nine regions (*trnK–rps16, ycf1, ndhF, petA–psbJ, petN–psbM, ccsA–ndhD, trnT–trnL, rpoB–trnC,* and *rpl32–ccsA*) among ten regions with high nucleotide diversity.



Appendix 10. Continued. Based on the top ten regions (*trnK–rps16, ycf1, ndhF, petA–psbJ, petN–psbM, ccsA–ndhD, trnT–trnL, rpoB–trnC, rpl32–ccsA,* and *rbcL–accD*) among ten regions with high nucleotide diversity.



Appendix 10. Continued. Based on the previous three genes (*ccsA*, *matK*, and *ndhF*) of Chapter 2.

APPENDICES FOR CHAPTER 4

Appendix 11. List of single-copy genes used in the probe design for Hyb-Seq.

Appendix 12. The voucher information of 96 samples used in Hyb-Seq analyses.

Appendix 13. The genomic DNA electrophoresis results for 96 samples.

Appendix 14. Statistics of sequencing and aseembly results by HybPiper.

Appendix 15. Information for supercontigs of 272 loci obtained using AMAS.

Appendix 16. A phylogenetic tree based on concatenated ML analysis using RAxML.

Appendix 17. The information on the shared quartet percentage for 272 loci calculated by Quartet (sorted in order of highest percentage).

Appendix 11. List of single-copy genes used in the probe design for Hyb-Seq.

Gene No. of *S. baicalensis*	*Arabidopsis* ID	Gene Function	*A. thaliana*	*O. sativa*	*P. trichocarpa*	*V. vinifera*	length
DN12361_c0_g3_i1	AT1G31860	histidine biosynthesis bifunctional protein (HISIE)	o		o	o	835
DN16740_c0_g1_i2	AT3G07890	Ypt/Rab-GAP domain of gyp1p superfamily protein	o	o		o	1091
DN19565_c0_g2_i4	AT5G64370	beta-ureidopropionase	o	o	o	o	1233
DN19628_c1_g1_i9	AT1G31410	putrescine-binding periplasmic protein-related	o			o	1584
DN19063_c2_g1_i9	AT5G11450	Mog1/PsbP/DUF1795-like photosystem II reaction center PsbP family protein			o	o	787
DN16484_c0_g4_i1	AT3G52200	Dihydrolipoamide acetyltransferase, long form protein		o	o	o	1527
DN16935_c0_g1_i3	AT2G02710	PAS/LOV protein B	o			o	943
DN12489_c0_g2_i1	AT5G27920	F-box family protein	o	o		o	1101
DN14441_c0_g2_i1	AT1G29260	peroxin 7	o	o	o		885
DN19154_c1_g1_i1	AT5G39590	TLD-domain containing nucleolar protein	o			o	1617
DN17379_c3_g3_i6	AT4G13360	ATP-dependent caseinolytic (Clp) protease/crotonase family protein			o	o	1245
DN18614_c0_g1_i9	AT3G54170	FKBP12 interacting protein 37	o		o		1027
DN18699_c1_g2_i7	AT1G60230	Radical SAM superfamily protein	o	o	o	o	885
DN18895_c2_g2_i8	AT5G15880	UNKNOWN	o			o	1059
DN15455_c0_g1_i1	AT3G24030	hydroxyethylthiazole kinase family protein	o		o		867
DN15349_c0_g1_i5	AT5G39830	Trypsin family protein with PDZ domain	o	o	o	o	1420
DN12096_c0_g1_i1	AT5G11540	D-arabinono-1,4-lactone oxidase family protein	o	o	o	o	1767
DN17681_c3_g5_i1	AT4G29720	polyamine oxidase 5			o	o	1581
DN13018_c0_g1_i2	AT5G08640	flavonol synthase 1	o	o		o	987
DN19099_c0_g4_i2	AT1G35190	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein	o	o		o	1074
DN15038_c0_g2_i2	AT4G34215	Domain of unknown function (DUF303)	o	o	o		747
DN17907_c0_g5_i1	AT5G67030	zeaxanthin epoxidase (ZEP) (ABA1)	o			o	1986
DN17544_c1_g3_i1	AT4G25750	ABC-2 type transporter family protein		o	o		1331
DN17790_c1_g1_i3	AT5G21060	Glyceraldehyde-3-phosphate dehydrogenase-like family protein	o		o	o	1149
DN9709_c0_g2_i1	AT3G51930	Transducin/WD40 repeat-like superfamily protein	o	o		o	1026
DN19273_c2_g5_i1	AT4G27340	Met-10+ like family protein			o	o	1225
DN10551_c0_g1_i1	AT2G21860	violaxanthin de-epoxidase-related	o	o			1620
DN12329_c0_g1_i1	AT5G44000	Glutathione S-transferase family protein	o		o		984
DN18571_c1_g3_i3	AT1G63180	UDP-D-glucose/UDP-D-galactose 4-epimerase 3		o	o	o	960
DN19202_c1_g2_i6	AT5G08720	UNKNOWN	o		o	o	2178
DN17093_c3_g1_i4	AT2G19430	DWD (DDB1-binding WD40 protein) hypersensitive to ABA 1	o		o	o	1140
DN20961_c0_g1_i1	AT2G33560	BUB1-related (BUB1: budding uninhibited by benzymidazol 1)	o			o	1173
DN12002_c0_g2_i1	AT3G54720	Peptidase M28 family protein	o		o		1803
DN15303_c0_g1_i9	AT1G16650	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein	o			o	1614
DN19286_c0_g1_i1	AT2G43950	chloroplast outer envelope protein 37	o	o		o	984
DN14847_c0_g2_i1	AT1G01910	P-loop containing nucleoside triphosphate hydrolases superfamily protein	o	o		o	903
DN19554_c0_g1_i4	AT2G31040	ATP synthase protein I -related	o	o		o	1023
DN10173_c0_g2_i2	AT2G04740	ankyrin repeat family protein	o	o	o		1749
DN17740_c0_g2_i1	AT2G40760	Rhodanese/Cell cycle control phosphatase superfamily protein	o		o	o	1263
DN19315_c0_g1_i2	AT1G59600	ZCW7	o		o		996
DN13002_c0_g1_i1	AT1G18480	Calcineurin-like metallo-phosphoesterase superfamily protein	o		o		1188
DN19202_c1_g2_i19	AT1G64600	methyltransferasescopper ion binding	o	o		o	1560
DN14633_c1_g1_i5	AT1G04900	Protein of unknown function (DUF185)	o	o	o		1350
DN16173_c1_g1_i1	AT3G25410	Sodium Bile acid symporter family			o	o	1275
DN11281_c0_g2_i1	AT3G48425	DNAse I-like superfamily protein	o	o	o	o	969
DN18993_c1_g9_i2	AT5G14720	Protein kinase superfamily protein	o	o	o	o	1725

DN16271_c0_g1_i5	AT2G03390	uvrB/uvrC motif−containing protein	o			o	1086
DN17537_c0_g6_i1	AT1G20410	Pseudouridine synthase family protein		o	o	o	1504
DN16947_c0_g3_i2	AT5G47090	UNKNOWN	o		o		744
DN16461_c0_g1_i1	AT3G22290	Endoplasmic reticulum vesicle transporter protein	o	o		o	1003
DN18525_c0_g1_i1	AT5G03555	permease, cytosine/purines, uracil, thiamine, allantoin family protein		o	o		1563
DN6834_c0_g1_i2	AT5G10460	Haloacid dehalogenase−like hydrolase (HAD) superfamily protein	o	o	o	o	948
DN15242_c0_g1_i4	AT2G45770	signal recognition particle receptor protein, chloroplast (FTSY)	o			o	1116
DN19024_c0_g1_i5	AT3G18860	transducin family protein / WD−40 repeat family protein	o	o		o	2289
DN14018_c0_g1_i2	AT5G39900	Small GTP−binding protein	o	o	o		1995
DN16698_c0_g1_i1	AT5G45170	Haloacid dehalogenase−like hydrolase (HAD) superfamily protein	o		o	o	1170
DN13544_c0_g1_i1	AT5G44520	NagB/RpiA/CoA transferase−like superfamily protein			o	o	774
DN12801_c0_g1_i2	AT1G16020	Protein of unknown function (DUF1712)			o	o	1578
DN11308_c0_g1_i1	AT3G61620	3'−5'−exoribonuclease family protein	o	o	o	o	810
DN10161_c0_g1_i1	AT2G35720	DNAJ heat shock N−terminal domain−containing protein	o	o		o	1623
DN19243_c1_g1_i3	AT4G30310	FGGY family of carbohydrate kinase	o	o	o	o	1821
DN17093_c3_g2_i1	AT1G74470	Pyridine nucleotide−disulphide oxidoreductase family protein	o	o	o		1398
DN10838_c0_g1_i1	AT3G55260	beta−hexosaminidase 1	o	o	o	o	1659
DN17500_c0_g1_i1	AT1G73180	Eukaryotic translation initiation factor eIF2A family protein	o	o	o	o	1536
DN12765_c0_g1_i3	AT1G67940	non−intrinsic ABC protein 3	o			o	801
DN16233_c0_g1_i2	AT4G11120	translation elongation factor Ts (EF−Ts), putative	o		o		1293
DN12731_c0_g1_i1	AT2G27450	nitrilase−like protein 1	o			o	966
DN12737_c0_g1_i1	AT5G25480	DNA methyltransferase−2	o		o	o	828
DN17736_c0_g1_i4	AT5G08170	porphyromonas−type peptidyl−arginine deiminase family protein		o	o	o	1116
DN13880_c0_g1_i6	AT3G13170	Spo11/DNA topoisomerase VI, subunit A protein	o	o	o	o	1086
DN18681_c1_g1_i2	AT1G12410	CLP protease proteolytic subunit 2	o	o	o	o	858
DN14530_c1_g2_i2	AT5G04770	cationic amino acid transporter 6		o	o	o	1758
DN18419_c0_g1_i3	AT1G31190	myo−inositol monophosphatase like 1	o			o	1137
DN17565_c2_g2_i3	AT2G38680	5'−nucleotidasesmagnesium ion binding	o			o	1047
DN19232_c0_g2_i7	AT2G26510	Xanthine/uracil permease family protein	o	o		o	1596
DN16171_c0_g1_i4	AT1G76130	alpha−amylase−like 2		o	o	o	1230
DN18502_c1_g1_i1	AT2G18876	Afadin/alpha−actinin−binding protein		o	o	o	1101
DN16657_c0_g1_i1	AT5G43600	ureidoglycolate amidohydrolase	o	o	o	o	1422
DN18665_c0_g1_i1	AT3G13450	Transketolase family protein		o	o	o	1104
DN15884_c0_g2_i9	AT2G04660	anaphase−promoting complex/cyclosome 2	o	o	o	o	2652
DN19060_c0_g1_i1	AT3G07700	Protein kinase superfamily protein	o	o	o		2184
DN11531_c0_g1_i2	AT5G04070	NAD(P)−binding Rossmann−fold superfamily protein			o	o	942
DN15231_c0_g2_i1	AT3G51820	UbiA prenyltransferase family protein	o	o		o	1126
DN16307_c0_g2_i1	AT1G13770	Protein of unknown function, DUF647	o		o	o	1245
DN14928_c0_g1_i2	AT3G01920	DHBP synthase RibB−like alpha/beta domain	o		o		903
DN16369_c0_g1_i2	AT2G40090	ABC2 homolog 9	o		o	o	1764
DN11182_c0_g1_i1	AT4G34910	P−loop containing nucleoside triphosphate hydrolases superfamily protein	o	o	o		1842
DN13037_c0_g1_i1	AT5G15550	Transducin/WD40 repeat−like superfamily protein	o	o	o	o	1264
DN11837_c0_g2_i1	AT4G02750	Tetratricopeptide repeat (TPR)−like superfamily protein	o	o	o	o	2011
DN13040_c0_g1_i1	AT1G70480	Domain of unknown function (DUF220)	o		o	o	978
DN19029_c0_g7_i1	AT3G59380	farnesyltransferase A	o		o	o	1055
DN15733_c1_g2_i3	AT1G69380	Protein of unknown function (DUF155)	o			o	1131
DN15904_c0_g2_i1	AT2G45500	AAA−type ATPase family protein	o		o	o	1239
DN17335_c0_g1_i1	AT4G14210	phytoene desaturase 3	o			o	1710
DN19676_c0_g3_i3	AT5G55220	trigger factor type chaperone family protein			o	o	1484
DN12972_c0_g2_i1	AT4G32990	Transducin/WD40 repeat−like superfamily protein		o	o	o	1068
DN18237_c2_g8_i1	AT3G02660	Tyrosyl−tRNA synthetase, class Ib, bacterial/mitochondrial	o	o			1425

DN17085_c0_g1_i1	AT3G14860	NHL domain-containing protein	o			o	1443
DN12055_c0_g3_i1	AT4G38240	alpha-1,3-mannosyl-glycoprotein beta-1,2-N-acetylglucosaminyltransferase, putative	o	o		o	1181
DN17117_c2_g1_i4	AT4G15545	UNKNOWN	o		o	o	967
DN17670_c0_g2_i7	AT5G16120	alpha/beta-Hydrolases superfamily protein	o	o	o		1098
DN13681_c0_g1_i1	AT3G09180	UNKNOWN	o	o	o	o	1137
DN13314_c0_g1_i1	AT4G09730	RH39	o	o			1950
DN16872_c0_g1_i1	AT4G34460	GTP binding protein beta 1	o	o		o	1134
DN15144_c0_g1_i1	AT3G33520	actin-related protein 6	o	o	o	o	993
DN16636_c0_g1_i3	AT5G40610	NAD-dependent glycerol-3-phosphate dehydrogenase family protein	o	o	o	o	1149
DN13882_c0_g1_i3	AT1G72640	NAD(P)-binding Rossmann-fold superfamily protein	o	o	o	o	747
DN15155_c0_g1_i1	AT2G27680	NAD(P)-linked oxidoreductase superfamily protein	o	o	o	o	1290
DN15758_c0_g1_i4	AT1G80770	P-loop containing nucleoside triphosphate hydrolases superfamily protein	o		o	o	1341
DN11960_c0_g2_i1	AT2G47210	myb-like transcription factor family protein	o	o	o	o	1220
DN11744_c0_g1_i1	AT4G05210	Trimeric LpxA-like enzymes superfamily protein			o	o	924
DN12874_c0_g1_i2	AT1G14620	decoy	o	o	o	o	720
DN18263_c0_g2_i2	AT1G69070	UNKNOWN	o		o	o	2880
DN11436_c0_g1_i2	AT5G52030	TraB family protein	o			o	780
DN17245_c0_g1_i1	AT1G53280	Class I glutamine amidotransferase-like superfamily protein	o	o	o	o	1308
DN17795_c0_g1_i5	AT4G33120	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein			o	o	1092
DN13036_c0_g1_i3	AT3G52390	TatD related DNase	o			o	972
DN13908_c0_g1_i2	AT1G51730	Ubiquitin-conjugating enzyme family protein	o	o		o	756
DN18970_c2_g4_i2	AT5G19150	pfkB-like carbohydrate kinase family protein	o	o		o	1038
DN19647_c0_g7_i3	AT5G08380	alpha-galactosidase 1	o	o	o	o	1020
DN15150_c0_g1_i1	AT3G47450	P-loop containing nucleoside triphosphate hydrolases superfamily protein	o	o	o	o	1713
DN15524_c0_g1_i4	AT4G18465	RNA helicase family protein	o	o	o	o	2094
DN15244_c0_g1_i3	AT3G57680	Peptidase S41 family protein	o	o	o	o	1524
DN12176_c0_g3_i2	AT4G27250	NAD(P)-binding Rossmann-fold superfamily protein			o	o	783
DN13925_c0_g1_i2	AT3G53140	O-methyltransferase family protein	o	o	o		1056
DN18174_c0_g2_i1	AT1G30300	Metallo-hydrolase/oxidoreductase superfamily protein	o		o		968
DN16801_c0_g1_i10	AT1G04620	coenzyme F420 hydrogenase family / dehydrogenase, beta subunit family	o		o	o	1371
DN11731_c1_g1_i2	AT3G46980	phosphate transporter 43	o	o	o		1575
DN15613_c0_g1_i7	AT5G56130	Transducin/WD40 repeat-like superfamily protein	o	o	o	o	942
DN14856_c0_g1_i1	AT1G64550	general control non-repressible 3	o	o	o	o	2151
DN17625_c0_g1_i2	AT1G14810	semialdehyde dehydrogenase family protein	o	o			1203
DN18972_c3_g3_i2	AT5G03770	KDO transferase A	o		o		669
DN13513_c0_g1_i5	AT5G57140	purple acid phosphatase 28	o	o			1203
DN17797_c0_g1_i3	AT1G30320	Remorin family protein			o	o	1413
DN15482_c0_g1_i1	AT2G34640	plastid transcriptionally active 12	o			o	1593
DN17772_c0_g2_i4	AT1G50500	Membrane trafficking VPS53 family protein			o	o	2484
DN12234_c0_g2_i1	AT3G12100	Cation efflux family protein	o		o	o	999
DN17045_c0_g2_i3	AT2G34090	maternal effect embryo arrest 18	o	o		o	1260
DN15725_c0_g1_i3	AT5G06260	TLD-domain containing nucleolar protein	o	o		o	1212
DN17902_c0_g6_i1	AT5G50375	cyclopropyl isomerase		o	o	o	864
DN12199_c0_g1_i1	AT1G55980	FAD/NAD(P)-binding oxidoreductase family protein			o	o	1194
DN17850_c5_g5_i1	AT5G36170	high chlorophyll fluorescent 109	o		o	o	1398
DN18568_c3_g4_i1	AT2G27590	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein	o		o	o	969
DN17032_c0_g2_i2	AT5G14220	Flavin containing amine oxidoreductase family	o	o	o	o	1572
DN19287_c1_g2_i5	AT5G51340	Tetratricopeptide repeat (TPR)-like superfamily protein	o		o	o	2169
DN7486_c0_g1_i1	AT3G05410	Photosystem II reaction center PsbP family protein	o			o	819
DN18877_c3_g5_i1	AT1G43800	Plant stearoyl-acyl-carrier-protein desaturase family protein	o		o		1149
DN17071_c0_g8_i2	AT5G50310	Galactose oxidase/kelch repeat superfamily protein	o		o	o	2028

DN12988_c0_g1_i1	AT5G50230	Transducin/WD40 repeat-like superfamily protein	o	o	o	o	1530
DN17465_c0_g2_i3	AT3G56740	Ubiquitin-associated (UBA) protein	o	o		o	886
DN16048_c0_g1_i3	AT1G69830	alpha-amylase-like 3	o		o	o	2208
DN16728_c3_g1_i5	AT4G27640	ARM repeat superfamily protein	o			o	3147
DN12536_c0_g1_i1	AT5G42470	UNKNOWN	o	o		o	1191
DN18884_c0_g2_i2	AT4G23840	Leucine-rich repeat (LRR) family protein			o	o	1755
DN18121_c2_g2_i4	AT3G07720	Galactose oxidase/kelch repeat superfamily protein	o		o		972
DN19490_c3_g3_i9	AT2G26200	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein	o		o		1653
DN18602_c0_g3_i6	AT5G48520	UNKNOWN	o	o			1863
DN18668_c0_g1_i1	AT5G42480	Chaperone DnaJ-domain superfamily protein	o	o	o	o	2361
DN14151_c0_g1_i1	AT3G52940	Ergosterol biosynthesis ERG4/ERG24 family	o			o	903
DN14366_c0_g1_i1	AT4G00560	NAD(P)-binding Rossmann-fold superfamily protein	o		o	o	966
DN18909_c4_g1_i2	AT1G65020	UNKNOWN	o		o		966
DN15601_c0_g2_i2	AT3G12080	GTP-binding family protein	o	o	o	o	1968
DN16178_c0_g1_i2	AT2G03510	SPFH/Band 7/PHB domain-containing membrane-associated protein family	o		o	o	972
DN8581_c0_g1_i1	AT5G47500	Pectin lyase-like superfamily protein	o		o	o	942
DN18369_c1_g3_i1	AT2G47760	asparagine-linked glycosylation 3			o	o	1103
DN16725_c0_g3_i2	AT3G27925	DegP protease 1			o	o	1293
DN11363_c0_g3_i2	AT3G18630	uracil dna glycosylase	o		o		987
DN13344_c0_g1_i1	AT1G72320	pumilio 23	o		o	o	2338
DN11620_c0_g2_i1	AT1G63110	GPI transamidase subunit PIG-U	o		o	o	1218
DN18402_c1_g2_i1	AT5G17410	Spc97 / Spc98 family of spindle pole body (SBP) component	o	o	o	o	1800
DN19552_c0_g3_i9	AT3G20790	NAD(P)-binding Rossmann-fold superfamily protein	o	o	o	o	1071
DN10655_c0_g1_i1	AT5G03070	importin alpha isoform 9	o	o		o	1617
DN13874_c0_g2_i1	AT5G57930	Arabidopsis thaliana protein of unknown function (DUF794)	o			o	996
DN17392_c3_g2_i4	AT4G20130	plastid transcriptionally active 14	o	o	o	o	1266
DN16981_c1_g6_i1	AT2G42810	protein phosphatase 5.2	o	o			1452
DN12756_c0_g1_i4	AT3G45640	mitogen-activated protein kinase 3	o		o	o	1158
DN17477_c0_g4_i1	AT5G65720	nitrogen fixation S (NIFS)-like 1	o	o	o		1365
DN19165_c1_g1_i1	AT2G37130	Peroxidase superfamily protein	o	o	o	o	990
DN15203_c0_g2_i5	AT2G17975	zinc finger (Ran-binding) family protein		o		o	1029
DN10263_c0_g1_i2	AT1G52155	UNKNOWN	o			o	858
DN16447_c0_g1_i1	AT4G26180	Mitochondrial substrate carrier family protein	o		o	o	830
DN15330_c0_g1_i2	AT5G60540	pyridoxine biosynthesis 2	o	o	o	o	939
DN13118_c0_g1_i1	AT1G28680	HXXXD-type acyl-transferase family protein			o	o	1392
DN16761_c0_g1_i3	AT3G04480	endoribonucleases	o			o	2262
DN19464_c3_g1_i2	AT5G19540	UNKNOWN		o		o	1356
DN17566_c0_g1_i2	AT3G54940	Papain family cysteine protease	o		o		1185
DN16606_c0_g1_i1	AT5G01400	HEAT repeat-containing protein			o	o	4004
DN19464_c3_g1_i1	AT5G19540	UNKNOWN	o		o		1398
DN19026_c3_g5_i1	AT1G14030	Rubisco methyltransferase family protein	o	o		o	1473
DN19564_c2_g1_i2	AT3G26570	phosphate transporter 21	o	o	o	o	1776
DN14916_c0_g1_i2	AT1G48430	Dihydroxyacetone kinase		o		o	1785
DN12355_c0_g1_i1	AT5G56740	histone acetyltransferase of the GNAT family 2			o	o	1389
DN7758_c0_g1_i3	AT5G41760	Nucleotide-sugar transporter family protein	o			o	1020
DN17133_c0_g3_i1	AT5G04660	cytochrome P450, family 77, subfamily A, polypeptide 4		o	o	o	1518
DN10242_c0_g1_i1	AT5G64470	Plant protein of unknown function (DUF828)	o	o		o	1218
DN21247_c0_g1_i1	AT3G50120	Plant protein of unknown function (DUF247)		o	o	o	1527
DN18917_c2_g4_i6	AT3G51830	SAC domain-containing protein 8	o	o	o	o	1530
DN15090_c0_g1_i1	AT4G09350	Chaperone DnaJ-domain superfamily protein			o	o	804
DN14278_c0_g1_i1	AT4G30950	fatty acid desaturase 6	o			o	1326

DN14543_c0_g3_i3	AT2G36990	RNApolymerase sigma−subunit F			o	o	1482
DN18089_c1_g1_i3	AT3G03310	lecithin	o		o	o	1104
DN13629_c0_g2_i4	AT5G59250	Major facilitator superfamily protein	o		o	o	1578
DN15500_c0_g1_i1	AT2G43770	Transducin/WD40 repeat−like superfamily protein	o	o			1026
DN13573_c0_g1_i1	AT1G65030	Transducin/WD40 repeat−like superfamily protein	o		o	o	1029
DN17394_c1_g1_i1	AT1G03000	peroxin 6	o		o	o	2781
DN16611_c2_g12_i3	AT3G54380	SAC3/GANP/Nin1/mts3/eIF−3 p25 family	o		o		1236
DN11824_c0_g1_i3	AT3G07080	EamA−like transporter family	o	o	o	o	1197
DN19656_c2_g3_i3	AT3G52190	phosphate transporter traffic facilitator1	o	o		o	1158
DN12228_c0_g1_i2	AT3G59710	NAD(P)−binding Rossmann−fold superfamily protein	o		o	o	921
DN19552_c0_g9_i1	AT5G22480	ZPR1 zinc−finger domain protein		o	o	o	1096
DN18356_c0_g3_i3	AT5G49970	pyridoxin (pyrodoxamine) 5'−phosphate oxidase	o			o	1602
DN12999_c0_g1_i1	AT2G39550	Prenyltransferase family protein	o			o	1071
DN8666_c0_g1_i1	AT3G16810	pumilio 24	o			o	1650
DN16908_c0_g1_i2	AT1G11870	Seryl−tRNA synthetase	o		o		1200
DN19208_c1_g2_i8	AT1G25570	Di−glucose binding protein with Leucine−rich repeat domain	o			o	1908
DN18912_c0_g2_i1	AT4G39010	glycosyl hydrolase 9B18	o	o		o	1293
DN14300_c0_g1_i1	AT3G10700	galacturonic acid kinase	o	o		o	1338
DN11899_c0_g2_i1	AT1G55760	BTB/POZ domain−containing protein	o	o	o		1014
DN14004_c0_g1_i1	AT5G15730	Protein kinase superfamily protein	o			o	1272
DN11368_c0_g1_i4	AT1G48310	chromatin remodeling factor18	o	o		o	2124
DN16529_c0_g1_i2	AT3G02710	ARM repeat superfamily protein	o	o	o		1615
DN12637_c0_g2_i1	AT2G03270	DNA−binding protein, putative	o	o	o	o	1660
DN17374_c0_g2_i1	AT4G21470	riboflavin kinase/FMN hydrolase	o	o			1152
DN15755_c0_g2_i1	AT2G38020	vacuoleless1 (VCL1)	o			o	2022
DN17651_c0_g1_i6	AT5G27970	ARM repeat superfamily protein	o			o	3902
DN11234_c0_g1_i1	AT1G04850	ubiquitin−associated (UBA)/TS−N domain−containing protein	o	o		o	1266
DN16093_c0_g1_i2	AT5G10730	NAD(P)−binding Rossmann−fold superfamily protein	o	o		o	891
DN19682_c1_g4_i5	AT1G29120	Hydrolase−like protein family	o		o		1389
DN15731_c0_g2_i2	AT5G16210	HEAT repeat−containing protein	o	o	o	o	2938
DN15514_c0_g1_i3	AT4G16180	UNKNOWN	o	o	o	o	2239
DN16839_c0_g1_i8	AT3G02300	Regulator of chromosome condensation (RCC1) family protein	o			o	1425
DN16185_c0_g1_i1	AT3G07140	GPI transamidase component Gpi16 subunit family protein	o		o	o	1977
DN18332_c0_g1_i2	AT5G11580	Regulator of chromosome condensation (RCC1) family protein	o	o	o		1752
DN10261_c0_g1_i1	AT4G22790	MATE efflux family protein	o	o			1431
DN15112_c0_g1_i1	AT1G50940	electron transfer flavoprotein alpha	o	o	o	o	1065
DN16826_c2_g2_i3	AT5G41150	Restriction endonuclease, type II−like superfamily protein	o	o			1197
DN16839_c0_g1_i4	AT3G02300	Regulator of chromosome condensation (RCC1) family protein		o	o		1485
DN13596_c0_g1_i2	AT2G26810	Putative methyltransferase family protein	o	o			336
DN16082_c0_g1_i1	AT2G36360	Galactose oxidase/kelch repeat superfamily protein	o	o	o	o	1548
DN14258_c0_g1_i1	AT2G40700	P−loop containing nucleoside triphosphate hydrolases superfamily protein	o	o	o	o	1794
DN8412_c0_g1_i1	AT3G54970	D−aminoacid aminotransferase−like PLP−dependent enzymes superfamily protein			o	o	1122
DN19396_c1_g6_i1	AT5G50840	UNKNOWN	o		o	o	1082
DN13461_c0_g1_i1	AT5G35400	Pseudouridine synthase family protein		o	o	o	1302
DN8407_c0_g2_i1	AT5G50390	Pentatricopeptide repeat (PPR−like) superfamily protein	o		o	o	1311
DN19586_c0_g5_i3	AT1G27980	dihydrosphingosine phosphate lyase		o	o	o	1281
DN12042_c0_g1_i2	AT2G45030	Translation elongation factor EFG/EF2 protein		o		o	2334
DN18531_c2_g1_i2	AT2G15860	UNKNOWN	o	o	o	o	1530
DN11902_c0_g1_i1	AT2G44970	alpha/beta−Hydrolases superfamily protein	o		o	o	1485
DN16136_c0_g1_i4	AT3G14910	UNKNOWN	o	o	o	o	1398
DN12024_c0_g2_i1	AT1G74030	enolase 1	o	o	o	o	1511

DN16137_c0_g1_i1	AT3G16840	P-loop containing nucleoside triphosphate hydrolases superfamily protein	o			o	2538
DN16168_c3_g1_i1	AT5G30510	ribosomal protein S1	o	o		o	1242
DN19185_c0_g1_i8	AT5G04360	limit dextrinase	o		o	o	2886
DN19033_c0_g4_i1	AT3G42630	Pentatricopeptide repeat (PPR) superfamily protein	o			o	1254
DN16640_c0_g2_i1	AT3G44600	cyclophilin71	o	o	o		1860
DN13755_c0_g1_i2	AT1G03110	Transducin/WD40 repeat-like superfamily protein	o		o	o	1314
DN18207_c0_g1_i1	AT5G09580	UNKNOWN	o	o	o	o	1095
DN12152_c0_g2_i1	AT3G20810	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein	o	o	o		1170
DN15986_c0_g1_i3	AT2G23390	UNKNOWN	o	o	o	o	1395
DN17124_c0_g1_i3	AT1G08460	histone deacetylase 8	o			o	1179
DN15499_c0_g1_i1	AT4G15850	RNA helicase 1	o	o	o	o	1587
DN13155_c0_g1_i1	AT1G09280	UNKNOWN	o			o	1476
DN19423_c2_g2_i3	AT5G24760	GroES-like zinc-binding dehydrogenase family protein	o	o	o	o	1149
DN17464_c0_g2_i1	AT1G71810	Protein kinase superfamily protein	o	o	o	o	1152
DN15130_c0_g1_i1	AT3G48470	embryo defective 2423			o	o	3054
DN13084_c0_g1_i2	AT1G02020	nitroreductase family protein	o		o	o	1860
DN13476_c0_g1_i1	AT2G23890	HAD-superfamily hydrolase, subfamily IG, 5'-nucleotidase	o	o			1686
DN13170_c0_g1_i1	AT2G25840	Nucleotidylyl transferase superfamily protein	o		o	o	1248
DN16386_c1_g1_i1	AT3G56040	UDP-glucose pyrophosphorylase 3	o	o	o	o	2586
DN11412_c0_g2_i2	AT4G17910	transferases, transferring acyl groups		o	o	o	966
DN15204_c0_g1_i1	AT3G59630	diphthamide synthesis DPH2 family protein	o		o	o	1524
DN14907_c0_g1_i1	AT5G58760	damaged DNA binding 2			o	o	1650
DN19395_c0_g2_i1	AT5G27270	Tetratricopeptide repeat (TPR)-like superfamily protein	o	o		o	2334
DN6748_c0_g1_i1	AT1G74460	GDSL-like Lipase/Acylhydrolase superfamily protein	o	o	o	o	1107
DN16415_c0_g1_i1	AT1G01770	UNKNOWN	o	o	o	o	1908
DN18676_c0_g1_i3	AT3G06510	Glycosyl hydrolase superfamily protein	o	o	o	o	1910
DN17430_c4_g3_i1	AT1G80680	SUPPRESSOR OF AUXIN RESISTANCE 3	o		o		3096
DN17688_c0_g6_i1	AT4G15110	cytochrome P450, family 97, subfamily B, polypeptide 3	o	o	o	o	1734
DN9996_c0_g1_i1	AT5G56780	effector of transcription2	o			o	972
DN18243_c0_g1_i1	AT5G32450	RNA binding (RRM/RBD/RNP motifs) family protein	o	o		o	718
DN16541_c1_g3_i1	AT2G30800	helicase in vascular tissue and tapetum		o	o	o	3633
DN11606_c0_g1_i1	AT4G31850	proton gradient regulation 3	o		o	o	1479
DN15597_c0_g1_i2	AT1G11880	transferases, transferring hexosyl groups	o	o	o	o	1491
DN17727_c0_g3_i1	AT2G03450	purple acid phosphatase 9		o	o		1968
DN11937_c0_g1_i3	AT5G66380	folate transporter 1	o	o	o	o	915
DN15512_c0_g1_i6	AT2G33205	Serinc-domain containing serine and sphingolipid biosynthesis protein	o	o	o	o	1227
DN13672_c0_g1_i1	NONE				o	o	2313
DN17681_c2_g1_i1	AT1G25375	Metallo-hydrolase/oxidoreductase superfamily protein	o			o	1526
DN12792_c0_g1_i1	AT5G11240	transducin family protein / WD-40 repeat family protein	o		o	o	1818
DN16919_c0_g1_i1	AT3G02760	Class II aaRS and biotin synthetases superfamily protein	o		o		2130
DN12467_c0_g1_i1	AT3G48500	Nucleic acid-binding, OB-fold-like protein	o		o	o	2067
DN15668_c0_g1_i2	AT4G31390	Protein kinase superfamily protein	o	o		o	1995
DN14800_c0_g1_i1	AT5G10920	L-Aspartase-like family protein	o	o	o	o	1542
DN12784_c0_g1_i2	AT4G00740	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein	o			o	1782
DN13905_c0_g1_i1	AT5G25150	TBP-associated factor 5	o	o		o	1718
DN19179_c0_g2_i9	AT5G53450	OBP3-responsive gene 1	o			o	1820
DN15225_c0_g1_i1	AT5G19350	RNA-binding (RRM/RBD/RNP motifs) family protein	o	o	o	o	1269
DN17417_c0_g2_i1	AT3G11830	TCP-1/cpn60 chaperonin family protein	o	o		o	1686
DN14642_c0_g1_i1	AT2G39930	isoamylase 1	o	o		o	2409
DN15095_c0_g1_i1	AT5G27660	Trypsin family protein with PDZ domain	o		o	o	1302
DN11804_c0_g2_i1	AT1G78590	NAD(H) kinase 3	o	o	o	o	978

DN12596_c0_g2_i2	AT1G79050	recA DNA recombination family protein			o	o	1281
DN14166_c0_g1_i1	AT1G12470	zinc ion binding	o	o	o	o	2406
DN12145_c0_g1_i1	AT3G63170	Chalcone−flavanone isomerase family protein	o		o	o	864
DN14134_c0_g1_i1	AT3G57180	P−loop containing nucleoside triphosphate hydrolases superfamily protein	o		o	o	2076
DN12835_c0_g1_i1	AT3G05675	BTB/POZ domain−containing protein	o		o		1305
DN14054_c0_g1_i2	AT1G71190	senescence associated gene 18	o	o		o	870
DN15269_c0_g2_i1	AT3G51050	FG−GAP repeat−containing protein	o	o		o	2094
DN19055_c2_g2_i8	AT3G45040	phosphatidate cytidylyltransferase family protein	o		o	o	1698
DN12450_c0_g1_i1	AT5G19660	SITE−1 protease	o		o	o	3084
DN13481_c0_g1_i3	AT5G56900	CwfJ−like family protein / zinc finger (CCCH−type) family protein	o	o	o	o	1542
DN17905_c0_g1_i2	AT2G13540	ARM repeat superfamily protein	o	o	o	o	2367
DN15333_c0_g1_i1	AT3G15620	DNA photolyase family protein	o	o	o	o	1674
DN13158_c0_g1_i1	AT1G65380	Leucine−rich repeat (LRR) family protein			o	o	2223
DN17639_c2_g7_i1	AT5G65490	UNKNOWN	o		o		1935
DN20935_c0_g1_i1	AT1G13040	Pentatricopeptide repeat (PPR−like) superfamily protein	o		o	o	970
DN11753_c0_g1_i1	AT5G06680	spindle pole body component 98	o			o	2829
DN12618_c0_g1_i1	AT5G57460	UNKNOWN	o	o		o	1980
DN17333_c0_g1_i3	AT3G14075	Mono−/di−acylglycerol lipase, N−terminalLipase, class 3	o	o	o		1989
DN14976_c0_g1_i1	AT5G61970	signal recognition particle−related / SRP−related	o	o		o	1809
DN11058_c0_g1_i2	AT2G37500	arginine biosynthesis protein ArgJ family	o	o		o	1413
DN15511_c0_g3_i4	AT2G35130	Tetratricopeptide repeat (TPR)−like superfamily protein			o	o	1731
DN15853_c0_g4_i2	AT5G66810	UNKNOWN			o	o	2094
DN7963_c0_g1_i1	AT3G05040	ARM repeat superfamily protein	o	o	o	o	3591
DN15511_c0_g3_i2	AT2G35130	Tetratricopeptide repeat (TPR)−like superfamily protein	o	o			1662
DN15467_c1_g1_i2	AT1G67550	urease	o		o	o	2039
DN11877_c0_g1_i1	AT1G74850	plastid transcriptionally active 2	o	o	o	o	2496
DN18029_c4_g1_i3	AT3G17940	Galactose mutarotase−like superfamily protein	o	o	o		1032
DN11809_c0_g1_i2	AT5G24300	Glycogen/starch synthases, ADP−glucose type	o		o		1980
DN19519_c0_g1_i1	AT3G04460	peroxin−12	o	o		o	1179
DN16130_c0_g1_i2	AT5G20140	SOUL heme−binding family protein			o	o	1064
DN4882_c0_g1_i1	AT3G20420	RNAse THREE−like protein 2			o	o	1071
DN13141_c0_g1_i2	AT1G73720	transducin family protein / WD−40 repeat family protein	o	o	o	o	1563
DN13516_c0_g1_i3	AT2G21960	UNKNOWN	o	o	o	o	969
DN18748_c3_g1_i13	AT2G14050	minichromosome maintenance 9	o	o	o	o	1455
DN15831_c0_g1_i4	AT1G04130	Tetratricopeptide repeat (TPR)−like superfamily protein	o	o	o	o	1104
DN19569_c0_g1_i1	AT1G63680	acid−amino acid ligasesligasesATP bindingATP bindingligases	o	o	o		2349
DN19782_c3_g1_i1	AT1G01040	dicer−like 1	o	o	o		5853
DN16818_c0_g1_i2	AT2G23540	GDSL−like Lipase/Acylhydrolase superfamily protein	o		o	o	1131
DN18621_c0_g1_i1	AT1G67960	UNKNOWN	o		o		1389
DN10075_c0_g1_i1	AT3G47860	chloroplastic lipocalin	o	o	o	o	1014
DN9873_c0_g1_i1	AT3G20440	Alpha amylase family protein	o	o		o	2796
DN18570_c0_g2_i2	AT5G51140	Pseudouridine synthase family protein		o	o	o	1164
DN18370_c4_g1_i4	AT2G19940	oxidoreductases, acting on the aldehyde or oxo group of donors, NADorNADPasacceptorcopperionbinding	o		o	o	1182
DN9684_c0_g1_i1	AT3G15150	RING/U−box superfamily protein	o		o		789
DN16844_c0_g2_i3	AT5G57655	xylose isomerase family protein	o	o	o	o	1340
DN15348_c0_g1_i4	AT1G12800	Nucleic acid−binding, OB−fold−like protein			o	o	2349
DN13485_c0_g1_i3	AT2G27760	tRNAisopentenyltransferase 2	o	o	o	o	699
DN10718_c0_g2_i1	AT4G21770	Pseudouridine synthase family protein	o		o	o	1283
DN18339_c0_g2_i2	NONE				o	o	1449
DN19331_c2_g1_i2	AT3G26670	Protein of unknown function (DUF803)	o	o	o	o	1338

DN18120_c0_g1_i2	AT1G03090	methylcrotonyl−CoA carboxylase alpha chain, mitochondrial /3−methylcrotonyl−CoAcarboxylase1(MCCA)	o	o	o	o	2232
DN6810_c0_g1_i1	AT1G08960	cation exchanger 11	o		o	o	1374
DN17305_c0_g1_i1	AT2G17033	pentatricopeptide (PPR) repeat−containing protein	o		o	o	1386
DN16098_c0_g1_i2	AT4G38890	FMN−linked oxidoreductases superfamily protein	o			o	2163
DN17255_c0_g1_i2	AT5G14260	Rubisco methyltransferase family protein	o		o		1548
DN8715_c0_g1_i1	AT2G29900	Presenilin−2	o		o	o	1086
DN6798_c0_g1_i1	AT2G28070	ABC−2 type transporter family protein	o	o		o	2044
DN16514_c0_g2_i1	AT5G02230	Haloacid dehalogenase−like hydrolase (HAD) superfamily protein	o	o	o		797
DN11857_c0_g1_i1	AT4G31790	Tetrapyrrole (Corrin/Porphyrin) Methylases	o	o	o	o	798
DN11552_c0_g1_i1	AT3G66658	aldehyde dehydrogenase 22A1	o		o	o	1776
DN15890_c0_g1_i1	AT1G73110	P−loop containing nucleoside triphosphate hydrolases superfamily protein	o	o	o	o	1293
DN11430_c0_g1_i1	AT1G16280	RNA helicase 36	o		o	o	1485
DN3835_c0_g1_i1	AT1G64350	Transducin/WD40 repeat−like superfamily protein	o	o	o	o	990
DN19205_c2_g3_i1	AT1G53770	O−fucosyltransferase family protein		o	o		1731
DN19448_c2_g2_i2	AT3G57790	Pectin lyase−like superfamily protein	o	o			1407
DN14540_c0_g1_i3	AT3G24430	ATP binding	o	o	o	o	1602
DN15885_c0_g1_i3	AT1G27150	Tetratricopeptide repeat (TPR)−like superfamily protein	o	o	o	o	1449
DN11257_c0_g1_i1	AT5G14600	S−adenosyl−L−methionine−dependent methyltransferases superfamily protein	o		o		927
DN18482_c0_g1_i1	AT5G18070	phosphoglucosamine mutase−related	o	o	o		1674
DN17712_c1_g5_i1	AT3G54460	SNF2 domain−containing protein / helicase domain−containing protein /F−boxfamilyprotein			o	o	4113
DN12424_c0_g1_i1	AT2G33640	DHHC−type zinc finger family protein		o		o	1722
DN19405_c1_g1_i1	AT5G64440	fatty acid amide hydrolase	o	o	o	o	1923
DN16991_c1_g1_i1	AT3G59040	Tetratricopeptide repeat (TPR)−like superfamily protein	o	o	o	o	1740
DN15760_c0_g1_i1	AT1G67420	Zn−dependent exopeptidases superfamily protein	o	o		o	2625
DN16176_c0_g3_i4	AT4G10050	esterase/lipase/thioesterase family protein	o			o	1044
DN13915_c0_g1_i1	AT5G54860	Major facilitator superfamily protein	o	o		o	1497
DN19090_c1_g1_i2	AT2G42700	UNKNOWN	o		o	o	2574
DN16316_c0_g1_i2	AT3G44190	FAD/NAD(P)−binding oxidoreductase family protein	o	o	o		1107
DN18034_c0_g1_i4	AT2G39910	ARM repeat superfamily protein	o		o	o	1362
DN16815_c0_g1_i2	AT2G31060	elongation factor family protein	o	o	o	o	2055
DN18474_c0_g1_i4	AT3G44880	Pheophorbide a oxygenase family protein with Rieske [2Fe−2S] domain	o			o	1632
DN12077_c0_g1_i1	AT1G03687	DTW domain−containing protein	o	o	o	o	966
DN16986_c1_g6_i2	AT3G04870	zeta−carotene desaturase	o			o	1725
DN11871_c0_g1_i1	AT5G01470	S−adenosyl−L−methionine−dependent methyltransferases superfamily protein	o	o	o	o	744
DN19588_c3_g5_i2	AT2G13840	Polymerase/histidinol phosphatase−like	o		o		1274
DN19515_c0_g1_i3	AT3G52030	F−box family protein with WD40/YVTN repeat doamin			o	o	1306
DN14901_c0_g1_i2	AT1G52630	O−fucosyltransferase family protein	o	o		o	1323
DN17731_c0_g1_i8	AT5G48330	Regulator of chromosome condensation (RCC1) family protein	o		o		1335
DN10505_c0_g2_i1	AT3G57190	peptide chain release factor, putative	o		o	o	1074
DN7552_c0_g2_i1	AT3G09720	P−loop containing nucleoside triphosphate hydrolases superfamily protein	o	o	o	o	1617
DN16623_c0_g1_i1	AT5G63890	histidinol dehydrogenase	o	o			1350
DN20512_c0_g1_i1	AT1G50710	UNKNOWN	o	o		o	1353
DN17850_c5_g2_i1	AT3G47610	transcription regulatorszinc ion binding	o		o	o	1146
DN14715_c0_g1_i3	AT2G47830	Cation efflux family protein	o		o	o	1509
DN10117_c0_g1_i1	AT1G63810	UNKNOWN	o		o	o	3045
DN18312_c1_g6_i1	AT3G24190	Protein kinase superfamily protein	o	o			1878
DN18903_c0_g2_i1	AT4G24270	EMBRYO DEFECTIVE 140	o			o	2520
DN15331_c0_g1_i1	AT1G12050	fumarylacetoacetase, putative	o		o	o	1263
DN13213_c0_g1_i1	AT1G51965	ABA Overly−Sensitive 5	o	o		o	1944

DN17092_c0_g1_i4	AT3G02060	DEAD/DEAH box helicase, putative	o	o			1635
DN13833_c0_g1_i3	AT1G78010	tRNA modification GTPase, putative			o	o	1324
DN14140_c0_g1_i1	AT4G29830	Transducin/WD40 repeat-like superfamily protein	o	o	o	o	996
DN18825_c0_g8_i1	AT5G61070	histone deacetylase of the RPD3/HDA1 superfamily 18			o	o	1995
DN18034_c1_g1_i4	AT2G01460	P-loop containing nucleoside triphosphate hydrolases superfamily protein	o	o		o	1998
DN14069_c0_g1_i1	AT1G20810	FKBP-like peptidyl-prolyl cis-trans isomerase family protein			o	o	684
DN19224_c1_g1_i2	AT5G05310	TLC ATP/ADP transporter	o		o		1038
DN13497_c0_g1_i3	AT4G20930	6-phosphogluconate dehydrogenase family protein		o	o	o	1050
DN19125_c0_g1_i1	AT3G05350	Metallopeptidase M24 family protein	o	o	o		2145
DN10752_c0_g1_i1	AT4G14605	Mitochondrial transcription termination factor family protein			o	o	1455
DN19305_c1_g2_i6	AT2G25530	AFG1-like ATPase family protein		o	o	o	1508
DN16010_c0_g3_i2	AT1G56180	UNKNOWN	o			o	1152
DN15516_c0_g1_i1	AT5G40440	mitogen-activated protein kinase kinase 3	o	o	o	o	1578
DN19252_c0_g3_i2	AT4G12680	UNKNOWN	o	o	o	o	816
DN11395_c0_g11_i1	AT4G23440	Disease resistance protein (TIR-NBS class)	o	o	o	o	2460
DN17355_c0_g1_i1	AT5G57960	GTP-binding protein, HflX			o	o	1656
DN18692_c1_g1_i4	AT4G17100	UNKNOWN			o	o	1248
DN14455_c0_g1_i1	AT3G55480	protein affected traf&#64257cking 2			o	o	2979
DN16041_c0_g1_i2	AT5G18900	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein	o	o	o	o	879
DN14204_c0_g1_i1	AT1G16900	Alg9-like mannosyltransferase family	o	o	o	o	1767
DN18050_c3_g5_i1	AT1G08520	ALBINA 1	o	o			2289
DN16384_c0_g1_i4	AT3G19210	homolog of RAD54	o	o	o	o	2790
DN18538_c2_g1_i2	AT5G09860	nuclear matrix protein-related	o	o			1887
DN16453_c0_g1_i1	AT5G27560	Domain of unknown function (DUF1995)	o			o	1002
DN17692_c0_g2_i1	AT3G01720	UNKNOWN	o		o	o	2046
DN19360_c3_g4_i6	AT4G33630	Protein of unknown function (DUF3506)	o		o	o	2097
DN17821_c0_g1_i1	AT3G07670	Rubisco methyltransferase family protein	o	o	o	o	1578
DN19278_c0_g1_i2	AT4G34900	xanthine dehydrogenase 2		o		o	3171
DN11829_c0_g1_i2	AT3G21300	RNA methyltransferase family protein	o	o	o	o	1614
DN13265_c0_g3_i3	AT1G11780	oxidoreductase, 2OG-Fe(II) oxygenase family protein	o		o	o	1080
DN16910_c0_g1_i3	AT3G14890	phosphoesterase	o	o	o	o	1107
DN17586_c0_g1_i6	AT1G02145	homolog of asparagine-linked glycosylation 12	o	o	o	o	1140
DN14588_c0_g1_i2	AT5G65950	UNKNOWN	o			o	3546
DN16443_c0_g1_i7	AT1G15240	Phox-associated domainPhox-likeSorting nexin, C-terminal	o	o			3201
DN20148_c0_g1_i1	AT5G45790	Ubiquitin carboxyl-terminal hydrolase family protein	o		o		1322
DN13973_c0_g4_i1	AT2G40390	UNKNOWN	o	o	o	o	1479
DN8683_c0_g1_i2	AT3G57480	zinc finger (C2H2 type, AN1-like) family protein		o	o	o	746
DN17071_c0_g1_i15	AT3G19720	P-loop containing nucleoside triphosphate hydrolases superfamily protein	o	o			2253
DN13397_c0_g1_i1	AT4G27450	Aluminium induced protein with YGL and LRDR motifs	o			o	783
DN17071_c0_g1_i16	AT3G19720	P-loop containing nucleoside triphosphate hydrolases superfamily protein			o	o	2367
DN18728_c1_g2_i1	AT3G61080	Protein kinase superfamily protein	o	o	o	o	888
DN10654_c0_g1_i1	AT4G30840	Transducin/WD40 repeat-like superfamily protein	o		o		891
DN16221_c0_g1_i3	AT2G16860	GCIP-interacting family protein	o	o	o	o	960
DN14959_c0_g1_i1	AT1G31780	UNKNOWN	o	o	o	o	2088
DN12235_c0_g1_i1	AT3G58530	RNI-like superfamily protein	o	o	o	o	1083
DN17132_c1_g2_i2	AT5G12410	THUMP domain-containing protein	o	o	o	o	1140
DN14783_c0_g1_i1	AT2G14530	TRICHOME BIREFRINGENCE-LIKE 13		o	o		1245
DN14294_c0_g1_i1	AT5G13630	magnesium-chelatase subunit chlH, chloroplast, putative /Mg-protoporphyrinIXchelatase,putative(CHLH)	o		o	o	4197
DN12774_c0_g1_i1	AT3G55760	UNKNOWN	o		o		1827
DN15692_c0_g2_i1	AT5G55960	UNKNOWN	o	o	o		2022

DN10425_c0_g3_i1	AT3G26540	Tetratricopeptide repeat (TPR)-like superfamily protein	o		o		931
DN10457_c0_g1_i1	AT5G62030	diphthamide synthesis DPH2 family protein	o		o		1356
DN10594_c0_g1_i3	AT5G48730	Pentatricopeptide repeat (PPR) superfamily protein	o	o	o	o	1425
DN10678_c0_g1_i1	AT1G07740	Tetratricopeptide repeat (TPR)-like superfamily protein	o		o		1068
DN10751_c0_g1_i1	AT5G50280	Pentatricopeptide repeat (PPR) superfamily protein	o		o	o	2051
DN10784_c0_g1_i1	AT2G17670	Tetratricopeptide repeat (TPR)-like superfamily protein	o		o		949
DN10983_c0_g1_i1	AT3G61350	SKP1 interacting partner 4			o	o	1203
DN10998_c0_g1_i1	AT2G04850	Auxin-responsive family protein	o	o			1196
DN11000_c0_g1_i1	AT5G46220	Protein of unknown function (DUF616)			o	o	1191
DN11078_c0_g1_i1	AT1G22730	MA3 domain-containing protein	o	o	o		1401
DN11241_c0_g1_i1	AT1G63850	BTB/POZ domain-containing protein	o	o		o	1452
DN11349_c0_g1_i1	AT5G04200	metacaspase 9	o			o	900
DN11352_c0_g3_i1	AT1G62260	mitochondrial editing factor 9	o		o	o	1416
DN11371_c0_g1_i1	AT1G02420	Pentatricopeptide repeat (PPR) superfamily protein	o		o		1272
DN11381_c0_g1_i2	AT3G48460	GDSL-like Lipase/Acylhydrolase superfamily protein	o		o	o	1140
DN11385_c0_g1_i1	AT1G61870	pentatricopeptide repeat 336	o			o	1197
DN11432_c0_g1_i1	AT5G13520	peptidase M1 family protein	o	o			1917
DN11448_c0_g1_i1	AT1G04420	NAD(P)-linked oxidoreductase superfamily protein	o			o	1284
DN11508_c0_g1_i1	AT5G15100	Auxin efflux carrier family protein	o			o	996
DN11545_c0_g1_i1	AT1G55880	Pyridoxal-5'-phosphate-dependent enzyme family protein	o	o	o	o	1347
DN11550_c0_g1_i1	AT5G65860	ankyrin repeat family protein	o	o	o	o	943
DN11570_c0_g2_i1	AT2G02150	Tetratricopeptide repeat (TPR)-like superfamily protein	o		o		923
DN11661_c0_g2_i1	AT5G38380	UNKNOWN	o	o	o	o	1053
DN11860_c0_g1_i1	AT3G53700	Pentatricopeptide repeat (PPR) superfamily protein	o	o	o		999
DN11932_c0_g1_i3	AT1G10510	RNI-like superfamily protein	o		o	o	1845
DN1195_c0_g1_i1	AT4G38160	Mitochondrial transcription termination factor family protein		o	o	o	831
DN11998_c1_g1_i1	AT4G02820	Pentatricopeptide repeat (PPR) superfamily protein	o	o	o	o	1410
DN12046_c0_g4_i1	AT2G41720	Tetratricopeptide repeat (TPR)-like superfamily protein	o	o	o	o	2133
DN12099_c0_g2_i2	AT4G39820	Tetratricopeptide repeat (TPR)-like superfamily protein	o		o		1137
DN12125_c0_g1_i1	AT1G08550	non-photochemical quenching 1			o	o	1440
DN12159_c0_g1_i1	AT1G32520	UNKNOWN	o		o		699
DN12247_c0_g1_i1	AT1G11090	alpha/beta-Hydrolases superfamily protein	o	o	o	o	1032
DN12263_c0_g2_i2	AT4G24750	Rhodanese/Cell cycle control phosphatase superfamily protein	o			o	873
DN12286_c0_g3_i1	AT5G03160	homolog of mamallian P58IPK	o		o	o	1527
DN12419_c0_g1_i3	AT4G12740	HhH-GPD base excision DNA repair family protein		o	o	o	1479
DN12469_c0_g1_i1	AT1G15440	periodic tryptophan protein 2	o	o		o	2700
DN12477_c0_g1_i1	AT5G03860	malate synthase	o	o		o	1683
DN12558_c0_g1_i1	AT5G03700	D-mannose binding lectin protein with Apple-like carbohydrate-bindingdomain			o	o	1274
DN12602_c0_g1_i1	AT5G21040	F-box protein 2	o		o		1641
DN12701_c0_g2_i1	AT4G08940	Ubiquitin carboxyl-terminal hydrolase family protein	o			o	1113
DN12780_c0_g1_i1	AT4G28020	UNKNOWN	o	o	o		1173
DN12782_c0_g1_i1	AT3G08970	DNAJ heat shock N-terminal domain-containing protein			o	o	1743
DN12833_c0_g1_i1	AT3G46610	Pentatricopeptide repeat (PPR-like) superfamily protein	o		o		2235
DN12946_c0_g1_i1	AT1G77090	Mog1/PsbP/DUF1795-like photosystem II reaction center PsbP family protein			o	o	822
DN13101_c0_g1_i6	AT1G45474	photosystem I light harvesting complex gene 5			o	o	840
DN13188_c0_g1_i1	AT1G43580	Sphingomyelin synthetase family protein	o			o	1149
DN13237_c0_g1_i2	AT1G18460	alpha/beta-Hydrolases superfamily protein			o	o	2088
DN13289_c0_g1_i2	AT1G64680	UNKNOWN	o		o		786
DN13290_c0_g1_i1	AT5G63310	nucleoside diphosphate kinase 2			o	o	678
DN13292_c0_g1_i2	AT5G47760	2-phosphoglycolate phosphatase 2	o	o	o		939

DN13303_c0_g1_i1	AT2G40570	initiator tRNA phosphoribosyl transferase family protein	o	o	o	o	1212
DN13319_c0_g4_i1	AT1G07010	Calcineurin-like metallo-phosphoesterase superfamily protein			o	o	1036
DN1333_c0_g1_i1	AT2G44020	Mitochondrial transcription termination factor family protein	o	o	o		1482
DN13371_c1_g1_i2	AT3G24040	Core-2/I-branching beta-1,6-N-acetylglucosaminyltransferase family protein	o		o	o	1029
DN13404_c0_g1_i3	AT3G19553	Amino acid permease family protein	o			o	1651
DN13409_c0_g2_i2	AT1G69800	Cystathionine beta-synthase (CBS) protein	o	o			1246
DN13432_c0_g1_i2	AT5G51130	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein	o			o	873
DN13433_c0_g3_i1	AT5G42310	Pentatricopeptide repeat (PPR-like) superfamily protein	o			o	1684
DN13446_c0_g1_i1	AT2G21710	Mitochondrial transcription termination factor family protein	o	o	o		2022
DN13462_c0_g3_i1	AT5G56450	Mitochondrial substrate carrier family protein	o			o	1164
DN13487_c0_g1_i1	AT3G54690	Sugar isomerase (SIS) family protein	o	o		o	1038
DN13571_c0_g2_i2	AT3G18940	clast3-related	o	o	o	o	831
DN13604_c0_g1_i1	AT1G04010	phospholipid sterol acyl transferase 1	o	o	o	o	1950
DN13620_c0_g1_i2	AT2G35450	catalyticshydrolases	o		o	o	984
DN13631_c0_g1_i1	AT1G43860	sequence-specific DNA binding transcription factors	o	o		o	939
DN13639_c0_g2_i1	AT5G15390	tRNA/rRNA methyltransferase (SpoU) family protein	o			o	1101
DN13659_c0_g2_i1	AT5G17570	TatD related DNase	o	o	o	o	891
DN13712_c0_g1_i3	AT5G16890	Exostosin family protein	o	o		o	1581
DN13826_c0_g1_i5	AT3G57050	cystathionine beta-lyase	o		o	o	1422
DN13845_c0_g1_i1	AT2G25620	DNA-binding protein phosphatase 1	o			o	1182
DN13916_c0_g2_i1	AT3G45770	Polyketide synthase, enoylreductase family protein	o		o	o	1125
DN13994_c0_g1_i2	AT1G53120	RNA-binding S4 domain-containing protein	o		o	o	1020
DN14000_c0_g1_i1	AT2G39000	Acyl-CoA N-acyltransferases (NAT) superfamily protein		o		o	744
DN14001_c0_g1_i1	AT1G74240	Mitochondrial substrate carrier family protein	o	o	o	o	1059
DN14081_c0_g1_i1	AT5G67140	F-box/RNI-like superfamily protein	o		o	o	693
DN14129_c0_g1_i1	AT3G15180	ARM repeat superfamily protein			o	o	1572
DN14143_c0_g1_i1	AT4G28980	CDK-activating kinase 1AT	o		o	o	1401
DN14195_c0_g1_i2	AT4G01310	Ribosomal L5P family protein	o			o	798
DN14197_c0_g1_i1	AT3G10220	tubulin folding cofactor B	o	o	o	o	726
DN14329_c0_g1_i1	AT1G57770	FAD/NAD(P)-binding oxidoreductase family protein	o		o	o	1719
DN14430_c0_g1_i4	AT1G21480	Exostosin family protein	o			o	1386
DN14442_c0_g1_i1	AT1G62750	Translation elongation factor EFG/EF2 protein	o		o	o	2361
DN14446_c0_g1_i1	AT4G36470	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein	o		o	o	1092
DN14467_c0_g1_i1	AT1G70330	equilibrative nucleotide transporter 1	o	o			1413
DN14484_c0_g1_i1	AT1G48520	GLU-ADT subunit B	o	o			1647
DN14498_c0_g1_i1	AT2G33630	NAD(P)-binding Rossmann-fold superfamily protein	o	o		o	1443
DN14529_c0_g1_i2	AT3G09050	UNKNOWN	o		o		786
DN14660_c0_g2_i1	AT1G74640	alpha/beta-Hydrolases superfamily protein	o	o	o	o	1137
DN14670_c0_g1_i3	AT5G13400	Major facilitator superfamily protein	o		o	o	1914
DN14681_c0_g1_i1	AT5G42130	Mitochondrial substrate carrier family protein	o		o		1245
DN14795_c0_g1_i3	AT2G45990	UNKNOWN	o	o	o	o	807
DN1486_c0_g1_i1	AT4G25080	magnesium-protoporphyrin IX methyltransferase	o		o		1005
DN14870_c0_g1_i1	AT5G63080	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein	o			o	1407
DN14903_c0_g1_i2	AT3G18165	modifier of snc1,4	o	o	o		759
DN15117_c0_g1_i5	AT5G01510	Protein of unknown function, DUF647			o	o	876
DN15132_c0_g1_i1	AT1G03910	UNKNOWN	o	o	o	o	1806
DN15139_c0_g2_i3	AT3G48820	Glycosyltransferase family 29 (sialyltransferase) family protein	o	o	o	o	1323
DN15162_c0_g2_i2	AT5G66530	Galactose mutarotase-like superfamily protein	o		o	o	948
DN15178_c0_g1_i4	AT3G27000	actin related protein 2	o	o	o	o	1170
DN15237_c0_g2_i2	AT5G55000	potassium channel tetramerisation domain-containing protein /pentapeptiderepeat-containingprotein	o	o	o	o	903

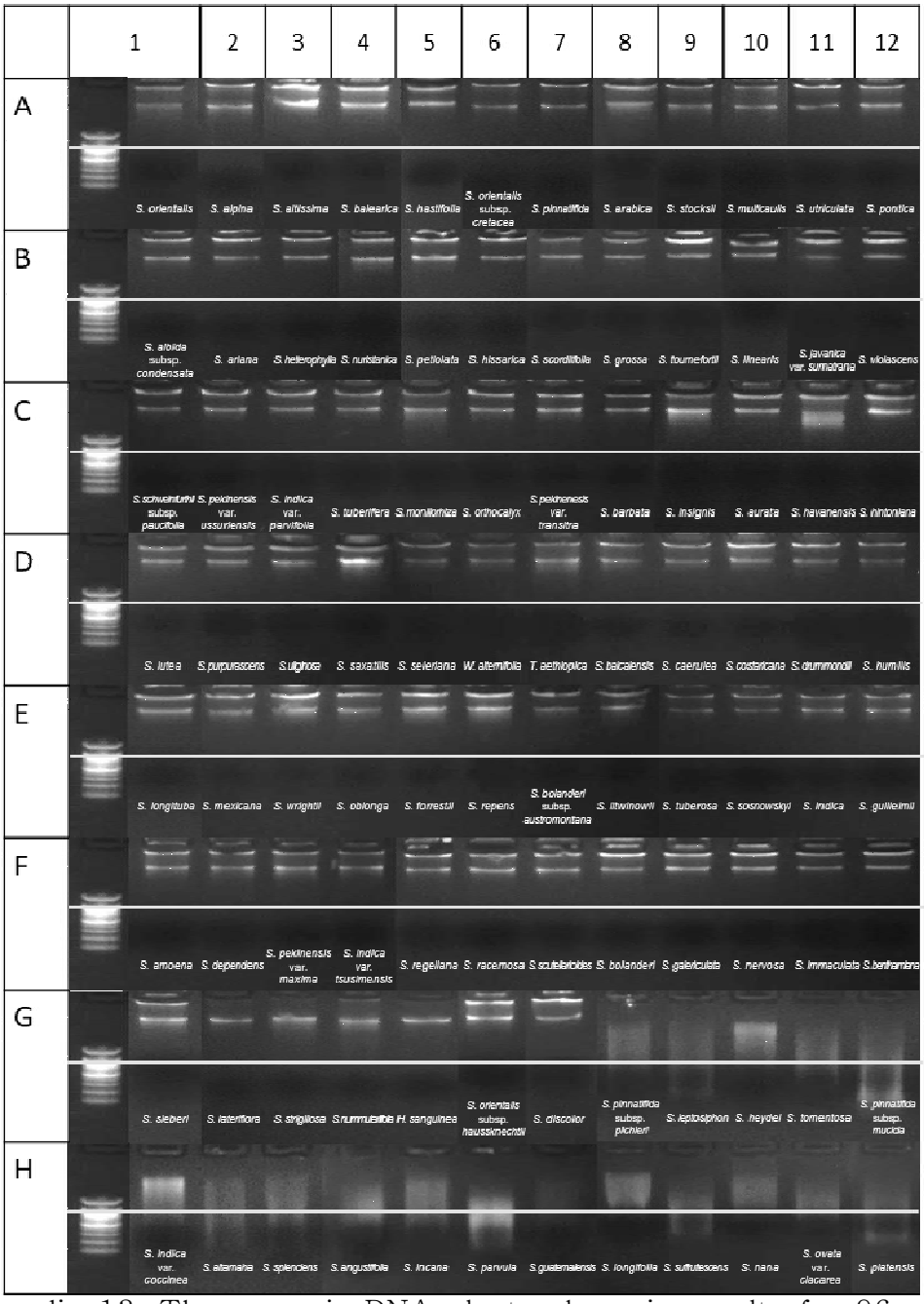
DN15314_c0_g1_i1	AT4G35850	Pentatricopeptide repeat (PPR) superfamily protein	o	o		o	1356
DN15404_c0_g2_i1	AT1G04690	potassium channel beta subunit 1	o	o	o		952
DN15419_c0_g1_i1	AT3G18390	CRS1 / YhbY (CRM) domain-containing protein			o	o	2472
DN15425_c0_g1_i1	AT5G46580	pentatricopeptide (PPR) repeat-containing protein	o		o		2190
DN15517_c0_g2_i1	AT1G54650	Methyltransferase family protein	o			o	984
DN15732_c0_g1_i1	AT5G45760	Transducin/WD40 repeat-like superfamily protein	o		o	o	1092
DN15736_c0_g1_i1	AT3G19630	Radical SAM superfamily protein	o	o	o	o	1113
DN15785_c0_g2_i2	AT2G22870	P-loop containing nucleoside triphosphate hydrolases superfamily protein	o		o		909
DN15851_c3_g2_i1	AT2G27790	RNA-binding (RRM/RBD/RNP motifs) family protein		o		o	678
DN15969_c0_g1_i5	AT1G78800	UDP-Glycosyltransferase superfamily protein	o	o		o	1227
DN16030_c0_g2_i2	AT4G12790	P-loop containing nucleoside triphosphate hydrolases superfamily protein	o			o	803
DN16055_c0_g1_i1	AT4G35250	NAD(P)-binding Rossmann-fold superfamily protein	o	o		o	1191
DN16070_c1_g1_i1	AT5G23340	RNI-like superfamily protein	o		o		1245
DN16086_c1_g1_i1	AT1G19800	trigalactosyldiacylglycerol 1			o	o	1242
DN16095_c0_g1_i1	AT4G36390	Methylthiotransferase	o	o		o	1899
DN16168_c1_g1_i1	AT4G08810	calcium ion binding	o	o	o	o	1695
DN16194_c1_g3_i3	AT3G19895	RING/U-box superfamily protein	o			o	1396
DN16229_c0_g1_i2	AT5G38530	tryptophan synthase beta type 2	o		o	o	1488
DN16275_c0_g1_i1	AT1G77030	hydrolases, acting on acid anhydrides, in phosphorus-containing anhydrides	o	o	o	o	2400
DN16317_c1_g2_i4	AT2G21280	NAD(P)-binding Rossmann-fold superfamily protein			o	o	1074
DN16388_c0_g1_i4	NONE			o		o	1227
DN16405_c0_g1_i5	AT1G71790	Subunits of heterodimeric actin filament capping protein Capz superfamily	o	o	o	o	777
DN16418_c0_g1_i5	AT4G01040	Glycosyl hydrolase superfamily protein	o		o	o	1302
DN16535_c0_g1_i5	AT3G62370	heme binding		o	o		1006
DN16733_c0_g1_i2	AT2G42450	alpha/beta-Hydrolases superfamily protein	o	o	o		1623
DN16738_c0_g1_i2	NONE				o	o	1122
DN16762_c0_g2_i2	AT5G24120	sigma factor E	o		o	o	1408
DN16782_c0_g1_i1	AT1G09960	sucrose transporter 4	o			o	1554
DN16786_c0_g1_i1	AT2G28100	alpha-L-fucosidase 1	o	o	o		1689
DN16868_c0_g1_i4	AT5G37360	UNKNOWN	o		o		924
DN16890_c0_g1_i1	AT4G33060	Cyclophilin-like peptidyl-prolyl cis-trans isomerase family protein	o	o	o	o	1473
DN16894_c2_g1_i2	AT3G25900	Homocysteine S-methyltransferase family protein	o		o	o	1131
DN16959_c0_g1_i3	AT1G07230	non-specific phospholipase C1	o	o			1734
DN16985_c1_g1_i1	AT1G09680	Pentatricopeptide repeat (PPR) superfamily protein	o		o	o	1164
DN17066_c0_g1_i2	AT2G42920	Pentatricopeptide repeat (PPR-like) superfamily protein	o		o	o	1415
DN17136_c2_g2_i1	AT1G62780	UNKNOWN	o		o	o	717
DN17160_c0_g1_i2	AT2G39670	Radical SAM superfamily protein	o	o		o	1254
DN17217_c0_g1_i1	AT3G52640	Zn-dependent exopeptidases superfamily protein	o	o		o	2010
DN17266_c0_g2_i2	AT5G62530	aldehyde dehydrogenase 12A1	o		o	o	1650
DN17286_c0_g1_i2	AT4G34030	3-methylcrotonyl-CoA carboxylase	o	o	o	o	1719
DN17366_c0_g12_i1	AT2G38270	CAX-interacting protein 2	o			o	891
DN17369_c0_g1_i3	AT3G28460	methyltransferases	o		o	o	933
DN17379_c3_g4_i7	AT5G44450	methyltransferases	o		o		929
DN17532_c0_g1_i1	AT4G15420	Ubiquitin fusion degradation UFD1 family protein	o	o	o		1809
DN17587_c1_g2_i4	AT4G12610	transcription activators/ DNA binding /RNApolymeraseIItranscriptionfactors/catalytics/transcriptioninitiationfactors		o		o	1617
DN17667_c0_g1_i1	AT5G02860	Pentatricopeptide repeat (PPR) superfamily protein	o	o		o	2445
DN17671_c0_g1_i4	AT5G48960	HAD-superfamily hydrolase, subfamily IG, 5'-nucleotidase	o			o	1908
DN17693_c0_g1_i2	AT3G62910	Peptide chain release factor 1			o	o	1254
DN17709_c0_g1_i1	AT1G21790	TRAM, LAG1 and CLN8 (TLC) lipid-sensing domain containing protein	o	o		o	792
DN17742_c0_g1_i1	AT1G76400	Ribophorin I	o			o	1851

DN17759_c3_g8_i1	NONE			o	o	o	921
DN17829_c1_g1_i1	AT5G43280	delta(3,5),delta(2,4)-dienoyl-CoA isomerase 1	o	o		o	888
DN17840_c0_g2_i1	AT1G67680	SRP72 RNA-binding domain		o	o		1844
DN17868_c2_g1_i3	AT4G37040	methionine aminopeptidase 1D	o	o	o	o	1062
DN17887_c2_g7_i1	AT2G15430	DNA-directed RNA polymerase family protein	o	o	o		918
DN17929_c0_g4_i1	AT4G19180	GDA1/CD39 nucleoside phosphatase family protein	o		o	o	2292
DN17957_c0_g1_i4	AT1G43620	UDP-Glycosyltransferase superfamily protein	o	o		o	1689
DN17966_c0_g1_i6	AT4G05090	Inositol monophosphatase family protein	o		o	o	1278
DN18021_c1_g1_i5	AT3G11210	SGNH hydrolase-type esterase superfamily protein	o	o			795
DN18032_c0_g1_i1	AT1G05060	UNKNOWN	o		o	o	768
DN18037_c4_g1_i1	AT1G64770	NDH-dependent cyclic electron flow 1		o	o	o	1041
DN18056_c2_g5_i2	AT4G17370	Oxidoreductase family protein	o		o		1041
DN18064_c0_g2_i1	AT5G48220	Aldolase-type TIM barrel family protein	o		o	o	1194
DN18105_c4_g5_i2	AT5G43430	electron transfer flavoprotein beta	o	o		o	756
DN18206_c0_g2_i1	AT1G05940	cationic amino acid transporter 9	o	o	o	o	1692
DN18221_c0_g9_i2	AT5G54080	homogentisate 1,2-dioxygenase	o	o	o	o	1383
DN18237_c2_g11_i1	AT5G39710	Tetratricopeptide repeat (TPR)-like superfamily protein	o		o		1943
DN18388_c0_g1_i1	AT2G01320	ABC-2 type transporter family protein	o	o	o	o	1898
DN18395_c0_g3_i3	AT1G08030	tyrosylprotein sulfotransferase	o			o	1428
DN18512_c0_g1_i2	AT3G10150	purple acid phosphatase 16	o	o	o		1131
DN18514_c0_g6_i1	AT3G56940	dicarboxylate diiron protein, putative (Crd1)	o			o	1025
DN18541_c0_g2_i2	AT1G61520	photosystem I light harvesting complex gene 3	o		o		822
DN18570_c2_g3_i1	NONE			o	o	o	1656
DN18808_c0_g4_i1	AT4G37380	Tetratricopeptide repeat (TPR)-like superfamily protein	o	o		o	1875
DN18853_c0_g2_i1	AT5G07910	Leucine-rich repeat (LRR) family protein	o		o	o	486
DN19014_c0_g2_i1	AT3G23800	selenium-binding protein 3		o		o	1458
DN19072_c0_g3_i1	AT5G03430	phosphoadenosine phosphosulfate (PAPS) reductase family protein	o	o	o	o	1530
DN19315_c1_g7_i6	AT3G48880	RNI-like superfamily protein	o	o		o	918
DN19396_c0_g1_i1	AT2G04560	transferases, transferring glycosyl groups	o		o	o	828
DN19521_c2_g1_i1	AT2G16880	Pentatricopeptide repeat (PPR) superfamily protein	o		o		1956
DN19680_c1_g1_i1	AT4G12700	UNKNOWN			o	o	1596
DN20511_c0_g1_i1	AT5G58480	O-Glycosyl hydrolases family 17 protein	o	o	o		1494
DN21706_c0_g1_i1	AT5G02820	Spo11/DNA topoisomerase VI, subunit A protein	o		o	o	1132
DN4521_c0_g1_i1	AT5G48840	homolog of bacterial PANC	o	o	o	o	942
DN5392_c0_g2_i1	AT5G04520	Protein of unknown function DUF455	o			o	891
DN6792_c0_g2_i1	AT5G59900	Pentatricopeptide repeat (PPR) superfamily protein	o		o	o	922
DN7031_c0_g1_i1	AT1G26500	Pentatricopeptide repeat (PPR) superfamily protein	o			o	1536
DN713_c0_g1_i1	AT3G20780	topoisomerase 6 subunit B	o	o	o	o	2020
DN7441_c0_g3_i1	AT2G18940	Tetratricopeptide repeat (TPR)-like superfamily protein	o	o	o	o	1669
DN8570_c0_g1_i1	AT3G29240	Protein of unknown function (DUF179)	o		o	o	924
DN8665_c0_g1_i1	AT2G20790	clathrin adaptor complexes medium subunit family protein	o	o	o	o	1062
DN8687_c0_g2_i1	AT1G69523	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein		o	o	o	927
DN8707_c0_g1_i1	AT1G32160	Protein of unknown function (DUF760)	o	o		o	1227
DN9477_c0_g1_i1	AT3G23020	Tetratricopeptide repeat (TPR)-like superfamily protein	o	o	o		1953
DN9658_c0_g4_i1	AT3G06920	Tetratricopeptide repeat (TPR)-like superfamily protein	o	o	o	o	1257
DN9710_c0_g2_i1	AT1G52780	Protein of unknown function (DUF2921)	o		o	o	2856
DN9853_c0_g1_i1	AT1G09300	Metallopeptidase M24 family protein	o	o	o	o	1400
DN9887_c0_g1_i1	AT3G09650	Tetratricopeptide repeat (TPR)-like superfamily protein	o		o	o	2199
DN9927_c0_g1_i1	AT5G65000	Nucleotide-sugar transporter family protein	o	o		o	996

Appendix 12. The voucher information of 96 samples used in Hyb-Seq analyses.

Taxon name	Country	Collector and collector number (Herbarium code)
Holmskioldia sanguinea Retz.	Korea	Y. Lee 2019-0001 (SWU)
Scutellaria albida subsp. *condensata* (Rech.f.) J. R. Edm.	Iraq	R Wheeler Haines s.n. (K)
Scutellaria alpina L.	Spain	Halliwell, Mason & Smallcombe s.n. (K)
Scutellaria altamaha Small	USA	R. Kral 51180 (MO)
Scutellaria altissima L.	Macedonia	J.D.A.Stainton 7826 (K)
Scutellaria amoena C. H. Wright	China	s. n. 65-0185 (KUN)
Scutellaria angustifolia Pursh	USA	R. Olmstead 558 (NYBG)
Scutellaria arabica Jaub. & Spach	Saudi Arabia	S.Collenette 6659 (K)
Scutellaria ariana Hedge	Afghanistan	P.Furse 5873 (K)
Scutellaria aurata Lem.	Ecuador	D. Neill & H. Vargar 12251 (MO)
Scutellaria baicalensis Georgi	USA	R. Olmstead 98-54 (WTU)
Scutellaria balearica Barcelo	Spain	Marcos 67 (K)
Scutellaria barbata D. Don	Korea	S.Kim 2019-056 (SWU)
Scutellaria benthamiana (Mansf.) Epling	Peru	Irene Shonle 2 (MO)
Scutellaria bolanderi A.Gray	USA	R. Olmstead 395 (WTU)
Scutellaria bolanderi subsp. *austromontana* Epling	USA	T. S. Eilias 10156 (PE)
Scutellaria caerulea Moc. & Sesse ex Benth.	Mexico	M. Egger & D. Tank 1191 (WTU)
Scutellaria costaricana H. Wendl.	USA	R. Olmstead s.n. (WTU)
Scutellaria dependens Maxim.	Japan	K. Yonekura 12880 (KYO)
Scutellaria discolor Colebr.	Myanmar	J. Murata & al. 25593 (KYO)
Scutellaria drummondii Benth.	USA	J. Bacon 2117 (WTU)
Scutellaria forrestii Diels	China	Liu Zhenwen & Zhoujing 1041 (KH)
Scutellaria galericulata L.	USA	R. Olmstead s.n. (WTU)
Scutellaria grossa Wall.	Nepal	C.Grey-Wilson & Phillips 359 (K)
Scutellaria guatemalensis Leonard	Mexico	Demetrio Alvarez 11028 (NYBG)
Scutellaria guilielmii A. Gray	Japan	Taku Miyazaki 605186 (KH)
Scutellaria hastifolia L.	Czech republic	M.Dvorakova, M.Smejkel & J.Vichar 1552 (K)
Scutellaria havanensis Jacq.	Bahamas	D. S. Correll 46132 (MO)
Scutellaria heterophylla Montbret & Aucher ex Benth.	Turkey	B.Balansa 498 (K)
Scutellaria heydei Hook. f.	India	M.Grace et al. 83 (K)
Scutellaria hintoniana Epling	Mexico	A. de Avila 662 (MO)
Scutellaria hissarica B. Fedtsch.	Tajikistan	L.P.Markova et al. s.n. (K)
Scutellaria humilis R.Br.	Australia	B. Lepschi & Lally 1717 (WTU)
Scutellaria immaculata Nevski ex Juz.	Uzbekistan	Vatolkina 4757 (K)
Scutellaria incana Spreng.	USA	R. Dale Thomas 150388 (NYBG)
Scutellaria indica L.	Taiwan	B.J.Conn 4109 (K)
Scutellaria indica var. *coccinea* S. Kim & S. Lee	Korea	S. Kim 2015-0298 (SWU)
Scutellaria indica var. *parvifolia* Makino	Arkansas	F. H. Utech M.Ohara 83-446 (KUN)
Scutellaria indica var. *tsusimensis* H. Hara	Korea	SWU S. Kim 2008-008 (SWU)
Scutellaria insignis Nakai	Korea	S. Kim 2015-0110 (SWU)
Scutellaria javanica var. *sumatrana* (Miq.) Bock	Malaysia	A.Lamb & A.Surat 375 (K)
Scutellaria lateriflora L.	USA	George F. buddell II 2352 (NYBG)
Scutellaria leptosiphon Nevski	Tadjikistan	R. Markova et al s.n. (K)
Scutellaria linearis Benth.	India	R.W.Parker 3396 (K)
Scutellaria litwinowii Bornm. & Sint.	Soviet	M. Popov 233 (PE)
Scutellaria longifolia Benth.	s. n.	J. J. Castillo & J. M. Vargas 2722 (NYBG)
Scutellaria longituba Koidz.	USA	R. Olmstead s.n. (WTU)
Scutellaria lutea J. D. Sm.	Guatemala	M. Perez 1532 (MO)
Scutellaria mexicana (Torr.) A. J. Paton	USA	W. Messinger 257 (WTU)

Scutellaria moniliorhiza Kom.	Korea	J. Kim & Y. Lee 2019-002 (SWU)
Scutellaria multicaulis Boiss.	Afghanistan	Rechinger 19308 (K)
Scutellaria nana A. Gray	USA	M. Bourell 6012 (PE)
Scutellaria nervosa Pursh	USA	S. Wagstaff 88-60 (WTU)
Scutellaria nummulariifolia Hook. f.	Argentina	R. G. Olmstead 2016-1 42370 (MO)
Scutellaria nuristanica Rech. f.	Afghanistan	Podlech 16150 (K)
Scutellaria oblonga Benth.	Ceylon	Wight 2141 (K)
Scutellaria orientalis L.	Georgia	P. Davis 33797 (K)
Scutellaria orientalis subsp. *cretacea* (Boiss. & Hausskn.) J.R.Edm.	Iraq	J.B.Gillett 10946 (K)
Scutellaria orientalis subsp. *haussknechtii* (Boiss.) J.R.Edm.	Turkey	P. Sintenis 1134 (K)
Scutellaria orthocalyx Hand.-Mazz.	Korea	Y. Lee 2019-003 (SWU)
Scutellaria ovata var. *clacarea* (Epling) Gleason	China	R. C. Clark 22239 (PE)
Scutellaria parvula Michx.	USA	R. Dale Thomas & Cecil Slaughter 103883 (NYBG)
Scutellaria pekinenesis var. *transitra* (Makino) H. Hara	Korea	S. Kim 2015-0109 (SWU)
Scutellaria pekinensis var. *maxima* S. Kim & S. Lee	Korea	S. Kim 2015-0108 (SWU)
Scutellaria pekinensis var. *ussuriensis* (Regel) Hand.-Mazz.	Russia	P. G. Gorovoy s. n. (KH)
Scutellaria petiolata Hemsl. ex Lace & Prain	Pakistan	Crookshank 155 (K)
Scutellaria pinnatifida A.Ham.	Turkey	P.Davis 18802 (K)
Scutellaria pinnatifida subsp. *mucida* (Stapf) Rech.f.	Persia	K. H. Rechinger 43494 (K)
Scutellaria pinnatifida subsp. *pichleri* (Stapf) Rech.f.	Iraq	A.Rawi 12254 (K)
Scutellaria platensis Speg.	Argentina	B. Sparre 341 (K)
Scutellaria pontica K.Koch	Turkey	P.Davis & I. Hedge s.n. (K)
Scutellaria purpurascens Sw.	Bolivia	J. S. Solomo 8558 (MO)
Scutellaria racemosa Pers.	Paraguay	M. Pena-Chocarro et al. 2256 (MO)
Scutellaria regeliana Nakai	Korea	s. n. (KYO)
Scutellaria repens Buch.-Ham. ex D.Don	Nepal	C. S. Chang, B. Y. Choi, S. P. Hong, K. S. Jang, U. Kang, S. Y. Kim, J. S. Lee, J. H. Park NE020279 (KH)
Scutellaria saxatilis Riddell	USA	Q. Metheny s.n. None (MO)
Scutellaria schweinfurthii subsp. *paucifolia* (Baker) A.J.Paton	Zambia	Sanane 844 (K)
Scutellaria scordiifolia Fisch. ex Schrank	Mongolia	C.Jeffey 1427 (K)
Scutellaria scutellarioides (Kunth) Harley	Ecuador	J. A. Hart 1616 (MO)
Scutellaria seleriana Loes.	Mexico	N. Martinez Melendez 1126 1126 (MO)
Scutellaria sieberi Benth.	Greece	J.K.Morton s.n. (K)
Scutellaria sosnowskyi Takht.	Turkey	Mitchell 3205 (K)
Scutellaria splendens Link, Klotzsch & Otto	Mexico	F. J. Rendon Sandoval 571 571 (MO)
Scutellaria stocksii Boiss.	Afghanistan	RB & E. Gibbons 242 (K)
Scutellaria strigillosa Hemsl.	Korea	S. Kim 20140705-1 (SWU)
Scutellaria suffrutescens S. Watson	Mexico	James L. Reveal 3390 (NYBG)
Scutellaria tomentosa Bertol.	Iran	Rechinger 46863 (K)
Scutellaria tournefortii Benth.	Iran	T.F.Hewer 3967 (K)
Scutellaria tuberifera C. Y. Wu & C. Chen	Korea	S. Kim 2019-025 (SWU)
Scutellaria tuberosa Benth.	USA	R. Olmstead 660 (WTU)
Scutellaria uliginosa A. St.-Hil. ex Benth.	Paraguay	E. M. Zardini & L. Guerrero 56594 (MO)
Scutellaria utriculata Labill.	Lebanon	G.Samulsson 1586 (K)
Scutellaria violascens Gurke	Ethiopia	Friis et al. 12332 (K)
Scutellaria wrightii A. Gray	USA	J. Bacon 2117A (WTU)
Tinnea aethiopica Kotschy ex Hook. f.	South Africa	s. n. (WTU)
Wenchengia alternifolia C. Y. Wu & S. Chow	China	S.K. Liu 28220 (WTU)



Appendix 13. The genomic DNA electrophoresis results for 96 samples (1 ul of gDNA was electrophoresed for 10 minutes at 135 V on a 1.3% agarose gel). Grey lines indicate 1,000 bp, and samples representing G8-H12 were reduced shearing.

Appendix 14. Statistics of sequencing and aseembly results by HybPiper.

Taxa	No. of paired reads	No. of reads	Yield(bases)	Over Q30-Bases(%)	No. of trimmed reads	Trimmed %	Reads Mapped	PctOn Target	Genes Mapped	Genes Mapped %	Genes With Contigs	Genes With Seqs	GenesAt 25pct	GenesAt 50pct	GenesAt 75pct	Genesat 150pct	Paralog Warnings
H.sanguinea	4,985,152	9,970,304	1,495,545,600	83.2	9,708,018	97.4	5,842,541	64.9	641	100	641	640	640	635	620	0	25
S. albida subsp. *condensata*	501,113	1,002,226	150,333,900	79.6	982,406	98.0	195,444	21.9	641	100	636	635	631	590	426	0	5
S. alpina	376,798	753,596	113,039,400	80.4	736,138	97.7	83,250	12.4	641	100	432	414	253	124	48	0	0
S. altamaha	5,832,372	11,664,744	1,749,711,600	77.1	11,487,964	98.5	8,337,887	81.3	641	100	641	629	622	612	588	0	133
S. altissima	666,101	1,332,202	199,830,300	79.2	1,305,796	98.0	290,436	22.2	640	99.8	637	637	636	625	563	0	6
S. amoena	504,868	1,009,736	151,460,400	79.8	984,958	97.6	176,505	19.6	641	100	232	209	68	16	4	0	0
S. angustifolia	4,076,564	8,153,128	1,222,969,200	79.4	8,023,050	98.4	5,832,495	80.3	641	100	641	640	640	639	635	0	11
S. arabica	4,605,083	9,210,166	1,381,524,900	82.0	8,957,802	97.3	5,653,368	67.9	641	100	637	637	635	615	529	0	4
S. ariana	586,272	1,172,544	175,881,600	78.0	1,149,326	98.0	47,716	4.6	641	100	217	206	97	39	16	0	0
S. aurata	387,154	774,308	116,146,200	72.6	760,268	98.2	184,603	27.9	641	100	637	637	623	549	364	0	6
S. baicalensis	492,882	985,764	147,864,600	73.5	968,238	98.2	243,511	28.7	641	100	640	640	640	635	564	0	7
S. balearica	425,149	850,298	127,544,700	78.7	829,600	97.6	53,573	7.1	641	100	352	336	193	82	36	0	2
S. barbata	6,276,144	12,552,288	1,882,843,200	81.6	12,310,756	98.1	8,591,860	73.1	641	100	641	641	641	641	624	0	15
S. benthamiana	309,349	618,698	92,804,700	76.7	604,342	97.7	100,535	18.6	639	99.7	362	350	230	99	25	0	0
S. bolanderi	327,076	654,152	98,122,800	74.3	640,786	98.0	73,116	13.0	641	100	388	382	259	132	58	0	0
S. bolanderi subsp. *austromontana*	234,037	468,074	70,211,100	78.1	458,032	97.9	29,277	7.1	640	99.8	82	67	33	9	3	0	0
S. caerulea	329,016	658,032	98,704,800	75.9	644,022	97.9	22,322	3.9	635	99.1	112	98	26	8	1	0	0
S. costaricana	419,742	839,484	125,922,600	79.2	821,990	97.9	213,219	28.7	641	100	639	639	639	614	508	0	6
S. dependens	17,372,627	34,745,254	5,211,788,100	83.5	33,919,706	97.6	23,538,422	73.4	641	100	641	638	638	636	622	0	11
S. discolor	419,806	839,612	125,941,800	77.7	822,412	98.0	331,376	44.8	641	100	637	636	631	600	469	0	31
S. drummondii	487,299	974,598	146,189,700	77.1	958,896	98.4	262,066	30.6	641	100	635	633	623	585	463	0	28
S. forrestii	305,205	610,410	91,561,500	76.8	597,368	97.9	107,128	20.1	641	100	614	611	550	392	211	0	1
S. galericulata	286,937	573,874	86,081,100	80.0	561,076	97.8	166,636	32.5	641	100	335	325	225	117	47	0	0
S. grossa	13,208,701	26,417,402	3,962,610,300	82.0	25,810,922	97.7	17,201,218	71.9	641	100	640	640	639	635	625	1	8
S. guatemalensis	5,200,682	10,401,364	1,560,204,600	80.3	10,218,398	98.2	7,349,091	79.0	641	100	641	640	639	638	635	0	22
S. guilielmii	6,805,087	13,610,174	2,041,526,100	84.1	13,279,264	97.6	8,353,062	67.2	641	100	626	623	608	532	383	0	6
S. hastifolia	424,976	849,952	127,492,800	78.8	829,866	97.6	78,015	10.4	558	87.1	62	55	26	13	5	0	0
S. havanensis	608,949	1,217,898	182,684,700	76.7	1,196,184	98.2	324,674	30.3	629	98.1	298	283	148	51	11	0	1
S. heterophylla	445,338	890,676	133,601,400	75.3	870,750	97.8	14,644	1.9	623	97.2	42	33	4	0	0	0	0
S. heydei	6,554,409	13,108,818	1,966,322,700	80.9	12,867,588	98.2	9,576,240	81.4	641	100	641	641	640	640	637	0	16

S. hintoniana	290,953	581,906	87,285,900	67.1	571,126	98.2	84,154	17.5	641	100	589	586	485	303	156	0	3
S. hissarica	652,390	1,304,780	195,717,000	77.8	1,279,512	98.1	267,014	23.3	640	99.8	639	639	637	625	537	0	6
S. humilis	246,743	493,486	74,022,900	76.8	486,324	98.6	119,880	27.4	640	99.8	437	427	236	81	28	0	1
S. immaculata	1,264,964	2,529,928	379,489,200	81.8	2,467,780	97.5	1,608,809	70.4	641	100	641	641	641	639	617	0	4
S. incana	4,693,609	9,387,218	1,408,082,700	82.2	9,209,712	98.1	6,577,375	77.6	641	100	641	634	633	624	606	1	147
S. indica	2,722,453	5,444,906	816,735,900	70.8	5,353,416	98.3	2,907,227	63.0	641	100	641	641	640	638	626	0	21
S. indica var. *coccinea*	6,131,290	12,262,580	1,839,387,000	77.4	12,029,932	98.1	8,504,224	78.9	641	100	641	641	640	640	638	0	22
S. indica var. *parvifolia*	632,409	1,264,818	189,722,700	71.4	1,238,988	98.0	144,291	13.5	641	100	632	631	595	466	270	0	4
S. indica var. *tsusimensis*	13,741,431	27,482,862	4,122,429,300	83.6	26,774,014	97.4	17,501,394	70.1	641	100	641	641	641	641	635	0	7
S. insignis	28,847,668	57,695,336	8,654,300,400	83.3	56,210,058	97.4	37,365,796	71.5	641	100	641	641	641	641	638	0	9
S. javanica var. *sumatrana*	536,390	1,072,780	160,917,000	72.3	1,052,196	98.1	9,639	1.1	557	86.9	37	10	1	0	0	0	0
S. lateriflora	12,466,490	24,932,980	3,739,947,000	83.9	24,320,150	97.5	16,589,466	73.2	641	100	641	641	641	638	618	1	21
S. leptosiphon	9,096,253	18,192,506	2,728,875,900	68.1	17,940,824	98.6	13,084,615	86.1	641	100	641	640	639	639	635	0	13
S. linearis	515,233	1,030,466	154,569,900	74.6	1,012,944	98.3	13,020	1.5	625	97.5	59	51	6	0	0	0	0
S. litwinowii	376,237	752,474	112,871,100	78.6	734,342	97.6	205,228	30.2	641	100	586	582	493	319	150	0	2
S. longifolia	5,309,876	10,619,752	1,592,962,800	82.8	10,403,500	98.0	7,367,633	76.6	641	100	641	641	640	638	634	0	19
S. longituba	228,027	456,054	68,408,100	78.7	449,338	98.5	79,534	19.6	641	100	502	496	369	194	91	0	2
S. lutea	340,023	680,046	102,006,900	75.0	666,384	98.0	101,948	17.3	641	100	585	583	489	316	156	0	6
S. mexicana	325,233	650,466	97,569,900	78.2	636,226	97.8	38,463	6.7	640	99.8	171	142	35	6	3	0	0
S. moniliorhiza	35,837,523	71,675,046	10,751,256,900	84.3	70,046,632	97.7	48,590,504	74.2	641	100	641	640	640	640	638	0	7
S. multicaulis	2,183,214	4,366,428	654,964,200	75.2	4,288,746	98.2	2,765,326	71.8	641	100	641	641	641	640	635	0	12
S. nana	5,336,091	10,672,182	1,600,827,300	81.5	10,465,680	98.1	7,453,984	77.7	641	100	641	641	641	641	639	0	16
S. nervosa	313,523	627,046	94,056,900	74.6	611,698	97.6	88,173	16.3	641	100	534	531	411	256	127	0	3
S. nummulariifolia	12,649,934	25,299,868	3,794,980,200	82.4	24,825,410	98.1	18,335,976	80.2	641	100	641	613	591	555	510	0	282
S. nuristanica	413,697	827,394	124,109,100	79.5	811,418	98.1	14,069	1.9	624	97.3	37	32	10	3	2	0	0
S. oblonga	390,199	780,398	117,059,700	77.3	759,652	97.3	4,150	0.6	281	43.8	9	1	0	0	0	0	0
S. orientalis	22,101,081	44,202,162	6,630,324,300	83.2	43,140,018	97.6	30,863,736	75.9	641	100	641	641	639	637	630	0	3
S. orientalis subsp. cretacea	446,519	893,038	133,955,700	70.1	876,374	98.1	257,302	34.2	641	100	639	639	639	630	550	0	5
S. orientalis subsp. *haussknechtii*	260,217	520,434	78,065,100	68.3	510,890	98.2	110,272	25.4	641	100	612	611	537	373	192	0	5
S. orthocalyx	13,103,257	26,206,514	3,930,977,100	82.8	25,620,752	97.8	17,626,155	74.2	641	100	641	641	641	640	638	0	10
S. ovata var. *clacarea*	5,658,810	11,317,620	1,697,643,000	77.8	11,156,670	98.6	8,185,226	81.9	641	100	641	628	621	609	589	1	123
S. parvula	3,984,475	7,968,950	1,195,342,500	75.3	7,854,922	98.6	5,721,247	82.5	641	100	641	641	641	640	631	0	13
S. pekinenesis var. *transitra*	8,421,937	16,843,874	2,526,581,100	83.2	16,434,928	97.6	10,618,735	69.5	641	100	641	641	639	637	630	1	10
S. pekinensis var. *maxima*	21,267,792	42,535,584	6,380,337,600	83.4	41,471,330	97.5	28,239,230	73.1	641	100	641	641	641	640	635	0	8
S. pekinensis var. *ussuriensis*	359,684	719,368	107,905,200	74.0	705,454	98.1	16,575	2.7	606	94.5	73	53	5	0	0	0	0

S. petiolata	493,307	986,614	147,992,100	77.4	967,062	98.0	54,990	6.4	639	99.7	418	407	259	106	41	0	0
S. pinnatifida	819,052	1,638,104	245,715,600	78.4	1,604,856	98.0	810,918	55.0	641	100	640	640	638	637	587	0	10
S. pinnatifida subsp. *mucida*	3,099,948	6,199,896	929,984,400	53.3	6,133,416	98.9	4,139,223	87.5	641	100	641	640	640	639	633	0	9
S. pinnatifida subsp. *pichleri*	21,095,845	42,191,690	6,328,753,500	78.0	41,419,140	98.2	30,207,099	81.1	641	100	641	641	639	638	636	0	18
S. platensis	3,081,408	6,162,816	924,422,400	59.4	6,093,594	98.9	4,103,156	84.0	641	100	641	641	641	639	625	1	26
S. pontica	344,109	688,218	103,232,700	76.0	672,844	97.8	24,967	4.2	628	98.0	89	70	30	12	5	0	0
S. purpurascens	397,648	795,296	119,294,400	77.3	779,890	98.1	193,905	27.8	641	100	634	631	619	543	366	0	4
S. racemosa	1,406,773	2,813,546	422,031,900	82.4	2,751,698	97.8	1,751,902	68.7	641	100	641	641	641	637	617	0	9
S. regeliana	59,545	119,090	17,863,500	63.6	117,118	98.3	62,806	65.1	641	100	515	508	288	82	21	0	0
S. repens	190,822	381,644	57,246,600	79.4	373,748	97.9	45,070	13.3	640	99.8	200	193	78	40	8	0	0
S. saxatilis	1,333,542	2,667,084	400,062,600	81.9	2,607,948	97.8	1,298,042	52.8	641	100	640	640	639	617	511	0	5
S. schweinfurthii subsp. *paucifolia*	14,253	28,506	4,275,900	82.3	27,806	97.5	18,985	73.6	638	99.5	85	74	13	3	1	0	0
S. scordiifolia	756,963	1,513,926	227,088,900	76.6	1,487,820	98.3	537,676	40.5	641	100	641	641	641	637	618	0	7
S. scutellarioides	337,559	675,118	101,267,700	70.7	662,466	98.1	214,685	37.6	641	100	616	615	582	453	249	0	3
S. seleriana	466,818	933,636	140,045,400	78.1	913,528	97.9	259,548	31.5	641	100	636	636	632	599	474	0	9
S. sieberi	4,085,699	8,171,398	1,225,709,700	81.4	7,972,054	97.6	5,409,027	72.7	641	100	639	639	638	635	625	0	11
S. sosnowskyi	397,182	794,364	119,154,600	74.1	778,902	98.1	145,597	21.3	641	100	632	631	608	495	310	0	6
S. splendens	8,972,600	17,945,200	2,691,780,000	80.5	17,582,492	98.0	12,456,145	77.7	641	100	641	641	641	640	638	0	19
S. stocksii	33,557,085	67,114,170	10,067,125,500	83.4	65,466,270	97.5	45,055,299	73.6	641	100	641	641	641	641	638	1	2
S. strigillosa	9,247,060	18,494,120	2,774,118,000	81.8	18,104,182	97.9	13,405,267	80.5	641	100	641	641	641	640	634	1	15
S. suffrutescens	3,661,295	7,322,590	1,098,388,500	62.7	7,241,026	98.9	4,994,747	84.3	641	100	641	634	623	597	559	0	76
S. tomentosa	6,813,830	13,627,660	2,044,149,000	74.4	13,432,912	98.6	10,079,401	85.2	641	100	641	640	639	638	631	0	13
S. tournefortii	737,491	1,474,982	221,247,300	78.8	1,446,354	98.1	527,464	40.4	641	100	640	640	640	637	613	0	9
S. tuberifera	733,731	1,467,462	220,119,300	77.6	1,432,806	97.6	188,301	14.6	641	100	638	638	633	594	448	0	5
S. tuberosa	616,175	1,232,350	184,852,500	82.0	1,201,880	97.5	603,871	54.3	641	100	558	557	475	283	134	0	0
S. uliginosa	333,701	667,402	100,110,300	77.6	653,632	97.9	137,643	23.5	641	100	633	632	607	530	340	0	5
S. utriculata	477,295	954,590	143,188,500	79.8	934,596	97.9	7,943	0.9	581	90.6	14	11	0	0	0	0	0
S. violascens	466,720	933,440	140,016,000	76.8	915,498	98.1	169,191	20.7	641	100	639	638	631	577	414	0	5
S. wrightii	381,810	763,620	114,543,000	78.4	751,000	98.4	123,045	18.2	641	100	588	583	525	356	185	0	11
T. aethiopica	18,120,214	36,240,428	5,436,064,200	82.8	35,258,746	97.3	21,210,474	64.9	641	100	641	641	641	640	632	0	16
W. alternifolia	328,563	657,126	98,568,900	80.3	640,102	97.4	4,970	0.9	535	83.5	13	2	0	0	0	0	0

Appendix 15. Information for supercontigs of 272 loci obtained using AMAS.

Alignment_name	Arabidopsis ID	Function	No. of taxa	Alignment length	Total matrix cells	Undetermined characters	Missing percent	No. of variable sites	Proportion of variable sites	Parsimony informative sites	Proportion parsimony informative	AT content	GC content	A	C	G	T	K	M	R	Y	S	W	B	V	H	D	X	N	O	-	?
DN10117_c0_g1_i1	AT1G63810	UNKNOWN	79	7672	606088	170056	28.06	4677	0.61	2501	0.33	0.636	0.364	121395	72745	85979	155913	0	0	0	0	0	0	0	0	0	0	0	0	0	170056	0
DN10161_c0_g1_i1	AT2G35720	DNAJ heat shock N-terminal domain-containing protein	59	4383	258597	87207	33.72	2429	0.55	1295	0.30	0.64	0.36	47529	27845	33414	62602	0	0	0	0	0	0	0	0	0	0	0	0	0	87207	0
DN10173_c0_g2_i2	AT2G04740	ankyrin repeat family protein	61	2335	142435	35632	25.02	1051	0.45	547	0.23	0.596	0.404	29448	21679	21336	34340	0	0	0	0	0	0	0	0	0	0	0	0	0	35632	0
DN10505_c0_g2_i1	AT3G57190	peptide chain release factor, putative	77	2332	179564	30541	17.01	1426	0.61	869	0.37	0.622	0.378	43725	24776	31230	49292	0	0	0	0	0	0	0	0	0	0	0	0	0	30541	0
DN10654_c0_g1_i1	AT4G30840	Transducin/WD40 repeat-like superfamily protein	61	2036	124196	28315	22.80	1221	0.60	624	0.31	0.57	0.43	22814	20268	20559	32240	0	0	0	0	0	0	0	0	0	0	0	0	0	28315	0
DN10655_c0_g1_i1	AT5G03070	importin alpha isoform 9	78	4183	326274	97129	29.77	2844	0.68	1589	0.38	0.619	0.381	63908	37403	49029	78805	0	0	0	0	0	0	0	0	0	0	0	0	0	97129	0
DN10718_c0_g2_i1	AT4G21770	Pseudouridine synthase family protein	78	3117	243126	57450	23.63	1929	0.62	1177	0.38	0.626	0.374	50701	34919	34420	65636	0	0	0	0	0	0	0	0	0	0	0	0	0	57450	0
DN10838_c0_g1_i1	AT3G55260	beta-hexosaminidase 1	64	3769	241216	69290	28.73	2058	0.55	1146	0.30	0.617	0.383	47628	30746	34243	59309	0	0	0	0	0	0	0	0	0	0	0	0	0	69290	0
DN11234_c0_g1_i1	AT1G04850	ubiquitin-associated (UBA)/TS-N domain-containing protein	52	4582	238264	75873	31.84	2486	0.54	1319	0.29	0.627	0.373	46978	28062	32101	55250	0	0	0	0	0	0	0	0	0	0	0	0	0	75873	0
DN11241_c0_g1_i1	AT1G63850	BTB/POZ domain-containing protein	75	2146	160950	32676	20.30	1118	0.52	562	0.26	0.552	0.448	34205	25560	32009	36500	0	0	0	0	0	0	0	0	0	0	0	0	0	32676	0
DN11281_c0_g2_i1	AT3G48425	DNAse I-like superfamily protein	71	3087	219177	70639	32.23	1930	0.63	1201	0.39	0.635	0.365	45749	24577	29030	49182	0	0	0	0	0	0	0	0	0	0	0	0	0	70639	0
DN11363_c0_g3_i2	AT3G18630	uracil dna glycosylase	64	2981	190784	60537	31.73	1808	0.61	1112	0.37	0.626	0.374	38179	24094	24234	43740	0	0	0	0	0	0	0	0	0	0	0	0	0	60537	0
DN11368_c0_g1_i4	AT1G48310	chromatin remodeling factor18	75	7420	556500	189376	34.03	5071	0.68	2819	0.38	0.635	0.365	101460	64486	68581	132597	0	0	0	0	0	0	0	0	0	0	0	0	0	189376	0
DN11381_c0_g1_i2	AT3G48460	GDSL-like Lipase/Acylhydrolase superfamily protein	58	1400	81200	19096	23.52	816	0.58	514	0.37	0.513	0.487	15821	18276	12178	15829	0	0	0	0	0	0	0	0	0	0	0	0	0	19096	0
DN11430_c0_g1_i1	AT1G16280	RNA helicase 36	61	2732	166652	45595	27.36	1356	0.50	753	0.28	0.588	0.412	33806	22951	27073	37227	0	0	0	0	0	0	0	0	0	0	0	0	0	45595	0
DN11432_c0_g1_i1	AT5G13520	peptidase M1 family protein	76	2893	219868	45440	20.67	1673	0.58	906	0.31	0.548	0.452	42812	39579	39363	52674	0	0	0	0	0	0	0	0	0	0	0	0	0	45440	0
DN11436_c0_g1_i2	AT5G52030	TraB family protein	74	1819	134606	33015	24.53	1038	0.57	636	0.35	0.615	0.385	27756	18474	20453	34908	0	0	0	0	0	0	0	0	0	0	0	0	0	33015	0
DN11545_c0_g1_i1	AT1G55880	Pyridoxal-5'-phosphate-dependent enzyme family protein	69	2764	190716	49976	26.20	1834	0.66	1015	0.37	0.603	0.397	38164	24001	31751	46824	0	0	0	0	0	0	0	0	0	0	0	0	0	49976	0
DN11550_c0_g1_i1	AT5G65860	ankyrin repeat family protein	60	2147	128820	24978	19.39	1151	0.54	715	0.33	0.607	0.393	28616	18090	22472	34664	0	0	0	0	0	0	0	0	0	0	0	0	0	24978	0
DN11552_c0_g1_i1	AT3G66658	aldehyde dehydrogenase 22A1	77	4545	349965	111964	31.99	2838	0.62	1553	0.34	0.621	0.379	67288	43229	46563	80921	0	0	0	0	0	0	0	0	0	0	0	0	0	111964	0
DN11731_c1_g1_i2	AT3G46980	phosphate transporter 43	57	2986	170202	40041	23.53	1504	0.50	786	0.26	0.589	0.411	33056	26201	26836	44068	0	0	0	0	0	0	0	0	0	0	0	0	0	40041	0
DN11877_c0_g1_i1	AT1G74850	plastid transcriptionally active 2	70	3507	245490	43395	17.68	1829	0.52	840	0.24	0.575	0.425	55892	36424	49257	60522	0	0	0	0	0	0	0	0	0	0	0	0	0	43395	0
DN12002_c0_g2_i1	AT3G54720	Peptidase M28 family protein	52	3189	165828	41281	24.89	1589	0.50	911	0.29	0.565	0.435	29901	25405	28213	41028	0	0	0	0	0	0	0	0	0	0	0	0	0	41281	0
DN12024_c0_g2_i1	AT1G74030	enolase 1	75	2806	210450	57846	27.49	1582	0.56	861	0.31	0.59	0.41	42328	31162	31611	47503	0	0	0	0	0	0	0	0	0	0	0	0	0	57846	0
DN12042_c0_g1_i2	AT2G45030	Translation elongation factor EFG/EF2 proteinAT1G45332	58	5689	329962	104214	31.58	3073	0.54	1682	0.30	0.603	0.397	59496	39522	48037	78693	0	0	0	0	0	0	0	0	0	0	0	0	0	104214	0
DN12055_c0_g3_i1	AT4G38240	alpha-1,3-mannosyl-glycoprotein beta-1,2-N-acetylglucosaminyltransferase, putative	70	4313	301910	87564	29.00	2288	0.53	1294	0.30	0.64	0.36	58304	37324	39523	79195	0	0	0	0	0	0	0	0	0	0	0	0	0	87564	0
DN12152_c0_g2_i1	AT3G20810	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein	56	2730	152880	43126	28.21	1491	0.55	930	0.34	0.602	0.398	29094	20824	22989	36847	0	0	0	0	0	0	0	0	0	0	0	0	0	43126	0
DN12199_c0_g1_i1	AT1G55980	FAD/NAD(P)-binding oxidoreductase family protein	67	2478	166026	47978	28.90	1383	0.56	830	0.34	0.609	0.391	31344	21424	24561	40719	0	0	0	0	0	0	0	0	0	0	0	0	0	47978	0
DN12228_c0_g1_i2	AT3G59710	NAD(P)-binding Rossmann-fold superfamily protein	55	1259	69245	19611	28.32	767	0.61	448	0.36	0.592	0.408	14699	8737	11192	15006	0	0	0	0	0	0	0	0	0	0	0	0	0	19611	0

DN12263_c0_g2_i2	AT4G24750	Rhodanese/Cell cycle control phosphatase superfamily protein	75	3193	239475	82723	34.54	2167	0.68	1316	0.41	0.637	0.363	43192	26263	30295	57002	0	0	0	0	0	0	0	0	0	0	0	0	0	82723	0
DN12329_c0_g1_i1	AT5G44000	Glutathione S−transferase family protein	57	1424	81168	18508	22.80	827	0.58	498	0.35	0.571	0.429	18093	13189	13909	17469	0	0	0	0	0	0	0	0	0	0	0	0	0	18508	0
DN12355_c0_g1_i1	AT5G56740	histone acetyltransferase of the GNAT family 2	70	2766	193620	66056	34.12	2009	0.73	1133	0.41	0.611	0.389	35414	24124	25151	42875	0	0	0	0	0	0	0	0	0	0	0	0	0	66056	0
DN12467_c0_g1_i1	AT3G48500	Nucleic acid−binding, OB−fold−like protein	77	5369	413413	111012	26.85	3138	0.58	1901	0.35	0.607	0.393	87712	53266	62498	98925	0	0	0	0	0	0	0	0	0	0	0	0	0	111012	0
DN12477_c0_g1_i1	AT5G03860	malate synthase	60	2568	154080	30168	19.58	1445	0.56	803	0.31	0.564	0.436	35845	21171	32486	34410	0	0	0	0	0	0	0	0	0	0	0	0	0	30168	0
DN12536_c0_g1_i1	AT5G42470	UNKNOWN	72	3327	239544	69196	28.89	2139	0.64	1149	0.35	0.629	0.371	44692	31786	31191	62679	0	0	0	0	0	0	0	0	0	0	0	0	0	69196	0
DN12637_c0_g2_i1	AT2G03270	DNA−binding protein, putative	61	4658	284138	86892	30.58	2350	0.51	1387	0.30	0.623	0.377	56261	33733	39903	67349	0	0	0	0	0	0	0	0	0	0	0	0	0	86892	0
DN12737_c0_g1_i1	AT5G25480	DNA methyltransferase−2	66	2704	178464	59715	33.46	1642	0.61	934	0.35	0.628	0.372	32408	22113	21642	42586	0	0	0	0	0	0	0	0	0	0	0	0	0	59715	0
DN12801_c0_g1_i2	AT1G16020	Protein of unknown function (DUF1712)	76	3621	275196	81311	29.55	2068	0.57	1158	0.32	0.618	0.382	49717	33317	39982	70869	0	0	0	0	0	0	0	0	0	0	0	0	0	81311	0
DN12874_c0_g1_i2	AT1G14620	decoy	52	1964	102128	34390	33.67	1039	0.53	557	0.28	0.622	0.378	17556	12483	13108	24591	0	0	0	0	0	0	0	0	0	0	0	0	0	34390	0
DN12972_c0_g2_i1	AT4G32990	Transducin/WD40 repeat−like superfamily protein	60	1887	113220	34444	30.42	1032	0.55	559	0.30	0.618	0.382	21307	11754	16577	29138	0	0	0	0	0	0	0	0	0	0	0	0	0	34444	0
DN12988_c0_g1_i1	AT5G50230	Transducin/WD40 repeat−like superfamily protein	62	2277	141174	30743	21.78	1048	0.46	617	0.27	0.587	0.413	31666	20403	24892	33470	0	0	0	0	0	0	0	0	0	0	0	0	0	30743	0
DN12999_c0_g1_i1	AT2G39550	Prenyltransferase family protein	59	3072	181248	54541	30.09	2113	0.69	1249	0.41	0.641	0.359	33045	21809	23780	48073	0	0	0	0	0	0	0	0	0	0	0	0	0	54541	0
DN13037_c0_g1_i1	AT5G15550	Transducin/WD40 repeat−like superfamily protein	56	3808	213248	65179	30.57	2119	0.56	1164	0.31	0.613	0.387	38797	25283	30989	53000	0	0	0	0	0	0	0	0	0	0	0	0	0	65179	0
DN13084_c0_g1_i2	AT1G02020	nitroreductase family protein	79	2231	176249	31071	17.63	1302	0.58	746	0.33	0.535	0.465	35906	31853	35266	42153	0	0	0	0	0	0	0	0	0	0	0	0	0	31071	0
DN13155_c0_g1_i1	AT1G09280	UNKNOWN	61	2843	173423	51238	29.55	1665	0.59	860	0.30	0.609	0.391	34289	22644	24417	40835	0	0	0	0	0	0	0	0	0	0	0	0	0	51238	0
DN13237_c0_g1_i2	AT1G18460	alpha/beta−Hydrolases superfamily protein	80	4181	334480	89968	26.90	2640	0.63	1651	0.40	0.601	0.399	68081	46296	49385	80750	0	0	0	0	0	0	0	0	0	0	0	0	0	89968	0
DN13265_c0_g3_i3	AT1G11780	oxidoreductase, 2OG−Fe(II) oxygenase family protein	53	2825	149725	41488	27.71	1576	0.56	834	0.30	0.633	0.367	30192	18068	21304	38673	0	0	0	0	0	0	0	0	0	0	0	0	0	41488	0
DN13289_c0_g1_i2	AT1G64680	UNKNOWN	51	1444	73644	21277	28.89	702	0.49	457	0.32	0.557	0.443	14513	9936	12569	15349	0	0	0	0	0	0	0	0	0	0	0	0	0	21277	0
DN13303_c0_g1_i1	AT2G40570	initiator tRNA phosphoribosyl transferase family protein	71	2102	149242	36103	24.19	1405	0.67	770	0.37	0.615	0.385	32410	20438	22673	37618	0	0	0	0	0	0	0	0	0	0	0	0	0	36103	0
DN13314_c0_g1_i1	AT4G09730	RH39	76	3701	281276	74315	26.42	2171	0.59	1110	0.30	0.586	0.414	58079	39117	45144	64621	0	0	0	0	0	0	0	0	0	0	0	0	0	74315	0
DN13371_c1_g1_i2	AT3G24040	Core−2/I−branching beta−1,6−N−acetylglucosaminyltransferase family protein	54	2150	116100	26106	22.49	1260	0.59	650	0.30	0.567	0.433	22430	20510	17959	29095	0	0	0	0	0	0	0	0	0	0	0	0	0	26106	0
DN13461_c0_g1_i1	AT5G35400	Pseudouridine synthase family protein	64	2833	181312	53842	29.70	1488	0.53	810	0.29	0.622	0.378	36781	21761	26025	42903	0	0	0	0	0	0	0	0	0	0	0	0	0	53842	0
DN13462_c0_g3_i1	AT5G56450	Mitochondrial substrate carrier family protein	66	1677	110682	25675	23.20	959	0.57	519	0.31	0.55	0.45	22759	16094	22069	24085	0	0	0	0	0	0	0	0	0	0	0	0	0	25675	0
DN13476_c0_g1_i1	AT2G23890	HAD−superfamily hydrolase, subfamily IG, 5'−nucleotidase	57	3861	220077	60647	27.56	2259	0.59	1188	0.31	0.638	0.362	46238	26923	30431	55838	0	0	0	0	0	0	0	0	0	0	0	0	0	60647	0
DN13481_c0_g1_i3	AT5G56900	CwfJ−like family protein / zinc finger (CCCH−type) family protein	70	3718	260260	68235	26.22	2068	0.56	1204	0.32	0.625	0.375	53136	33452	37813	67624	0	0	0	0	0	0	0	0	0	0	0	0	0	68235	0
DN13497_c0_g1_i3	AT4G20930	6−phosphogluconate dehydrogenase family protein	55	2783	153065	42716	27.91	1413	0.51	853	0.31	0.626	0.374	29745	19296	21858	39450	0	0	0	0	0	0	0	0	0	0	0	0	0	42716	0
DN13513_c0_g1_i5	AT5G57140	purple acid phosphatase 28	71	2297	163087	42403	26.00	1338	0.58	755	0.33	0.605	0.395	33970	20843	26264	39607	0	0	0	0	0	0	0	0	0	0	0	0	0	42403	0
DN13596_c0_g1_i2	AT2G26810	Putative methyltransferase family protein	52	1604	83408	23467	28.14	1063	0.66	585	0.37	0.643	0.357	17311	10310	11052	21268	0	0	0	0	0	0	0	0	0	0	0	0	0	23467	0
DN13631_c0_g1_i1	AT1G43860	sequence−specific DNA binding transcription factors	58	2375	137750	31678	23.00	1370	0.58	776	0.33	0.623	0.377	31416	17582	22395	34679	0	0	0	0	0	0	0	0	0	0	0	0	0	31678	0
DN13880_c0_g1_i6	AT3G13170	Spo11/DNA topoisomerase VI, subunit A protein	57	3337	190209	57916	30.45	1925	0.58	1067	0.32	0.642	0.358	38298	22669	24530	46796	0	0	0	0	0	0	0	0	0	0	0	0	0	57916	0
DN13905_c0_g1_i1	AT5G25150	TBP−associated factor 5	79	5047	398713	139736	35.05	3360	0.67	1948	0.39	0.624	0.376	68943	45944	49541	94549	0	0	0	0	0	0	0	0	0	0	0	0	0	139736	0
DN13915_c0_g1_i1	AT5G54860	Major facilitator superfamily protein	72	3504	252288	68013	26.96	1964	0.56	1079	0.31	0.626	0.374	52606	33732	34766	63171	0	0	0	0	0	0	0	0	0	0	0	0	0	68013	0

DN13916_c0_g2_i1	AT3G45770	Polyketide synthase, enoylreductase family protein	72	3241	233352	53869	23.09	2008	0.62	1211	0.37	0.609	0.391	47076	33882	36142	62383	0	0	0	0	0	0	0	0	0	0	0	0	0	53869	0
DN13925_c0_g1_i2	AT3G53140	O-methyltransferase family protein	63	1682	105966	27605	26.05	989	0.59	661	0.39	0.557	0.443	22583	15958	18456	21364	0	0	0	0	0	0	0	0	0	0	0	0	0	27605	0
DN13973_c0_g4_i1	AT2G40390	UNKNOWN	69	1802	124338	29427	23.67	1066	0.59	682	0.38	0.526	0.474	25601	23057	21805	24448	0	0	0	0	0	0	0	0	0	0	0	0	0	29427	0
DN14001_c0_g1_i1	AT1G74240	Mitochondrial substrate carrier family protein	58	2372	137576	38106	27.70	1131	0.48	695	0.29	0.618	0.382	27514	16756	21043	34157	0	0	0	0	0	0	0	0	0	0	0	0	0	38106	0
DN14081_c0_g1_i1	AT5G67140	F-box/RNI-like superfamily protein	60	1877	112620	31781	28.22	1289	0.69	748	0.40	0.621	0.379	22490	14874	15526	27949	0	0	0	0	0	0	0	0	0	0	0	0	0	31781	0
DN14129_c0_g1_i1	AT3G15180	ARM repeat superfamily protein	58	4385	254330	89806	35.31	2669	0.61	1401	0.32	0.634	0.366	42873	28806	31692	61153	0	0	0	0	0	0	0	0	0	0	0	0	0	89806	0
DN14140_c0_g1_i1	AT4G29830	Transducin/WD40 repeat-like superfamily protein	56	1606	89936	20285	22.56	695	0.43	408	0.25	0.537	0.463	17191	15314	16537	20609	0	0	0	0	0	0	0	0	0	0	0	0	0	20285	0
DN14166_c0_g1_i1	AT1G12470	zinc ion binding	61	6791	414251	131355	31.71	3421	0.50	1774	0.26	0.639	0.361	79202	47581	53897	102216	0	0	0	0	0	0	0	0	0	0	0	0	0	131355	0
DN14195_c0_g1_i2	AT4G01310	Ribosomal L5P family protein	75	1961	147075	40246	27.36	1411	0.72	763	0.39	0.592	0.408	31278	21173	22188	32190	0	0	0	0	0	0	0	0	0	0	0	0	0	40246	0
DN14300_c0_g1_i1	AT3G10700	galacturonic acid kinase	54	3654	197316	65054	32.97	2474	0.68	1189	0.33	0.639	0.361	38754	23034	24146	46328	0	0	0	0	0	0	0	0	0	0	0	0	0	65054	0
DN14329_c0_g1_i1	AT1G57770	FAD/NAD(P)-binding oxidoreductase family protein	59	2797	165023	34098	20.66	1460	0.52	868	0.31	0.585	0.415	35725	25390	28497	41313	0	0	0	0	0	0	0	0	0	0	0	0	0	34098	0
DN14430_c0_g1_i4	AT1G21480	Exostosin family protein	78	3418	266604	73391	27.53	2657	0.78	1517	0.44	0.61	0.39	54813	35474	38296	64630	0	0	0	0	0	0	0	0	0	0	0	0	0	73391	0
DN14446_c0_g1_i1	AT4G36470	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein	56	1468	82208	21729	26.43	864	0.59	573	0.39	0.606	0.394	18939	10794	13097	17649	0	0	0	0	0	0	0	0	0	0	0	0	0	21729	0
DN14455_c0_g1_i1	AT3G55480	protein affected TRAFficKING	80	6085	486800	115726	23.77	3963	0.65	2204	0.36	0.632	0.368	104316	64571	70926	131261	0	0	0	0	0	0	0	0	0	0	0	0	0	115726	0
DN14467_c0_g1_i1	AT1G70330	equilibrative nucleotide transporter 1	70	2047	143290	25269	17.64	1112	0.54	692	0.34	0.575	0.425	29646	22118	27850	38407	0	0	0	0	0	0	0	0	0	0	0	0	0	25269	0
DN14484_c0_g1_i1	AT1G48520	GLU-ADT subunit B	68	3452	234736	49544	21.11	2167	0.63	1110	0.32	0.622	0.378	53927	31955	37717	61593	0	0	0	0	0	0	0	0	0	0	0	0	0	49544	0
DN14540_c0_g1_i3	AT3G24430	ATP binding	64	4154	265856	86908	32.69	1779	0.43	949	0.23	0.615	0.385	49008	32818	36112	61010	0	0	0	0	0	0	0	0	0	0	0	0	0	86908	0
DN14633_c1_g1_i5	AT1G04900	Protein of unknown function (DUF185)	61	3047	185867	58409	31.43	1682	0.55	1003	0.33	0.634	0.366	36634	23108	23380	44336	0	0	0	0	0	0	0	0	0	0	0	0	0	58409	0
DN14642_c0_g1_i1	AT2G39930	isoamylase 1	61	5875	358375	123950	34.59	2989	0.51	1624	0.28	0.631	0.369	64072	40915	45496	83942	0	0	0	0	0	0	0	0	0	0	0	0	0	123950	0
DN14660_c0_g2_i1	AT1G74640	alpha/beta-Hydrolases superfamily protein	64	1918	122752	26195	21.34	1012	0.53	517	0.27	0.58	0.42	26401	19165	21023	29968	0	0	0	0	0	0	0	0	0	0	0	0	0	26195	0
DN14670_c0_g1_i3	AT5G13400	Major facilitator superfamily protein	77	2478	190806	22912	12.01	1170	0.47	653	0.26	0.556	0.444	43305	36321	38079	50189	0	0	0	0	0	0	0	0	0	0	0	0	0	22912	0
DN14715_c0_g1_i3	AT2G47830	Cation efflux family protein	73	4040	294920	75987	25.77	2632	0.65	1569	0.39	0.624	0.376	60858	40215	41591	76269	0	0	0	0	0	0	0	0	0	0	0	0	0	75987	0
DN14783_c0_g1_i1	AT2G14530	TRICHOME BIREFRINGENCE-LIKE 13	56	2528	141568	39659	28.01	1429	0.57	807	0.32	0.601	0.399	27357	19170	20898	34484	0	0	0	0	0	0	0	0	0	0	0	0	0	39659	0
DN14847_c0_g2_i1	AT1G01910	P-loop containing nucleoside triphosphate hydrolases superfamily protein	51	1938	98838	31998	32.37	1112	0.57	697	0.36	0.619	0.381	18055	11691	13614	23480	0	0	0	0	0	0	0	0	0	0	0	0	0	31998	0
DN14856_c0_g1_i1	AT1G64550	general control non-repressible 3	68	6777	460836	143710	31.19	4042	0.60	2160	0.32	0.625	0.375	87479	54385	62947	112315	0	0	0	0	0	0	0	0	0	0	0	0	0	143710	0
DN14903_c0_g1_i2	AT3G18165	modifier of snc1,4	66	2220	146520	39693	27.09	1226	0.55	781	0.35	0.595	0.405	29000	20687	22156	34984	0	0	0	0	0	0	0	0	0	0	0	0	0	39693	0
DN14907_c0_g1_i1	AT5G58760	damaged DNA binding 2	63	3242	204246	56487	27.66	1719	0.53	964	0.30	0.616	0.384	42496	25064	31549	48650	0	0	0	0	0	0	0	0	0	0	0	0	0	56487	0
DN14916_c0_g1_i2	AT1G48430	Dihydroxyacetone kinase	57	5399	307743	97781	31.77	3180	0.59	1561	0.29	0.611	0.389	55452	37244	43351	73915	0	0	0	0	0	0	0	0	0	0	0	0	0	97781	0
DN14928_c0_g1_i2	AT3G01920	DHBP synthase RibB-like alpha/beta domain	67	2314	155038	42752	27.58	1544	0.67	913	0.40	0.613	0.387	29820	21234	21861	39371	0	0	0	0	0	0	0	0	0	0	0	0	0	42752	0
DN14959_c0_g1_i1	AT1G31780	UNKNOWN	64	4030	257920	70986	27.52	1971	0.49	1035	0.26	0.6	0.4	50928	37417	36733	61856	0	0	0	0	0	0	0	0	0	0	0	0	0	70986	0
DN14976_c0_g1_i1	AT5G61970	signal recognition particle-related / SRP-related	67	5195	348065	104812	30.11	2981	0.57	1618	0.31	0.621	0.379	68178	41442	49734	83899	0	0	0	0	0	0	0	0	0	0	0	0	0	104812	0
DN15038_c0_g2_i2	AT4G34215	Domain of unknown function (DUF303)	33	1321	43593	11004	25.24	761	0.58	500	0.38	0.538	0.462	9300	6298	8480	8511	0	0	0	0	0	0	0	0	0	0	0	0	0	11004	0
DN15095_c0_g1_i1	AT5G27660	Trypsin family protein with PDZ domain	59	3140	185260	62674	33.83	1756	0.56	1011	0.32	0.618	0.382	32661	21569	24692	43664	0	0	0	0	0	0	0	0	0	0	0	0	0	62674	0
DN15130_c0_g1_i1	AT3G48470	embryo defective 2423	67	5588	374396	99578	26.60	3198	0.57	1555	0.28	0.611	0.389	78493	49825	55868	90632	0	0	0	0	0	0	0	0	0	0	0	0	0	99578	0
DN15144_c0_g1_i1	AT3G33520	actin-related protein 6	70	2203	154210	33163	21.51	1183	0.54	699	0.32	0.593	0.407	31675	23712	25523	40137	0	0	0	0	0	0	0	0	0	0	0	0	0	33163	0
DN15150_c0_g1_i1	AT3G47450	P-loop containing nucleoside triphosphate hydrolases superfamily protein	66	3526	232716	60081	25.82	2330	0.66	1307	0.37	0.592	0.408	46373	33341	36475	56446	0	0	0	0	0	0	0	0	0	0	0	0	0	60081	0
DN15203_c0_g2_i5	AT2G17975	zinc finger (Ran-binding) family protein	68	3302	224536	55319	24.64	2130	0.65	1220	0.37	0.601	0.399	45809	30832	35708	56868	0	0	0	0	0	0	0	0	0	0	0	0	0	55319	0
DN15204_c0_g1_i1	AT3G59630	diphthamide synthesis DPH2 family protein	68	3733	253844	79014	31.13	2330	0.62	1259	0.34	0.62	0.38	49220	30470	35541	59599	0	0	0	0	0	0	0	0	0	0	0	0	0	79014	0

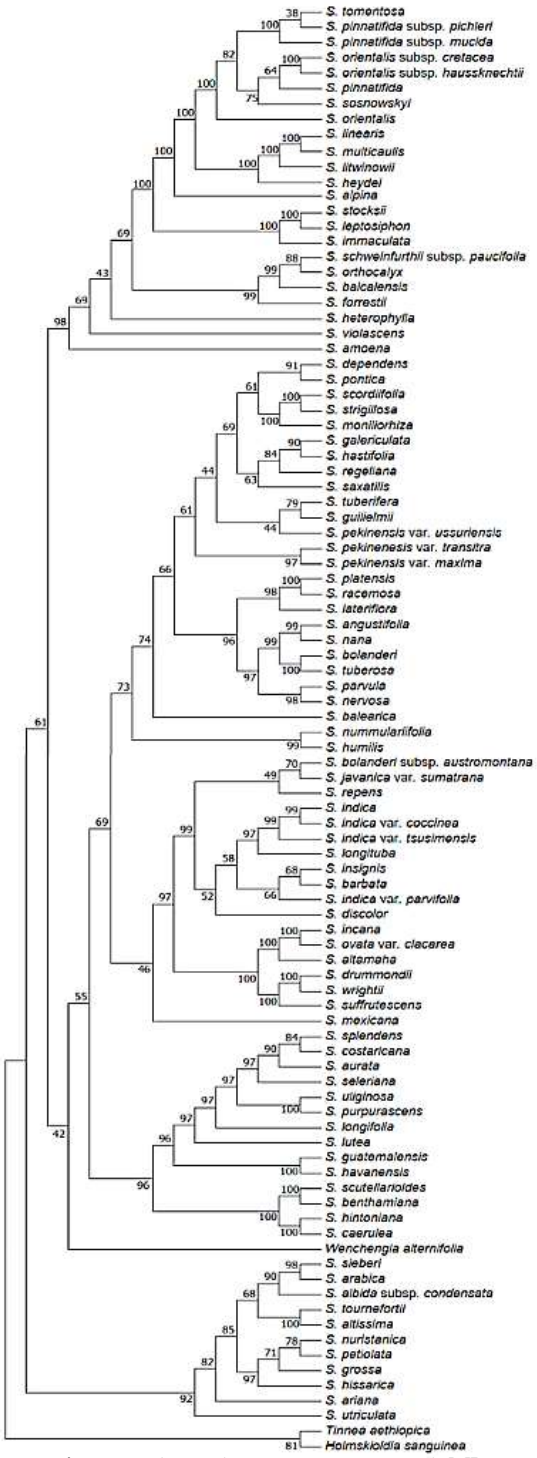
DN15237_c0_g2_i2	AT5G55000	potassium channel tetramerisation domain-containing protein	60	2510	150600	45165	29.99	1435	0.57	802	0.32	0.637	0.363	28239	18487	19610	39099	0	0	0	0	0	0	0	0	0	0	0	0	0	45165	0
DN15269_c0_g2_i1	AT3G51050	/pentapeptiderepeat-containingprotein FG-GAP repeat-containing protein	78	5522	430716	132695	30.81	3379	0.61	1861	0.34	0.605	0.395	82299	54844	60157	100721	0	0	0	0	0	0	0	0	0	0	0	0	0	132695	0
DN15314_c0_g1_i1	AT4G35850	Pentatricopeptide repeat (PPR) superfamily protein	69	3725	257025	77885	30.30	2083	0.56	1177	0.32	0.634	0.366	49324	30300	35094	64422	0	0	0	0	0	0	0	0	0	0	0	0	0	77885	0
DN15330_c0_g1_i2	AT5G60540	pyridoxine biosynthesis 2	54	2046	110484	33803	30.60	1040	0.51	606	0.30	0.594	0.406	19928	15269	15748	25736	0	0	0	0	0	0	0	0	0	0	0	0	0	33803	0
DN15349_c0_g1_i5	AT5G39830	Trypsin family protein with PDZ domain	53	2466	130698	29585	22.64	1260	0.51	718	0.29	0.607	0.393	25470	20578	18987	36078	0	0	0	0	0	0	0	0	0	0	0	0	0	29585	0
DN15425_c0_g1_i1	AT5G46580	pentatricopeptide (PPR) repeat-containing protein	73	2500	182500	16869	9.24	1184	0.47	611	0.24	0.551	0.449	44662	32376	42005	46588	0	0	0	0	0	0	0	0	0	0	0	0	0	16869	0
DN15467_c1_g1_i2	AT1G67550	urease	73	4593	335289	100170	29.88	2462	0.54	1381	0.30	0.632	0.368	69274	38948	47342	79555	0	0	0	0	0	0	0	0	0	0	0	0	0	100170	0
DN15482_c0_g1_i1	AT2G34640	plastid transcriptionally active 12	68	2855	194140	49230	25.36	1490	0.52	866	0.30	0.602	0.398	43762	25553	31492	44103	0	0	0	0	0	0	0	0	0	0	0	0	0	49230	0
DN15511_c0_g3_i2	AT2G35130	Tetratricopeptide repeat (TPR)-like superfamily protein	56	2594	145264	32391	22.30	1676	0.65	812	0.31	0.602	0.398	33691	20425	23931	34826	0	0	0	0	0	0	0	0	0	0	0	0	0	32391	0
DN15511_c0_g3_i4	AT2G35130	Tetratricopeptide repeat (TPR)-like superfamily protein	74	2943	217782	43132	19.81	1600	0.54	986	0.34	0.61	0.39	52531	31203	36703	54213	0	0	0	0	0	0	0	0	0	0	0	0	0	43132	0
DN15516_c0_g1_i1	AT5G40440	mitogen-activated protein kinase kinase 3	67	3473	232691	53789	23.12	2084	0.60	988	0.28	0.627	0.373	51136	31343	34946	61477	0	0	0	0	0	0	0	0	0	0	0	0	0	53789	0
DN15524_c0_g1_i4	AT4G18465	RNA helicase family protein	75	6893	516975	182798	35.36	4531	0.66	2728	0.40	0.644	0.356	92405	55412	62889	123471	0	0	0	0	0	0	0	0	0	0	0	0	0	182798	0
DN15601_c0_g2_i2	AT3G12080	GTP-binding family protein	66	4680	308880	94266	30.52	2564	0.55	1440	0.31	0.622	0.378	63153	36819	44172	70470	0	0	0	0	0	0	0	0	0	0	0	0	0	94266	0
DN15732_c0_g1_i1	AT5G45760	Transducin/WD40 repeat-like superfamily protein	55	2826	155430	49663	31.95	1623	0.57	925	0.33	0.627	0.373	26975	18351	21082	39359	0	0	0	0	0	0	0	0	0	0	0	0	0	49663	0
DN15736_c0_g1_i1	AT3G19630	Radical SAM superfamily protein	58	2149	124642	25392	20.37	1456	0.68	804	0.37	0.612	0.388	27481	19646	18874	33249	0	0	0	0	0	0	0	0	0	0	0	0	0	25392	0
DN15755_c0_g2_i1	AT2G38020	vacuoleless1 (VCL1)	63	4220	265860	80507	30.28	2346	0.56	1378	0.33	0.591	0.409	46817	36543	38751	63242	0	0	0	0	0	0	0	0	0	0	0	0	0	80507	0
DN15758_c0_g1_i4	AT1G80770	P-loop containing nucleoside triphosphate hydrolases superfamily protein	73	4749	346677	113537	32.75	2904	0.61	1599	0.34	0.628	0.372	66637	40496	45881	80126	0	0	0	0	0	0	0	0	0	0	0	0	0	113537	0
DN15760_c0_g1_i1	AT1G67420	Zn-dependent exopeptidases superfamily protein	75	7087	531525	166917	31.40	3865	0.55	2163	0.31	0.634	0.366	96807	64053	69160	134588	0	0	0	0	0	0	0	0	0	0	0	0	0	166917	0
DN15853_c0_g4_i2	AT5G66810	UNKNOWN	61	4063	247843	75061	30.29	2268	0.56	1122	0.28	0.604	0.396	47480	32081	36212	57009	0	0	0	0	0	0	0	0	0	0	0	0	0	75061	0
DN15890_c0_g1_i1	AT1G73110	P-loop containing nucleoside triphosphate hydrolases superfamily protein	77	3734	287518	79925	27.80	2309	0.62	1361	0.36	0.633	0.367	62585	35017	40397	69594	0	0	0	0	0	0	0	0	0	0	0	0	0	79925	0
DN15904_c0_g2_i1	AT2G45500	AAA-type ATPase family protein	72	3321	239112	65777	27.51	1851	0.56	1004	0.30	0.613	0.387	48034	30880	35279	59142	0	0	0	0	0	0	0	0	0	0	0	0	0	65777	0
DN15969_c0_g1_i5	AT1G78800	UDP-Glycosyltransferase superfamily protein	65	2504	162760	34853	21.41	1302	0.52	756	0.30	0.635	0.365	35737	23538	22891	45741	0	0	0	0	0	0	0	0	0	0	0	0	0	34853	0
DN15986_c0_g1_i3	AT2G23390	UNKNOWN	77	3828	294756	80780	27.41	2741	0.72	1485	0.39	0.631	0.369	61972	37819	40561	73624	0	0	0	0	0	0	0	0	0	0	0	0	0	80780	0
DN16048_c0_g1_i3	AT1G69830	alpha-amylase-like 3	79	4940	390260	102097	26.16	2841	0.58	1628	0.33	0.619	0.381	83640	49711	59035	95777	0	0	0	0	0	0	0	0	0	0	0	0	0	102097	0
DN16070_c1_g1_i1	AT5G23340	RNI-like superfamily protein	72	1823	131256	23963	18.26	843	0.46	473	0.26	0.572	0.428	28104	20953	25183	33053	0	0	0	0	0	0	0	0	0	0	0	0	0	23963	0
DN16093_c0_g1_i2	AT5G10730	NAD(P)-binding Rossmann-fold superfamily protein	65	3013	195845	55384	28.28	1462	0.49	871	0.29	0.618	0.382	34768	25489	27751	52453	0	0	0	0	0	0	0	0	0	0	0	0	0	55384	0
DN16137_c0_g1_i1	AT3G16840	P-loop containing nucleoside triphosphate hydrolases superfamily protein	64	4134	264576	77794	29.40	2707	0.66	1357	0.33	0.613	0.387	54939	31008	41241	59594	0	0	0	0	0	0	0	0	0	0	0	0	0	77794	0
DN16171_c0_g1_i4	AT1G76130	alpha-amylase-like 2	58	2959	171622	45993	26.80	1457	0.49	859	0.29	0.632	0.368	36033	21428	24493	43675	0	0	0	0	0	0	0	0	0	0	0	0	0	45993	0
DN16176_c0_g3_i4	AT4G10050	esterase/lipase/thioesterase family protein	73	3609	263457	76321	28.97	2170	0.60	1321	0.37	0.632	0.368	50743	32600	36087	67706	0	0	0	0	0	0	0	0	0	0	0	0	0	76321	0
DN16185_c0_g1_i1	AT3G07140	GPI transamidase component Gpi16 subunit family protein	71	3489	247719	51818	20.92	1864	0.53	1047	0.30	0.604	0.396	55091	37449	39837	63524	0	0	0	0	0	0	0	0	0	0	0	0	0	51818	0
DN16194_c1_g3_i3	AT3G19895	RING/U-box superfamily protein	72	3737	269064	87445	32.50	2238	0.60	1190	0.32	0.606	0.394	46777	36355	34979	63508	0	0	0	0	0	0	0	0	0	0	0	0	0	87445	0
DN16233_c0_g1_i2	AT4G11120	translation elongation factor Ts (EF-Ts), putative	65	2637	171405	38456	22.44	1493	0.57	882	0.33	0.626	0.374	37576	22720	26582	46071	0	0	0	0	0	0	0	0	0	0	0	0	0	38456	0

DN16275_c0_g1_i1	AT1G77030	hydrolases, acting on acid anhydrides, in phosphorus−containing anhydrides/ ATP−dependenthelicases/nucleicacidbinding /ATPbinding/RNAbinding/helicases	66	4611	304326	85607	28.13	2245	0.49	1227	0.27	0.604	0.396	64746	39435	46804	67734	0	0	0	0	0	0	0	0	0	0	0	0	0	85607	0
DN16369_c0_g1_i2	AT2G40090	ABC2 homolog 9	69	4004	276276	72467	26.23	2538	0.63	1362	0.34	0.633	0.367	56742	35160	39524	72383	0	0	0	0	0	0	0	0	0	0	0	0	0	72467	0
DN16384_c0_g1_i4	AT3G19210	homolog of RAD54	64	5945	380480	119274	31.35	3110	0.52	1645	0.28	0.62	0.38	74416	47078	51715	87997	0	0	0	0	0	0	0	0	0	0	0	0	0	119274	0
DN16386_c1_g1_i1	AT3G56040	UDP−glucose pyrophosphorylase 3	83	6533	542239	155089	28.60	4091	0.63	2246	0.34	0.628	0.372	107749	64441	77799	137161	0	0	0	0	0	0	0	0	0	0	0	0	0	155089	0
DN16415_c0_g1_i1	AT1G01770	UNKNOWN	71	4147	294437	77663	26.38	2576	0.62	1344	0.32	0.609	0.391	60203	37980	46051	72540	0	0	0	0	0	0	0	0	0	0	0	0	0	77663	0
DN16418_c0_g1_i5	AT4G01040	Glycosyl hydrolase superfamily protein	58	3351	194358	54273	27.92	2049	0.61	1045	0.31	0.613	0.387	38055	25044	28617	48369	0	0	0	0	0	0	0	0	0	0	0	0	0	54273	0
DN16443_c0_g1_i7	AT1G15240	Phox−associated domainPhox−likeSorting nexin, C−terminal	81	5529	447849	99038	22.11	3124	0.57	1733	0.31	0.605	0.395	101869	61931	72778	112233	0	0	0	0	0	0	0	0	0	0	0	0	0	99038	0
DN16453_c0_g1_i1	AT5G27560	Domain of unknown function (DUF1995)	52	3183	165516	48261	29.16	1997	0.63	1036	0.33	0.615	0.385	32038	21729	22833	40655	0	0	0	0	0	0	0	0	0	0	0	0	0	48261	0
DN16461_c0_g1_i1	AT3G22290	Endoplasmic reticulum vesicle transporter protein	70	3619	253330	79440	31.36	2248	0.62	1305	0.36	0.647	0.353	46553	28972	32394	65971	0	0	0	0	0	0	0	0	0	0	0	0	0	79440	0
DN16529_c0_g1_i2	AT3G02710	ARM repeat superfamily protein	77	2831	217987	47523	21.80	1560	0.55	890	0.31	0.577	0.423	44907	34182	36746	54629	0	0	0	0	0	0	0	0	0	0	0	0	0	47523	0
DN16541_c1_g3_i1	AT2G30800	helicase in vascular tissue and tapetum	75	6065	454875	115654	25.43	3308	0.55	1817	0.30	0.607	0.393	97772	60486	72495	108468	0	0	0	0	0	0	0	0	0	0	0	0	0	115654	0
DN16606_c0_g1_i1	AT5G01400	HEAT repeat−containing protein	72	8292	597024	162557	27.23	4544	0.55	2384	0.29	0.618	0.382	119473	79177	85564	150253	0	0	0	0	0	0	0	0	0	0	0	0	0	162557	0
DN16611_c2_g12_i3	AT3G54380	SAC3/GANP/Nin1/mts3/eIF−3 p25 family	65	2741	178165	53262	29.90	1501	0.55	898	0.33	0.624	0.376	35108	23365	23470	42960	0	0	0	0	0	0	0	0	0	0	0	0	0	53262	0
DN16640_c0_g2_i1	AT3G44600	cyclophilin71	72	4209	303048	79567	26.26	2662	0.63	1321	0.31	0.631	0.369	63314	39929	42259	77979	0	0	0	0	0	0	0	0	0	0	0	0	0	79567	0
DN16698_c0_g1_i1	AT5G45170	Haloacid dehalogenase−like hydrolase (HAD) superfamily protein	69	3876	267444	85256	31.88	2560	0.66	1491	0.39	0.634	0.366	51743	28539	37224	64682	0	0	0	0	0	0	0	0	0	0	0	0	0	85256	0
DN16725_c0_g3_i2	AT3G27925	DegP protease 1	55	3080	169400	49333	29.12	1480	0.48	760	0.25	0.602	0.398	31168	22831	24250	41818	0	0	0	0	0	0	0	0	0	0	0	0	0	49333	0
DN16728_c3_g1_i5	AT4G27640	ARM repeat superfamily protein	62	5561	344782	94854	27.51	3150	0.57	1400	0.25	0.597	0.403	64981	49662	50300	84985	0	0	0	0	0	0	0	0	0	0	0	0	0	94854	0
DN16761_c0_g1_i3	AT3G04480	endoribonucleases	71	4547	322837	85045	26.34	2227	0.49	1197	0.26	0.623	0.377	65804	43681	45602	82705	0	0	0	0	0	0	0	0	0	0	0	0	0	85045	0
DN16801_c0_g1_i1	AT1G04620	coenzyme F420 hydrogenase family / dehydrogenase, beta subunit family	77	5135	395395	125145	31.65	3443	0.67	1947	0.38	0.644	0.356	76090	45520	50643	97997	0	0	0	0	0	0	0	0	0	0	0	0	0	125145	0
DN16815_c0_g1_i2	AT2G31060	elongation factor family protein	77	6726	517902	174227	33.64	4345	0.65	2480	0.37	0.622	0.378	89925	60529	67722	125499	0	0	0	0	0	0	0	0	0	0	0	0	0	174227	0
DN16839_c0_g1_i4	AT3G02300	Regulator of chromosome condensation (RCC1) family protein	58	3011	174638	51007	29.21	1483	0.49	785	0.26	0.586	0.414	32439	21581	28858	40753	0	0	0	0	0	0	0	0	0	0	0	0	0	51007	0
DN16894_c2_g1_i2	AT3G25900	Homocysteine S−methyltransferase family protein	52	1779	92508	24020	25.97	912	0.51	560	0.32	0.585	0.415	19490	12966	15358	20674	0	0	0	0	0	0	0	0	0	0	0	0	0	24020	0
DN16910_c0_g1_i3	AT3G14890	phosphoesterase	52	3182	165464	51765	31.29	1852	0.58	1008	0.32	0.631	0.369	32722	18603	22454	39920	0	0	0	0	0	0	0	0	0	0	0	0	0	51765	0
DN16947_c0_g3_i2	AT5G47090	UNKNOWN	53	1510	80030	16860	21.07	704	0.47	404	0.27	0.608	0.392	19471	8752	15700	19247	0	0	0	0	0	0	0	0	0	0	0	0	0	16860	0
DN16959_c0_g1_i3	AT1G07230	non−specific phospholipase C1	57	2552	145464	32102	22.07	1320	0.52	753	0.30	0.544	0.456	28945	26883	24706	32828	0	0	0	0	0	0	0	0	0	0	0	0	0	32102	0
DN16981_c1_g6_i1	AT2G42810	protein phosphatase 5.2	63	3208	202104	71951	35.60	2262	0.71	1277	0.40	0.643	0.357	36861	20861	25473	46958	0	0	0	0	0	0	0	0	0	0	0	0	0	71951	0
DN16986_c1_g6_i2	AT3G04870	zeta−carotene desaturase	62	4359	270258	87051	32.21	2389	0.55	1214	0.28	0.607	0.393	48765	33079	38568	62795	0	0	0	0	0	0	0	0	0	0	0	0	0	87051	0
DN17032_c0_g2_i2	AT5G14220	Flavin containing amine oxidoreductase family	60	4525	271500	95044	35.01	2314	0.51	1202	0.27	0.633	0.367	47422	30485	34111	64438	0	0	0	0	0	0	0	0	0	0	0	0	0	95044	0
DN17071_c0_g1_i15	AT3G19720	P−loop containing nucleoside triphosphate hydrolases superfamily protein	55	5446	299530	98680	32.95	3050	0.56	1602	0.29	0.615	0.385	54593	37808	38851	69598	0	0	0	0	0	0	0	0	0	0	0	0	0	98680	0
DN17071_c0_g1_i16	AT3G19720	P−loop containing nucleoside triphosphate hydrolases superfamily protein	74	5866	434084	138815	31.98	3775	0.64	1995	0.34	0.625	0.375	82638	54562	56545	101524	0	0	0	0	0	0	0	0	0	0	0	0	0	138815	0
DN17071_c0_g8_i2	AT5G50310	Galactose oxidase/kelch repeat superfamily protein	75	5599	419925	123018	29.30	3249	0.58	1897	0.34	0.632	0.368	86857	46835	61311	101904	0	0	0	0	0	0	0	0	0	0	0	0	0	123018	0
DN17092_c0_g1_i4	AT3G02060	DEAD/DEAH box helicase, putative	61	4443	271023	82650	30.50	2199	0.50	1151	0.26	0.635	0.365	52998	32445	35927	67003	0	0	0	0	0	0	0	0	0	0	0	0	0	82650	0
DN17093_c3_g1_i4	AT2G19430	DWD (DDB1−binding WD40 protein) hypersensitive to ABA 1	69	3347	230943	80486	34.85	1940	0.58	1018	0.30	0.638	0.362	39517	25687	28827	56426	0	0	0	0	0	0	0	0	0	0	0	0	0	80486	0
DN17124_c0_g1_i3	AT1G08460	histone deacetylase 8	71	2003	142213	32102	22.57	1205	0.60	768	0.38	0.587	0.413	30673	20828	24329	34281	0	0	0	0	0	0	0	0	0	0	0	0	0	32102	0
DN17136_c2_g2_i1	AT1G62780	UNKNOWN	54	2202	118908	29740	25.01	1157	0.53	714	0.32	0.63	0.37	26198	15363	17552	30055	0	0	0	0	0	0	0	0	0	0	0	0	0	29740	0

DN17160_c0_g1_i2	AT2G39670	Radical SAM superfamily protein	53	3311	175483	49382	28.14	1551	0.47	899	0.27	0.608	0.392	34941	23117	25691	42352	0	0	0	0	0	0	0	0	0	0	0	0	49382	0
DN17217_c0_g1_i1	AT3G52640	Zn-dependent exopeptidases superfamily protein	61	4978	303658	99378	32.73	2979	0.60	1643	0.33	0.63	0.37	56098	36271	38402	73509	0	0	0	0	0	0	0	0	0	0	0	0	99378	0
DN17355_c0_g1_i1	AT5G57960	GTP-binding protein, HflX	56	3899	218344	65510	30.00	1960	0.50	1098	0.28	0.611	0.389	42366	26503	32168	51797	0	0	0	0	0	0	0	0	0	0	0	0	65510	0
DN17366_c0_g12_i1	AT2G38270	CAX-interacting protein 2	68	1512	102816	21631	21.04	909	0.60	511	0.34	0.565	0.435	23094	17335	17765	22991	0	0	0	0	0	0	0	0	0	0	0	0	21631	0
DN17379_c3_g3_i6	AT4G13360	ATP-dependent caseinolytic (Clp) protease/crotonase family protein	58	1945	112810	27712	24.57	1089	0.56	641	0.33	0.602	0.398	21287	16047	17943	29821	0	0	0	0	0	0	0	0	0	0	0	0	27712	0
DN17379_c3_g4_i7	AT5G44450	methyltransferases	64	2572	164608	46986	28.54	1551	0.60	927	0.36	0.616	0.384	31677	19220	25366	41359	0	0	0	0	0	0	0	0	0	0	0	0	46986	0
DN17392_c3_g2_i4	AT4G20130	plastid transcriptionally active 14	66	4319	285054	75783	26.59	2628	0.61	1262	0.29	0.633	0.367	60741	37116	39224	72190	0	0	0	0	0	0	0	0	0	0	0	0	75783	0
DN17394_c1_g1_i1	AT1G03000	peroxin 6	75	6186	463950	132319	28.52	2925	0.47	1789	0.29	0.619	0.381	92511	60586	64310	114224	0	0	0	0	0	0	0	0	0	0	0	0	132319	0
DN17464_c0_g2_i1	AT1G71810	Protein kinase superfamily protein	47	3823	179681	55283	30.77	2003	0.52	1106	0.29	0.63	0.37	35792	22164	23296	43146	0	0	0	0	0	0	0	0	0	0	0	0	55283	0
DN17465_c0_g2_i3	AT3G56740	Ubiquitin-associated (UBA) protein	57	2585	147345	42715	28.99	1916	0.74	957	0.37	0.64	0.36	27164	18496	18790	40180	0	0	0	0	0	0	0	0	0	0	0	0	42715	0
DN17500_c0_g1_i1	AT1G73180	Eukaryotic translation initiation factor eIF2A family protein	55	3711	204105	59437	29.12	2117	0.57	1050	0.28	0.621	0.379	39740	25369	28915	50644	0	0	0	0	0	0	0	0	0	0	0	0	59437	0
DN17587_c1_g2_i4	AT4G12610	transcription activators/DNA binding/RNA polymerase II transcriptionfactors/catalytics/transcriptioni nitiationfactors	60	3502	210120	60605	28.84	1873	0.54	1021	0.29	0.601	0.399	44623	25993	33286	45613	0	0	0	0	0	0	0	0	0	0	0	0	60605	0
DN17625_c0_g1_i2	AT1G14810	semialdehyde dehydrogenase family protein	66	2289	151074	36263	24.00	1573	0.69	805	0.35	0.564	0.436	30988	26033	23256	34534	0	0	0	0	0	0	0	0	0	0	0	0	36263	0
DN17639_c2_g7_i1	AT5G65490	UNKNOWN	73	3604	263092	67774	25.76	2401	0.67	1306	0.36	0.596	0.404	52606	34648	44409	63655	0	0	0	0	0	0	0	0	0	0	0	0	67774	0
DN17651_c0_g1_i6	AT5G27970	ARM repeat superfamily protein	75	10724	804300	256753	31.92	6089	0.57	3238	0.30	0.638	0.362	151759	95601	101052	199135	0	0	0	0	0	0	0	0	0	0	0	0	256753	0
DN17670_c0_g2_i7	AT5G16120	alpha/beta-Hydrolases superfamily protein	62	2535	157170	42877	27.28	1613	0.64	916	0.36	0.619	0.381	32879	19138	24176	38100	0	0	0	0	0	0	0	0	0	0	0	0	42877	0
DN17671_c0_g1_i4	AT5G48960	HAD-superfamily hydrolase, subfamily IG, 5'-nucleotidase	67	4721	316307	87854	27.78	2626	0.56	1335	0.28	0.627	0.373	64103	40020	44716	79614	0	0	0	0	0	0	0	0	0	0	0	0	87854	0
DN17681_c2_g1_i1	AT1G25375	Metallo-hydrolase/oxidoreductase superfamily protein	58	4523	262334	87100	33.20	2902	0.64	1495	0.33	0.637	0.363	49982	30653	32445	62154	0	0	0	0	0	0	0	0	0	0	0	0	87100	0
DN17681_c3_g5_i1	AT4G29720	polyamine oxidase 5	63	2559	161217	53694	33.31	1330	0.52	752	0.29	0.546	0.454	29404	21869	27062	29188	0	0	0	0	0	0	0	0	0	0	0	0	53694	0
DN17688_c0_g6_i1	AT4G15110	cytochrome P450, family 97, subfamily B, polypeptide 3	58	4767	276486	87381	31.60	2458	0.52	1372	0.29	0.638	0.362	55367	31353	36520	65865	0	0	0	0	0	0	0	0	0	0	0	0	87381	0
DN17712_c1_g5_i1	AT3G54460	SNF2 domain-containing protein / helicase domain-containing protein / F-box family protein	73	5449	397777	90194	22.68	2408	0.44	1272	0.23	0.59	0.41	87346	57528	68425	94284	0	0	0	0	0	0	0	0	0	0	0	0	90194	0
DN17736_c0_g1_i4	AT5G08170	porphyromonas-type peptidyl-arginine deiminase family protein	70	2899	202930	60007	29.57	2099	0.72	1303	0.45	0.629	0.371	39847	25043	27566	50467	0	0	0	0	0	0	0	0	0	0	0	0	60007	0
DN17740_c0_g2_i1	AT2G40760	Rhodanese/Cell cycle control phosphatase superfamily protein	71	2303	163513	38870	23.77	1466	0.64	826	0.36	0.582	0.418	34677	23288	28433	38245	0	0	0	0	0	0	0	0	0	0	0	0	38870	0
DN17742_c0_g1_i1	AT1G76400	Ribophorin I	62	3216	199392	50480	25.32	1466	0.46	849	0.26	0.615	0.385	40793	26562	30556	51001	0	0	0	0	0	0	0	0	0	0	0	0	50480	0
DN17772_c0_g2_i4	AT1G50500	Membrane trafficking VPS53 family protein	63	6032	380016	125499	33.03	2805	0.47	1499	0.25	0.622	0.378	70679	45165	50020	88653	0	0	0	0	0	0	0	0	0	0	0	0	125499	0
DN17790_c1_g1_i3	AT5G21060	Glyceraldehyde-3-phosphate dehydrogenase-like family protein	56	2326	130256	38007	29.18	1239	0.53	715	0.31	0.617	0.383	25790	16503	18707	31249	0	0	0	0	0	0	0	0	0	0	0	0	38007	0
DN17821_c0_g1_i1	AT3G07670	Rubisco methyltransferase family protein	57	3955	225435	67393	29.90	2335	0.59	1459	0.37	0.601	0.399	45441	29239	33107	50255	0	0	0	0	0	0	0	0	0	0	0	0	67393	0
DN17905_c0_g1_i2	AT2G13540	ARM repeat superfamily protein	67	4613	309071	99479	32.19	2854	0.62	1431	0.31	0.608	0.392	54901	36742	43192	74757	0	0	0	0	0	0	0	0	0	0	0	0	99479	0
DN17966_c0_g1_i6	AT4G05090	Inositol monophosphatase family protein	69	3459	238671	69252	29.02	2391	0.69	1217	0.35	0.609	0.391	44571	29308	36526	59014	0	0	0	0	0	0	0	0	0	0	0	0	69252	0
DN18034_c1_g1_i4	AT2G01460	P-loop containing nucleoside triphosphate hydrolases superfamily protein	67	5905	395635	106120	26.82	2911	0.49	1685	0.29	0.627	0.373	79492	52487	55085	102451	0	0	0	0	0	0	0	0	0	0	0	0	106120	0
DN18050_c3_g5_i1	AT1G08520	ALBINA 1	62	4549	282038	86590	30.70	2331	0.51	1213	0.27	0.605	0.395	56707	34772	42009	61960	0	0	0	0	0	0	0	0	0	0	0	0	86590	0
DN18089_c1_g1_i3	AT3G03310	lecithin/cholesterol acyltransferase 3	71	3677	261067	81865	31.36	2311	0.63	1331	0.36	0.638	0.362	51008	32199	32388	63607	0	0	0	0	0	0	0	0	0	0	0	0	81865	0

DN18105_c4_g5_i2	AT5G43430	electron transfer flavoprotein beta	64	1842	117888	25386	21.53	1160	0.63	714	0.39	0.598	0.402	26217	16863	20265	29157	0	0	0	0	0	0	0	0	0	0	0	0	25386	0
DN18120_c0_g1_i2	AT1G03090	methylcrotonyl−CoA carboxylase alpha chain, mitochondrial /3−methylcrotonyl−CoAcarboxylase1(MCCA)	74	5327	394198	126010	31.97	3003	0.56	1606	0.30	0.618	0.382	73087	48550	53349	93202	0	0	0	0	0	0	0	0	0	0	0	0	126010	0
DN18174_c0_g2_i1	AT1G30300	Metallo−hydrolase/oxidoreductase superfamily protein	75	2872	215400	69361	32.20	1843	0.64	979	0.34	0.63	0.37	39039	27017	26925	53058	0	0	0	0	0	0	0	0	0	0	0	0	69361	0
DN18206_c0_g2_i1	AT1G05940	cationic amino acid transporter 9	59	2966	174994	46953	26.83	1579	0.53	1017	0.34	0.595	0.405	31231	26285	24955	45570	0	0	0	0	0	0	0	0	0	0	0	0	46953	0
DN18243_c0_g1_i1	AT5G32450	RNA binding (RRM/RBD/RNP motifs) family protein	53	1789	94817	28062	29.60	854	0.48	463	0.26	0.647	0.353	20407	10170	13248	22930	0	0	0	0	0	0	0	0	0	0	0	0	28062	0
DN18263_c0_g2_i2	AT1G69070	UNKNOWN	73	5288	386024	98852	25.61	2996	0.57	1721	0.33	0.623	0.377	90188	46960	61085	88939	0	0	0	0	0	0	0	0	0	0	0	0	98852	0
DN18332_c0_g1_i2	AT5G11580	Regulator of chromosome condensation (RCC1) family protein	60	2535	152100	30523	20.07	1245	0.49	743	0.29	0.574	0.426	33626	21983	29766	36202	0	0	0	0	0	0	0	0	0	0	0	0	30523	0
DN18339_c0_g2_i2	NONE		77	2191	168707	24369	14.45	1363	0.62	766	0.35	0.593	0.407	39627	25057	33745	45909	0	0	0	0	0	0	0	0	0	0	0	0	24369	0
DN18356_c0_g3_i3	AT5G49970	pyridoxin (pyrodoxamine) 5'−phosphate oxidase	65	3594	233610	82997	35.53	2650	0.74	1572	0.44	0.615	0.385	41534	28524	28848	51707	0	0	0	0	0	0	0	0	0	0	0	0	82997	0
DN18369_c1_g3_i1	AT2G47760	asparagine−linked glycosylation 3	51	3440	175440	52275	29.80	2016	0.59	963	0.28	0.635	0.365	30689	22782	21995	47699	0	0	0	0	0	0	0	0	0	0	0	0	52275	0
DN18388_c0_g1_i1	AT2G01320	ABC−2 type transporter family protein	60	4387	263220	78668	29.89	1903	0.43	1046	0.24	0.609	0.391	50065	34878	36910	62699	0	0	0	0	0	0	0	0	0	0	0	0	78668	0
DN18502_c1_g1_i1	AT2G18876	Afadin/alpha−actinin−binding protein	75	2966	222450	56282	25.30	1949	0.66	1265	0.43	0.637	0.363	52964	26187	33686	53331	0	0	0	0	0	0	0	0	0	0	0	0	56282	0
DN18512_c0_g1_i2	AT3G10150	purple acid phosphatase 16	58	2050	118900	34828	29.29	1270	0.62	776	0.38	0.529	0.471	21489	16387	23265	22931	0	0	0	0	0	0	0	0	0	0	0	0	34828	0
DN18538_c2_g1_i2	AT5G09860	nuclear matrix protein−related	77	6235	480095	159519	33.23	4251	0.68	2426	0.39	0.641	0.359	90334	53766	59775	116701	0	0	0	0	0	0	0	0	0	0	0	0	159519	0
DN18568_c3_g4_i1	AT2G27590	S−adenosyl−L−methionine−dependent methyltransferases superfamily protein	70	2088	146160	41526	28.41	1412	0.68	789	0.38	0.606	0.394	28998	18564	22414	34658	0	0	0	0	0	0	0	0	0	0	0	0	41526	0
DN18571_c1_g3_i3	AT1G63180	UDP−D−glucose/UDP−D−galactose 4−epimerase 1	59	2114	124726	32073	25.72	1157	0.55	680	0.32	0.606	0.394	25213	16714	19367	31359	0	0	0	0	0	0	0	0	0	0	0	0	32073	0
DN18602_c0_g3_i6	AT5G48520	UNKNOWN	70	5489	384230	129395	33.68	3572	0.65	1717	0.31	0.622	0.378	68287	47483	48305	90760	0	0	0	0	0	0	0	0	0	0	0	0	129395	0
DN18614_c0_g1_i9	AT3G54170	FKBP12 interacting protein 37	74	2689	198986	52761	26.52	1900	0.71	1249	0.46	0.643	0.357	42099	23290	28634	52202	0	0	0	0	0	0	0	0	0	0	0	0	52761	0
DN18621_c0_g1_i1	AT1G67960	UNKNOWN	76	3275	248900	74568	29.96	2001	0.61	1052	0.32	0.609	0.391	48121	30898	36472	58841	0	0	0	0	0	0	0	0	0	0	0	0	74568	0
DN18692_c1_g1_i4	AT4G17100	UNKNOWN	71	1726	122546	28891	23.58	1000	0.58	506	0.29	0.611	0.389	26916	16121	19886	30732	0	0	0	0	0	0	0	0	0	0	0	0	28891	0
DN18699_c1_g2_i7	AT1G60230	Radical SAM superfamily protein	59	1518	89562	20900	23.34	788	0.52	405	0.27	0.628	0.372	19330	10881	14715	23736	0	0	0	0	0	0	0	0	0	0	0	0	20900	0
DN18825_c0_g8_i1	AT5G61070	histone deacetylase of the RPD3/HDA1 superfamily 18	59	4118	242962	68353	28.13	2372	0.58	1271	0.31	0.62	0.38	48508	31124	34692	60285	0	0	0	0	0	0	0	0	0	0	0	0	68353	0
DN18903_c0_g2_i1	AT4G24270	EMBRYO DEFECTIVE 140	71	6215	441265	143183	32.45	4471	0.72	2380	0.38	0.625	0.375	84818	52272	58767	102225	0	0	0	0	0	0	0	0	0	0	0	0	143183	0
DN18909_c4_g1_i2	AT1G65020	UNKNOWN	50	2364	118200	37342	31.59	1339	0.57	752	0.32	0.627	0.373	22440	14679	15370	28369	0	0	0	0	0	0	0	0	0	0	0	0	37342	0
DN18917_c2_g4_i6	AT3G51830	SAC domain−containing protein 8	51	4346	221646	72620	32.76	1935	0.45	998	0.23	0.635	0.365	42062	24915	29054	52995	0	0	0	0	0	0	0	0	0	0	0	0	72620	0
DN18972_c3_g3_i2	AT5G03770	KDO transferase A	68	2614	177752	56438	31.75	1602	0.61	956	0.37	0.649	0.351	35597	19384	22822	43511	0	0	0	0	0	0	0	0	0	0	0	0	56438	0
DN18993_c1_g9_i2	AT5G14720	Protein kinase superfamily protein	75	4625	346875	107800	31.08	3152	0.68	1858	0.40	0.631	0.369	66633	41355	45840	85247	0	0	0	0	0	0	0	0	0	0	0	0	107800	0
DN19024_c0_g1_i5	AT3G18860	transducin family protein / WD−40 repeat family protein	76	7225	549100	177129	32.26	4838	0.67	2848	0.39	0.634	0.366	99784	61572	73479	137136	0	0	0	0	0	0	0	0	0	0	0	0	177129	0
DN19026_c3_g5_i1	AT1G14030	Rubisco methyltransferase family protein	71	2552	181192	35271	19.47	1287	0.50	730	0.29	0.604	0.396	40637	26625	31164	47495	0	0	0	0	0	0	0	0	0	0	0	0	35271	0
DN19029_c0_g7_i1	AT3G59380	farnesyltransferase A	66	2267	149622	42473	28.39	1309	0.58	785	0.35	0.606	0.394	29891	19967	21803	35488	0	0	0	0	0	0	0	0	0	0	0	0	42473	0
DN19033_c0_g4_i1	AT3G42630	Pentatricopeptide repeat (PPR) superfamily protein	63	1518	95634	10997	11.50	892	0.59	398	0.26	0.619	0.381	24093	12143	20116	28285	0	0	0	0	0	0	0	0	0	0	0	0	10997	0
DN19055_c2_g2_i8	AT3G45040	phosphatidate cytidylyltransferase family protein	77	4657	358589	114976	32.06	3169	0.68	1811	0.39	0.629	0.371	60972	45903	44585	92153	0	0	0	0	0	0	0	0	0	0	0	0	114976	0
DN19072_c0_g3_i1	AT5G03430	phosphoadenosine phosphosulfate (PAPS) reductase family protein	74	3751	277574	64881	23.37	2492	0.66	1522	0.41	0.626	0.374	58229	35196	44083	75185	0	0	0	0	0	0	0	0	0	0	0	0	64881	0
DN19090_c1_g1_i2	AT2G42700	UNKNOWN	64	3688	236032	51009	21.61	1702	0.46	915	0.25	0.594	0.406	51230	33031	40231	60531	0	0	0	0	0	0	0	0	0	0	0	0	51009	0

DN19099_c0_g4_i2	AT1G35190	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein	47	2946	138462	48040	34.70	1870	0.64	1040	0.35	0.635	0.365	26703	15444	17176	31099	0	0	0	0	0	0	0	0	0	0	0	0	48040	0
DN19125_c0_g1_i1	AT3G05350	Metallopeptidase M24 family protein	69	5066	349554	105315	30.13	3061	0.60	1718	0.34	0.616	0.384	68003	47141	46053	83042	0	0	0	0	0	0	0	0	0	0	0	0	105315	0
DN19154_c1_g1_i1	AT5G39590	TLD-domain containing nucleolar protein	53	2448	129744	31756	24.48	1497	0.61	783	0.32	0.618	0.382	28037	16351	21157	32443	0	0	0	0	0	0	0	0	0	0	0	0	31756	0
DN19185_c0_g1_i8	AT5G04360	limit dextrinase	59	6512	384208	130709	34.02	3374	0.52	1882	0.29	0.639	0.361	72597	47175	44336	89391	0	0	0	0	0	0	0	0	0	0	0	0	130709	0
DN19202_c1_g2_i1	AT1G64600	methyltransferasescopper ion binding	72	3838	276336	93293	33.76	2485	0.65	1332	0.35	0.618	0.382	53300	30721	38058	60964	0	0	0	0	0	0	0	0	0	0	0	0	93293	0
DN19202_c1_g2_i6	AT5G08720	UNKNOWN	64	3197	204608	51648	25.24	1778	0.56	1031	0.32	0.617	0.383	47748	25856	32355	47001	0	0	0	0	0	0	0	0	0	0	0	0	51648	0
DN19224_c1_g1_i2	AT5G05310	TLC ATP/ADP transporter	75	2980	223500	72268	32.34	1955	0.66	1010	0.34	0.637	0.363	39786	27385	27140	56921	0	0	0	0	0	0	0	0	0	0	0	0	72268	0
DN19232_c0_g2_i7	AT2G26510	Xanthine/uracil permease family protein	58	2996	173768	53212	30.62	1418	0.47	862	0.29	0.617	0.383	31670	21792	23056	44038	0	0	0	0	0	0	0	0	0	0	0	0	53212	0
DN19243_c1_g1_i3	AT4G30310	FGGY family of carbohydrate kinase	67	4892	327764	103964	31.72	3045	0.62	1553	0.32	0.623	0.377	62854	38208	45790	76948	0	0	0	0	0	0	0	0	0	0	0	0	103964	0
DN19273_c2_g5_i1	AT4G27340	Met-10+ like family protein	55	2721	149655	47601	31.81	1447	0.53	828	0.30	0.648	0.352	28682	16071	19891	37410	0	0	0	0	0	0	0	0	0	0	0	0	47601	0
DN19278_c0_g1_i2	AT4G34900	xanthine dehydrogenase 2AT4G34890	77	6057	466389	127729	27.39	3417	0.56	1915	0.32	0.614	0.386	94969	60787	69161	113743	0	0	0	0	0	0	0	0	0	0	0	0	127729	0
DN19286_c0_g1_i1	AT2G43950	chloroplast outer envelope protein 37	65	2279	148135	45852	30.95	1384	0.61	754	0.33	0.622	0.378	28476	18403	20317	35087	0	0	0	0	0	0	0	0	0	0	0	0	45852	0
DN19287_c1_g2_i5	AT5G51340	Tetratricopeptide repeat (TPR)-like superfamily protein	62	4098	254076	80084	31.52	1783	0.44	969	0.24	0.602	0.398	47388	32813	35846	57945	0	0	0	0	0	0	0	0	0	0	0	0	80084	0
DN19305_c1_g2_i6	AT2G25530	AFG1-like ATPase family protein	67	4427	296609	91420	30.82	2390	0.54	1320	0.30	0.629	0.371	58440	34991	40212	71546	0	0	0	0	0	0	0	0	0	0	0	0	91420	0
DN19315_c0_g1_i2	AT1G59600	ZCW7	77	2371	182567	45841	25.11	1636	0.69	921	0.39	0.603	0.397	37827	22577	31255	45067	0	0	0	0	0	0	0	0	0	0	0	0	45841	0
DN19331_c2_g1_i2	AT3G26670	Protein of unknown function (DUF803)	75	4198	314850	85758	27.24	2661	0.63	1519	0.36	0.64	0.36	62300	39913	41959	84920	0	0	0	0	0	0	0	0	0	0	0	0	85758	0
DN19360_c3_g4_i6	AT4G33630	Protein of unknown function (DUF3506)	61	4624	282064	73458	26.04	2406	0.52	1184	0.26	0.604	0.396	58380	34435	47157	68634	0	0	0	0	0	0	0	0	0	0	0	0	73458	0
DN19396_c1_g6_i1	AT5G50840	UNKNOWN	55	2894	159170	54947	34.52	1410	0.49	828	0.29	0.634	0.366	30205	17311	20116	36591	0	0	0	0	0	0	0	0	0	0	0	0	54947	0
DN19448_c2_g2_i2	AT3G57790	Pectin lyase-like superfamily protein	68	2092	142256	19557	13.75	1168	0.56	717	0.34	0.563	0.437	34077	25777	27703	35142	0	0	0	0	0	0	0	0	0	0	0	0	19557	0
DN19464_c3_g1_i1	AT5G19540	UNKNOWN	68	2467	167756	32550	19.40	1299	0.53	723	0.29	0.61	0.39	38155	25378	27305	44368	0	0	0	0	0	0	0	0	0	0	0	0	32550	0
DN19464_c3_g1_i2	AT5G19540	UNKNOWN	75	2190	164250	32674	19.89	1358	0.62	675	0.31	0.606	0.394	35779	24906	26904	43987	0	0	0	0	0	0	0	0	0	0	0	0	32674	0
DN19515_c0_g1_i3	AT3G52030	F-box family protein with WD40/YVTN repeat doamin	70	3460	242200	75770	31.28	2060	0.60	1172	0.34	0.617	0.383	45200	28796	34431	58003	0	0	0	0	0	0	0	0	0	0	0	0	75770	0
DN19519_c0_g1_i1	AT3G04460	peroxin-12	59	2610	153990	54605	35.46	1383	0.53	675	0.26	0.614	0.386	26642	18415	19439	34889	0	0	0	0	0	0	0	0	0	0	0	0	54605	0
DN19552_c0_g3_i9	AT3G20790	NAD(P)-binding Rossmann-fold superfamily protein	62	2720	168640	53720	31.86	1874	0.69	966	0.36	0.62	0.38	30731	20977	22204	41008	0	0	0	0	0	0	0	0	0	0	0	0	53720	0
DN19554_c0_g1_i4	AT2G31040	ATP synthase protein I -related	70	2192	153440	39629	25.83	1462	0.67	900	0.41	0.608	0.392	31155	22014	22139	38503	0	0	0	0	0	0	0	0	0	0	0	0	39629	0
DN19564_c2_g1_i2	AT3G26570	phosphate transporter 21	69	2422	167118	36808	22.03	1355	0.56	721	0.30	0.56	0.44	32248	28576	28632	40854	0	0	0	0	0	0	0	0	0	0	0	0	36808	0
DN19782_c3_g1_i1	AT1G01040	dicer-like 1	80	9627	770160	144911	18.82	5017	0.52	2529	0.26	0.608	0.392	177219	105052	138472	204506	0	0	0	0	0	0	0	0	0	0	0	0	144911	0
DN21247_c0_g1_i1	AT3G50120	Plant protein of unknown function (DUF247)AT3G50180	70	1847	129290	21538	16.66	985	0.53	613	0.33	0.512	0.488	29330	26124	26687	25611	0	0	0	0	0	0	0	0	0	0	0	0	21538	0
DN3835_c0_g1_i1	AT1G64350	Transducin/WD40 repeat-like superfamily protein	59	2848	168032	60807	36.19	1748	0.61	861	0.30	0.627	0.373	29663	16982	22865	37715	0	0	0	0	0	0	0	0	0	0	0	0	60807	0
DN6748_c0_g1_i1	AT1G74460	GDSL-like Lipase/Acylhydrolase superfamily protein	73	1937	141401	28296	20.01	1232	0.64	803	0.42	0.605	0.395	32903	22092	22413	35697	0	0	0	0	0	0	0	0	0	0	0	0	28296	0
DN6798_c0_g1_i1	AT2G28070	ABC-2 type transporter family protein	67	3776	252992	63948	25.28	1699	0.45	904	0.24	0.615	0.385	51276	34501	37784	65483	0	0	0	0	0	0	0	0	0	0	0	0	63948	0
DN6810_c0_g1_i1	AT1G08960	cation exchanger 11	81	2101	170181	32805	19.28	1090	0.52	615	0.29	0.562	0.438	30851	29030	30925	46570	0	0	0	0	0	0	0	0	0	0	0	0	32805	0
DN713_c0_g1_i1	AT3G20780	topoisomerase 6 subunit B	59	4666	275294	74779	27.16	2343	0.50	1283	0.28	0.626	0.374	58562	36198	38618	67137	0	0	0	0	0	0	0	0	0	0	0	0	74779	0
DN7486_c0_g1_i1	AT3G05410	Photosystem II reaction center PsbP family protein	69	1325	91425	28252	30.90	947	0.72	572	0.43	0.558	0.442	15957	12154	16145	18917	0	0	0	0	0	0	0	0	0	0	0	0	28252	0
DN7963_c0_g1_i1	AT3G05040	ARM repeat superfamily protein	76	7414	563464	129225	22.93	4644	0.63	2444	0.33	0.624	0.376	120834	76905	85483	151017	0	0	0	0	0	0	0	0	0	0	0	0	129225	0
DN8581_c0_g1_i1	AT5G47500	Pectin lyase-like superfamily protein	67	2236	149812	45813	30.58	1465	0.66	981	0.44	0.605	0.395	30106	20096	20312	33485	0	0	0	0	0	0	0	0	0	0	0	0	45813	0
DN8707_c0_g1_i1	AT1G32160	Protein of unknown function (DUF760)	60	1717	103020	18130	17.60	855	0.50	413	0.24	0.605	0.395	24237	14767	18652	27234	0	0	0	0	0	0	0	0	0	0	0	0	18130	0
DN9853_c0_g1_i1	AT1G09300	Metallopeptidase M24 family protein	73	4086	298278	97974	32.85	2239	0.55	1331	0.33	0.618	0.382	55316	34551	40957	69480	0	0	0	0	0	0	0	0	0	0	0	0	97974	0
DN9873_c0_g1_i1	AT3G20440	Alpha amylase family protein	78	6727	524706	152054	28.98	4299	0.64	2325	0.35	0.624	0.376	106106	66569	71884	128093	0	0	0	0	0	0	0	0	0	0	0	0	152054	0



Appendix 16. A phylogenetic tree based on concatenated ML analysis using RAxML.

Appendix 17. The information on the shared quartet percentage for 272 loci calculated by Quartet (sorted in order of highest percentage).

Gene No. of *S. baicalensis*	Gene Function	Gene No. of *Arabidopsis*	N	Q	s	Shared quartet Percentage	d	r1	r2	u
Concatenated matrix			910252	455126	455126	100.00	0	0	0	0
DN17712_c1_g5_i1	SNF2 domain-containing protein/helicase domain-containing protein/F-box family protein;	AT3G54460	848540	424270	397194	87.27	27076	0	0	0
DN6748_c0_g1_i1	GDSL-like Lipase/Acylhydrolase superfamily protein;	AT1G74460	848540	424270	392782	86.30	31488	0	0	0
DN18388_c0_g1_i1	ABC-2 type transporter family protein;	AT2G01320	848540	424270	385123	84.62	39147	0	0	0
DN18903_c0_g2_i1	EMBRYO DEFECTIVE 140;	AT4G24270	790020	395010	381986	83.93	13024	0	0	0
DN10161_c0_g1_i1	DNAJ heat shock N-terminal domain-containing protein;	AT2G35720	848540	424270	377723	82.99	46547	0	0	0
DN9853_c0_g1_i1	Metallopeptidase M24 family protein;	AT1G09300	910252	455126	371202	81.56	83924	0	0	0
DN19026_c3_g5_i1	Rubisco methyltransferase family protein;	AT1G14030	790020	395010	366602	80.55	28408	0	0	0
DN17366_c0_g12_i1	CAX-interacting protein 2;	AT2G38270	790020	395010	364115	80.00	30895	0	0	0
DN13880_c0_g1_i6	Spo11/DNA topoisomerase VI, subunit A protein;	AT3G13170	848540	424270	363376	79.84	60894	0	0	0
DN16384_c0_g1_i4	homolog of RAD54;	AT3G19210	848540	424270	363274	79.82	60996	0	0	0
DN15269_c0_g2_i1	FG-GAP repeat-containing protein;	AT3G51050	790020	395010	356834	78.40	38176	0	0	0
DN13303_c0_g1_i1	initiator tRNA phosphoribosyl transferase family protein;	AT2G40570	848540	424270	355434	78.10	68836	0	0	0
DN17651_c0_g1_i6	ARM repeat superfamily protein;	AT5G27970	848540	424270	355203	78.04	69067	0	0	0
DN17394_c1_g1_i1	peroxin 6;	AT1G03000	848540	424270	355081	78.02	69189	0	0	0
DN14446_c0_g1_i1	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein;	AT4G36470	848540	424270	354601	77.91	69669	0	0	0
DN18692_c1_g1_i4	UNKNOWN;	AT4G17100	848540	424270	353531	77.68	70739	0	0	0
DN17905_c0_g1_i2	ARM repeat superfamily protein;	AT2G13540	848540	424270	352342	77.42	71928	0	0	0
DN18263_c0_g2_i2	UNKNOWN;	AT1G69070	910252	455126	352142	77.37	102984	0	0	0
DN16541_c1_g3_i1	helicase in vascular tissue and tapetum;	AT2G30800	848540	424270	351153	77.16	73117	0	0	0
DN19278_c0_g1_i2	xanthine dehydrogenase 2;AT4G34890	AT4G34900	848540	424270	350513	77.01	73757	0	0	0
DN16461_c0_g1_i1	Endoplasmic reticulum vesicle transporter protein;	AT3G22290	848540	424270	349497	76.79	74773	0	0	0
DN14455_c0_g1_i1	protein affected TRAFfiCKING	AT3G55480	848540	424270	349478	76.79	74792	0	0	0
DN18339_c0_g2_i2	NONE	NONE	790020	395010	348208	76.51	46802	0	0	0
DN15758_c0_g1_i4	P-loop containing nucleoside triphosphate hydrolases superfamily protein;	AT1G80770	848540	424270	345883	76.00	78387	0	0	0
DN11234_c0_g1_i1	ubiquitin-associated (UBA)/TS-N domain-containing protein;	AT1G04850	848540	424270	344466	75.69	79804	0	0	0
DN11430_c0_g1_i1	RNA helicase 36;	AT1G16280	848540	424270	344278	75.64	79992	0	0	0
DN19782_c3_g1_i1	dicer-like 1;	AT1G01040	848540	424270	343970	75.58	80300	0	0	0
DN15853_c0_g4_i2	UNKNOWN;	AT5G66810	790020	395010	343693	75.52	51317	0	0	0
DN15349_c0_g1_i5	Trypsin family protein with PDZ domain;	AT5G39830	790020	395010	342307	75.21	52703	0	0	0
DN14903_c0_g1_i2	modifier of snc1,4;	AT3G18165	848540	424270	342280	75.21	81990	0	0	0
DN14633_c1_g1_i5	Protein of unknown function (DUF185);	AT1G04900	790020	395010	341697	75.08	53313	0	0	0
DN14928_c0_g1_i2	DHBP synthase RibB-like alpha/beta domain;	AT3G01920	790020	395010	340253	74.76	54757	0	0	0
DN19448_c2_g2_i2	Pectin lyase-like superfamily protein;	AT3G57790	734580	367290	340027	74.71	27263	0	0	0
DN14467_c0_g1_i1	equilibrative nucleotide transporter 1;	AT1G70330	734580	367290	339842	74.67	27448	0	0	0
DN12972_c0_g2_i1	Transducin/WD40 repeat-like superfamily protein;	AT4G32990	790020	395010	339409	74.57	55601	0	0	0
DN18538_c2_g1_i2	nuclear matrix protein-related;	AT5G09860	790020	395010	339149	74.52	55861	0	0	0
DN15986_c0_g1_i3	UNKNOWN;	AT2G23390	848540	424270	338864	74.45	85406	0	0	0
DN7486_c0_g1_i1	Photosystem II reaction center PsbP family protein;	AT3G05410	910252	455126	337749	74.21	117377	0	0	0
DN16529_c0_g1_i2	ARM repeat superfamily protein;	AT3G02710	790020	395010	337736	74.21	57274	0	0	0
DN16725_c0_g3_i2	DegP protease 1;	AT3G27925	790020	395010	337302	74.11	57708	0	0	0
DN12002_c0_g2_i1	Peptidase M28 family protein;	AT3G54720	790020	395010	337153	74.08	57857	0	0	0

DN13155_c0_g1_i1	UNKNOWN;	AT1G09280	790020	395010	335242	73.66	59768	0	0	0
DN15736_c0_g1_i1	Radical SAM superfamily protein;	AT3G19630	734580	367290	334292	73.45	32998	0	0	0
DN13476_c0_g1_i1	HAD-superfamily hydrolase, subfamily IG, 5'-nucleotidase;	AT2G23890	848540	424270	334101	73.41	90169	0	0	0
DN11731_c1_g1_i2	phosphate transporter 4;3;	AT3G46980	848540	424270	333795	73.34	90475	0	0	0
DN9873_c0_g1_i1	Alpha amylase family protein;	AT3G20440	790020	395010	332295	73.01	62715	0	0	0
DN18050_c3_g5_i1	ALBINA 1;	AT1G08520	790020	395010	331898	72.92	63112	0	0	0
DN17671_c0_g1_i4	HAD-superfamily hydrolase, subfamily IG, 5'-nucleotidase;	AT5G48960	790020	395010	331594	72.86	63416	0	0	0
DN18602_c0_g3_i6	UNKNOWN;	AT5G48520	790020	395010	331553	72.85	63457	0	0	0
DN16611_c2_g12_i3	SAC3/GANP/Nin1/mts3/eIF-3 p25 family;	AT3G54380	790020	395010	331038	72.74	63972	0	0	0
DN13497_c0_g1_i3	6-phosphogluconate dehydrogenase family protein;	AT4G20930	790020	395010	330353	72.58	64657	0	0	0
DN18174_c0_g2_i1	Metallo-hydrolase/oxidoreductase superfamily protein;	AT1G30300	790020	395010	329930	72.49	65080	0	0	0
DN14140_c0_g1_i1	Transducin/WD40 repeat-like superfamily protein;	AT4G29830	790020	395010	329317	72.36	65693	0	0	0
DN16369_c0_g1_i2	ABC2 homolog 9;	AT2G40090	790020	395010	329056	72.30	65954	0	0	0
DN17217_c0_g1_i1	Zn-dependent exopeptidases superfamily protein;	AT3G52640	790020	395010	328001	72.07	67009	0	0	0
DN19090_c1_g1_i2	UNKNOWN;	AT2G42700	790020	395010	327660	71.99	67350	0	0	0
DN16986_c1_g6_i2	zeta-carotene desaturase;	AT3G04870	734580	367290	327328	71.92	39962	0	0	0
DN12536_c0_g1_i1	UNKNOWN;	AT5G42470	734580	367290	327283	71.91	40007	0	0	0
DN17772_c0_g2_i4	Membrane trafficking VPS53 family protein;	AT1G50500	910252	455126	326709	71.78	128417	0	0	0
DN19202_c1_g2_i19	methyltransferases;copper ion binding;	AT1G64600	734580	367290	326495	71.74	40795	0	0	0
DN15425_c0_g1_i1	pentatricopeptide (PPR) repeat-containing protein;	AT5G46580	790020	395010	325655	71.55	69355	0	0	0
DN13037_c0_g1_i1	Transducin/WD40 repeat-like superfamily protein;	AT5G15550	790020	395010	325504	71.52	69506	0	0	0
DN15237_c0_g2_i2	potassium channel tetramerisation domain-containing protein / pentapeptide repeat-containing protein;	AT5G55000	790020	395010	323240	71.02	71770	0	0	0
DN19515_c0_g1_i3	F-box family protein with WD40/YVTN repeat doamin;	AT3G52030	734580	367290	322724	70.91	44566	0	0	0
DN11381_c0_g1_i2	GDSL-like Lipase/Acylhydrolase superfamily protein;	AT3G48460	734580	367290	321214	70.58	46076	0	0	0
DN15755_c0_g2_i1	vacuoleless1 (VCL1);	AT2G38020	734580	367290	320961	70.52	46329	0	0	0
DN19125_c0_g1_i1	Metallopeptidase M24 family protein;	AT3G05350	790020	395010	320365	70.39	74645	0	0	0
DN14660_c0_g2_i1	alpha/beta-Hydrolases superfamily protein;	AT1G74640	790020	395010	320064	70.32	74946	0	0	0
DN18621_c0_g1_i1	UNKNOWN;	AT1G67960	790020	395010	319093	70.11	75917	0	0	0
DN8707_c0_g1_i1	Protein of unknown function (DUF760);	AT1G32160	790020	395010	318505	69.98	76505	0	0	0
DN16418_c0_g1_i5	Glycosyl hydrolase superfamily protein;	AT4G01040	848540	424270	317975	69.87	106295	0	0	0
DN10173_c0_g2_i2	ankyrin repeat family protein;	AT2G04740	790020	395010	316773	69.60	78237	0	0	0
DN17742_c0_g1_i1	Ribophorin I;	AT1G76400	848540	424270	316710	69.59	107560	0	0	0
DN13289_c0_g1_i2	UNKNOWN;	AT1G64680	790020	395010	315990	69.43	79020	0	0	0
DN19360_c3_g4_i6	Protein of unknown function (DUF3506);	AT4G33630	734580	367290	313879	68.97	53411	0	0	0
DN15095_c0_g1_i1	Trypsin family protein with PDZ domain;	AT5G27660	848540	424270	313636	68.91	110634	0	0	0
DN15969_c0_g1_i5	UDP-Glycosyltransferase superfamily protein;	AT1G78800	790020	395010	313395	68.86	81615	0	0	0
DN19185_c0_g1_i8	limit dextrinase;	AT5G04360	734580	367290	311935	68.54	55355	0	0	0
DN14300_c0_g1_i1	galacturonic acid kinase;	AT3G10700	734580	367290	311911	68.53	55379	0	0	0
DN11363_c0_g3_i2	uracil dna glycosylase;	AT3G18630	734580	367290	311218	68.38	56072	0	0	0
DN15511_c0_g3_i4	Tetratricopeptide repeat (TPR)-like superfamily protein;	AT2G35130	734580	367290	310799	68.29	56491	0	0	0
DN713_c0_g1_i1	topoisomerase 6 subunit B;	AT3G20780	734580	367290	310306	68.18	56984	0	0	0
DN11436_c0_g1_i2	TraB family protein;	AT5G52030	848540	424270	309007	67.89	115263	0	0	0
DN19055_c2_g2_i8	phosphatidate cytidylyltransferase family protein;	AT3G45040	848540	424270	308768	67.84	115502	0	0	0
DN16728_c3_g1_i5	ARM repeat superfamily protein;	AT4G27640	790020	395010	306862	67.42	88148	0	0	0
DN16386_c1_g1_i1	UDP-glucose pyrophosphorylase 3;	AT3G56040	848540	424270	305717	67.17	118553	0	0	0
DN18332_c0_g1_i2	Regulator of chromosome condensation (RCC1) family protein;	AT5G11580	790020	395010	305117	67.04	89893	0	0	0
DN15524_c0_g1_i4	RNA helicase family protein;	AT4G18465	734580	367290	304421	66.89	62869	0	0	0

DN17740_c0_g2_i1	Rhodanese/Cell cycle control phosphatase superfamily protein;	AT2G40760	682110	341055	303267	66.63	37788	0	0	0
DN19315_c0_g1_i2	ZCW7;	AT1G59600	790020	395010	301867	66.33	93143	0	0	0
DN18120_c0_g1_i2	methylcrotonyl−CoA carboxylase alpha chain, mitochondrial / 3−methylcrotonyl−CoA carboxylase 1 (MCCA);	AT1G03090	734580	367290	301639	66.28	65651	0	0	0
DN15516_c0_g1_i1	mitogen−activated protein kinase kinase 3;	AT5G40440	734580	367290	301614	66.27	65676	0	0	0
DN19305_c1_g2_i6	AFG1−like ATPase family protein;	AT2G25530	682110	341055	301529	66.25	39526	0	0	0
DN17092_c0_g1_i4	DEAD/DEAH box helicase, putative;	AT3G02060	790020	395010	301170	66.17	93840	0	0	0
DN16194_c1_g3_i3	RING/U−box superfamily protein;	AT3G19895	682110	341055	299588	65.83	41467	0	0	0
DN16048_c0_g1_i3	alpha−amylase−like 3;	AT1G69830	682110	341055	298789	65.65	42266	0	0	0
DN17465_c0_g2_i3	Ubiquitin−associated (UBA) protein;	AT3G56740	682110	341055	298505	65.59	42550	0	0	0
DN17639_c2_g7_i1	UNKNOWN;	AT5G65490	848540	424270	297470	65.36	126800	0	0	0
DN14329_c0_g1_i1	FAD/NAD(P)−binding oxidoreductase family protein;	AT1G57770	734580	367290	296892	65.23	70398	0	0	0
DN17736_c0_g1_i4	porphyromonas−type peptidyl−arginine deiminase family protein;	AT5G08170	734580	367290	296751	65.20	70539	0	0	0
DN16176_c0_g3_i4	esterase/lipase/thioesterase family protein;	AT4G10050	682110	341055	296122	65.06	44933	0	0	0
DN19519_c0_g1_i1	peroxin−12;	AT3G04460	682110	341055	296079	65.05	44976	0	0	0
DN16815_c0_g1_i2	elongation factor family protein;	AT2G31060	682110	341055	295325	64.89	45730	0	0	0
DN13596_c0_g1_i2	Putative methyltransferase family protein;	AT2G26810	682110	341055	294365	64.68	46690	0	0	0
DN19202_c1_g2_i6	UNKNOWN;	AT5G08720	790020	395010	293940	64.58	101070	0	0	0
DN12152_c0_g2_i1	2−oxoglutarate (2OG) and Fe(II)−dependent oxygenase superfamily protein;	AT3G20810	734580	367290	293688	64.53	73602	0	0	0
DN15130_c0_g1_i1	embryo defective 2423;	AT3G48470	682110	341055	292973	64.37	48082	0	0	0
DN13973_c0_g4_i1	UNKNOWN;	AT2G40390	682110	341055	292797	64.33	48258	0	0	0
DN6798_c0_g1_i1	ABC−2 type transporter family protein;	AT2G28070	790020	395010	292301	64.22	102709	0	0	0
DN15203_c0_g2_i5	zinc finger (Ran−binding) family protein;	AT2G17975	632502	316251	292210	64.20	24041	0	0	0
DN19033_c0_g4_i1	Pentatricopeptide repeat (PPR) superfamily protein;	AT3G42630	790020	395010	291689	64.09	103321	0	0	0
DN19072_c0_g3_i1	phosphoadenosine phosphosulfate (PAPS) reductase family protein;	AT5G03430	682110	341055	291517	64.05	49538	0	0	0
DN12355_c0_g1_i1	histone acetyltransferase of the GNAT family 2;	AT5G56740	734580	367290	291492	64.05	75798	0	0	0
DN12467_c0_g1_i1	Nucleic acid−binding, OB−fold−like protein;	AT3G48500	790020	395010	291483	64.04	103527	0	0	0
DN10718_c0_g2_i1	Pseudouridine synthase family protein;	AT4G21770	632502	316251	291389	64.02	24862	0	0	0
DN14847_c0_g2_i1	P−loop containing nucleoside triphosphate hydrolases superfamily protein;	AT1G01910	682110	341055	291172	63.98	49883	0	0	0
DN10654_c0_g1_i1	Transducin/WD40 repeat−like superfamily protein;	AT4G30840	734580	367290	290990	63.94	76300	0	0	0
DN13905_c0_g1_i1	TBP−associated factor 5;	AT5G25150	682110	341055	290973	63.93	50082	0	0	0
DN12874_c0_g1_i2	decoy;	AT1G14620	682110	341055	290875	63.91	50180	0	0	0
DN17071_c0_g1_i16	P−loop containing nucleoside triphosphate hydrolases superfamily protein;	AT3G19720	790020	395010	289747	63.66	105263	0	0	0
DN15732_c0_g1_i1	Transducin/WD40 repeat−like superfamily protein;	AT5G45760	734580	367290	288392	63.37	78898	0	0	0
DN17124_c0_g1_i3	histone deacetylase 8;	AT1G08460	734580	367290	288197	63.32	79093	0	0	0
DN12024_c0_g2_i1	enolase 1;	AT1G74030	682110	341055	287863	63.25	53192	0	0	0
DN16415_c0_g1_i1	UNKNOWN;	AT1G01770	632502	316251	287027	63.07	29224	0	0	0
DN18034_c1_g1_i4	P−loop containing nucleoside triphosphate hydrolases superfamily protein;	AT2G01460	682110	341055	285999	62.84	55056	0	0	0
DN12228_c0_g1_i2	NAD(P)−binding Rossmann−fold superfamily protein;	AT3G59710	682110	341055	285878	62.81	55177	0	0	0
DN19243_c1_g1_i3	FGGY family of carbohydrate kinase;	AT4G30310	682110	341055	285455	62.72	55600	0	0	0
DN17587_c1_g2_i4	transcription activators;DNA binding;RNA polymerase II transcription factors; catalytics;transcription initiation factors;	AT4G12610	682110	341055	285434	62.72	55621	0	0	0
DN11550_c0_g1_i1	ankyrin repeat family protein;	AT5G65860	734580	367290	284824	62.58	82466	0	0	0
DN12737_c0_g1_i1	DNA methyltransferase−2;	AT5G25480	632502	316251	284787	62.57	31464	0	0	0
DN17688_c0_g6_i1	cytochrome P450, family 97, subfamily B, polypeptide 3;	AT4G15110	682110	341055	284696	62.55	56359	0	0	0
DN14959_c0_g1_i1	UNKNOWN;	AT1G31780	734580	367290	284677	62.55	82613	0	0	0
DN14129_c0_g1_i1	ARM repeat superfamily protein;	AT3G15180	632502	316251	284153	62.43	32098	0	0	0

DN14081_c0_g1_i1	F-box/RNI-like superfamily protein;	AT5G67140	682110	341055	283989	62.40	57066	0	0	0
DN16233_c0_g1_i2	translation elongation factor Ts (EF-Ts), putative;	AT4G11120	632502	316251	282222	62.01	34029	0	0	0
DN14715_c0_g1_i3	Cation efflux family protein;	AT2G47830	682110	341055	281254	61.80	59801	0	0	0
DN8581_c0_g1_i1	Pectin lyase-like superfamily protein;	AT5G47500	682110	341055	281174	61.78	59881	0	0	0
DN16093_c0_g1_i2	NAD(P)-binding Rossmann-fold superfamily protein;	AT5G10730	734580	367290	280559	61.64	86731	0	0	0
DN19287_c1_g2_i5	Tetratricopeptide repeat (TPR)-like superfamily protein;	AT5G51340	734580	367290	280333	61.59	86957	0	0	0
DN16137_c0_g1_i1	P-loop containing nucleoside triphosphate hydrolases superfamily protein;	AT3G16840	632502	316251	280236	61.57	36015	0	0	0
DN17670_c0_g2_i7	alpha/beta-Hydrolases superfamily protein;	AT5G16120	682110	341055	280048	61.53	61007	0	0	0
DN12477_c0_g1_i1	malate synthase;	AT5G03860	734580	367290	276549	60.76	90741	0	0	0
DN16801_c0_g1_i10	coenzyme F420 hydrogenase family / dehydrogenase, beta subunit family;	AT1G04620	632502	316251	276443	60.74	39808	0	0	0
DN14166_c0_g1_i1	zinc ion binding;	AT1G12470	682110	341055	275975	60.64	65080	0	0	0
DN18356_c0_g3_i3	pyridoxin (pyrodoxamine) 5'-phosphate oxidase;	AT5G49970	734580	367290	275902	60.62	91388	0	0	0
DN15760_c0_g1_i1	Zn-dependent exopeptidases superfamily protein;	AT1G67420	734580	367290	275267	60.48	92023	0	0	0
DN18993_c1_g9_i2	Protein kinase superfamily protein;	AT5G14720	682110	341055	274200	60.25	66855	0	0	0
DN17071_c0_g8_i2	Galactose oxidase/kelch repeat superfamily protein;	AT5G50310	682110	341055	272947	59.97	68108	0	0	0
DN17821_c0_g1_i1	Rubisco methyltransferase family protein;	AT3G07670	632502	316251	272685	59.91	43566	0	0	0
DN12999_c0_g1_i1	Prenyltransferase family protein;	AT2G39550	632502	316251	271597	59.68	44654	0	0	0
DN14642_c0_g1_i1	isoamylase 1;	AT2G39930	682110	341055	271271	59.60	69784	0	0	0
DN19024_c0_g1_i5	transducin family protein / WD-40 repeat family protein;	AT3G18860	734580	367290	270700	59.48	96590	0	0	0
DN15467_c1_g1_i2	urease;	AT1G67550	734580	367290	269308	59.17	97982	0	0	0
DN11432_c0_g1_i1	peptidase M1 family protein;	AT5G13520	682110	341055	269277	59.17	71778	0	0	0
DN13371_c1_g1_i2	Core-2/I-branching beta-1,6-N-acetylglucosaminyltransferase family protein;	AT3G24040	682110	341055	268226	58.93	72829	0	0	0
DN13925_c0_g1_i2	O-methyltransferase family protein;	AT3G53140	632502	316251	267910	58.87	48341	0	0	0
DN15890_c0_g1_i1	P-loop containing nucleoside triphosphate hydrolases superfamily protein;	AT1G73110	632502	316251	267785	58.84	48466	0	0	0
DN18825_c0_g8_i1	histone deacetylase of the RPD3/HDA1 superfamily 18	AT5G61070	632502	316251	267589	58.79	48662	0	0	0
DN11241_c0_g1_i1	BTB/POZ domain-containing protein;	AT1G63850	585650	292825	266721	58.60	26104	0	0	0
DN16171_c0_g1_i4	alpha-amylase-like 2;	AT1G76130	682110	341055	266537	58.56	74518	0	0	0
DN16981_c1_g6_i1	protein phosphatase 5.2;	AT2G42810	848540	424270	265957	58.44	158313	0	0	0
DN16640_c0_g2_i1	cyclophilin71;	AT3G44600	632502	316251	265655	58.37	50596	0	0	0
DN13462_c0_g3_i1	Mitochondrial substrate carrier family protein;	AT5G56450	734580	367290	265648	58.37	101642	0	0	0
DN10655_c0_g1_i1	importin alpha isoform 9;	AT5G03070	734580	367290	265536	58.34	101754	0	0	0
DN17032_c0_g2_i2	Flavin containing amine oxidoreductase family;	AT5G14220	734580	367290	264668	58.15	102622	0	0	0
DN19224_c1_g1_i2	TLC ATP/ADP transporter;	AT5G05310	682110	341055	263046	57.80	78009	0	0	0
DN16761_c0_g1_i3	endoribonucleases;	AT3G04480	682110	341055	261795	57.52	79260	0	0	0
DN15204_c0_g1_i1	diphthamide synthesis DPH2 family protein;	AT3G59630	632502	316251	261437	57.44	54814	0	0	0
DN17392_c3_g2_i4	plastid transcriptionally active 14;	AT4G20130	632502	316251	259675	57.06	56576	0	0	0
DN17355_c0_g1_i1	GTP-binding protein, HflX;	AT5G57960	632502	316251	258786	56.86	57465	0	0	0
DN10117_c0_g1_i1	UNKNOWN;	AT1G63810	734580	367290	258597	56.82	108693	0	0	0
DN19464_c3_g1_i2	UNKNOWN;	AT5G19540	682110	341055	257349	56.54	83706	0	0	0
DN18699_c1_g2_i7	Radical SAM superfamily protein;	AT1G60230	734580	367290	255962	56.24	111328	0	0	0
DN13084_c0_g1_i2	nitroreductase family protein;	AT1G02020	734580	367290	254963	56.02	112327	0	0	0
DN15511_c0_g3_i2	Tetratricopeptide repeat (TPR)-like superfamily protein;	AT2G35130	585650	292825	254459	55.91	38366	0	0	0
DN19554_c0_g1_i4	ATP synthase protein I -related;	AT2G31040	585650	292825	254403	55.90	38422	0	0	0
DN13481_c0_g1_i3	CwfJ-like family protein / zinc finger (CCCH-type) family protein;	AT5G56900	585650	292825	253251	55.64	39574	0	0	0
DN17966_c0_g1_i6	Inositol monophosphatase family protein;	AT4G05090	632502	316251	253093	55.61	63158	0	0	0
DN10505_c0_g2_i1	peptide chain release factor, putative;	AT3G57190	632502	316251	252949	55.58	63302	0	0	0
DN19232_c0_g2_i7	Xanthine/uracil permease family protein;	AT2G26510	585650	292825	252771	55.54	40054	0	0	0

DN17681_c2_g1_i1	Metallo−hydrolase/oxidoreductase superfamily protein;	AT1G25375	632502	316251	250143	54.96	66108	0	0	0
DN13461_c0_g1_i1	Pseudouridine synthase family protein;	AT5G35400	682110	341055	249458	54.81	91597	0	0	0
DN14907_c0_g1_i1	damaged DNA binding 2;	AT5G58760	632502	316251	247409	54.36	68842	0	0	0
DN11281_c0_g2_i1	DNAse I−like superfamily protein;	AT3G48425	632502	316251	246145	54.08	70106	0	0	0
DN16959_c0_g1_i3	non−specific phospholipase C1;	AT1G07230	541450	270725	245222	53.88	25503	0	0	0
DN15904_c0_g2_i1	AAA−type ATPase family protein;	AT2G45500	632502	316251	244994	53.83	71257	0	0	0
DN19331_c2_g1_i2	Protein of unknown function (DUF803);	AT3G26670	682110	341055	244219	53.66	96836	0	0	0
DN12055_c0_g3_i1	alpha−1,3−mannosyl−glycoprotein beta−1,2−N−acetylglucosaminyltransferase, putative;	AT4G38240	585650	292825	244156	53.65	48669	0	0	0
DN19564_c2_g1_i2	phosphate transporter 2;1;	AT3G26570	585650	292825	243590	53.52	49235	0	0	0
DN18614_c0_g1_i9	FKBP12 interacting protein 37;	AT3G54170	682110	341055	241435	53.05	99620	0	0	0
DN16275_c0_g1_i1	hydrolases, acting on acid anhydrides, in phosphorus−containing nhydrides; ATP−dependent helicases;nucleic acid binding;ATP binding;RNA binding;helicases;	AT1G77030	632502	316251	239979	52.73	76272	0	0	0
DN17500_c0_g1_i1	Eukaryotic translation initiation factor eIF2A family protein;	AT1G73180	632502	316251	239419	52.60	76832	0	0	0
DN15482_c0_g1_i1	plastid transcriptionally active 12;	AT2G34640	585650	292825	239173	52.55	53652	0	0	0
DN16185_c0_g1_i1	GPI transamidase component Gpi16 subunit family protein;	AT3G07140	541450	270725	237486	52.18	33239	0	0	0
DN18089_c1_g1_i3	lecithin/cholesterol acyltransferase 3;	AT3G03310	585650	292825	237430	52.17	55395	0	0	0
DN3835_c0_g1_i1	Transducin/WD40 repeat−like superfamily protein;	AT1G64350	585650	292825	237108	52.10	55717	0	0	0
DN15144_c0_g1_i1	actin−related protein 6;	AT3G33520	541450	270725	235665	51.78	35060	0	0	0
DN15038_c0_g2_i2	Domain of unknown function (DUF303);	AT4G34215	585650	292825	235474	51.74	57351	0	0	0
DN14195_c0_g1_i2	Ribosomal L5P family protein;	AT4G01310	541450	270725	234719	51.57	36006	0	0	0
DN16453_c0_g1_i1	Domain of unknown function (DUF1995);	AT5G27560	541450	270725	233829	51.38	36896	0	0	0
DN14430_c0_g1_i4	Exostosin family protein;	AT1G21480	585650	292825	232782	51.15	60043	0	0	0
DN19396_c1_g6_i1	UNKNOWN;	AT5G50840	585650	292825	231309	50.82	61516	0	0	0
DN19273_c2_g5_i1	Met−10+ like family protein;	AT4G27340	541450	270725	230607	50.67	40118	0	0	0
DN16606_c0_g1_i1	HEAT repeat−containing protein;	AT5G01400	632502	316251	230102	50.56	86149	0	0	0
DN15150_c0_g1_i1	P−loop containing nucleoside triphosphate hydrolases superfamily protein;	AT3G47450	541450	270725	230001	50.54	40724	0	0	0
DN17093_c3_g1_i4	DWD (DDB1−binding WD40 protein) hypersensitive to ABA 1;	AT2G19430	585650	292825	229983	50.53	62842	0	0	0
DN19552_c0_g3_i9	NAD(P)−binding Rossmann−fold superfamily protein;	AT3G20790	541450	270725	229682	50.47	41043	0	0	0
DN19029_c0_g7_i1	farnesyltransferase A;	AT3G59380	541450	270725	227373	49.96	43352	0	0	0
DN13237_c0_g1_i2	alpha/beta−Hydrolases superfamily protein;	AT1G18460	585650	292825	226749	49.82	66076	0	0	0
DN11552_c0_g1_i1	aldehyde dehydrogenase 22A1;	AT3G66658	632502	316251	226020	49.66	90231	0	0	0
DN17681_c3_g5_i1	polyamine oxidase 5;	AT4G29720	541450	270725	225389	49.52	45336	0	0	0
DN13916_c0_g2_i1	Polyketide synthase, enoylreductase family protein;	AT3G45770	541450	270725	224285	49.28	46440	0	0	0
DN19286_c0_g1_i1	chloroplast outer envelope protein 37;	AT2G43950	585650	292825	222917	48.98	69908	0	0	0
DN16698_c0_g1_i1	Haloacid dehalogenase−like hydrolase (HAD) superfamily protein;	AT5G45170	541450	270725	221615	48.69	49110	0	0	0
DN18972_c3_g3_i2	KDO transferase A;	AT5G03770	541450	270725	221142	48.59	49583	0	0	0
DN11545_c0_g1_i1	Pyridoxal−5'−phosphate−dependent enzyme family protein;	AT1G55880	499800	249900	220417	48.43	29483	0	0	0
DN19464_c3_g1_i1	UNKNOWN;	AT5G19540	585650	292825	220230	48.39	72595	0	0	0
DN17625_c0_g1_i2	semialdehyde dehydrogenase family protein;	AT1G14810	499800	249900	213842	46.99	36058	0	0	0
DN18502_c1_g1_i1	Afadin/alpha−actinin−binding protein;	AT2G18876	585650	292825	212981	46.80	79844	0	0	0
DN15314_c0_g1_i1	Pentatricopeptide repeat (PPR) superfamily protein;	AT4G35850	541450	270725	212909	46.78	57816	0	0	0
DN14540_c0_g1_i3	ATP binding;	AT3G24430	541450	270725	212179	46.62	58546	0	0	0
DN21247_c0_g1_i1	Plant protein of unknown function (DUF247);AT3G50180	AT3G50120	499800	249900	210744	46.30	39156	0	0	0
DN19154_c1_g1_i1	TLD−domain containing nucleolar protein;	AT5G39590	541450	270725	209864	46.11	60861	0	0	0
DN15601_c0_g2_i2	GTP−binding family protein;	AT3G12080	541450	270725	207934	45.69	62791	0	0	0
DN18243_c0_g1_i1	RNA binding (RRM/RBD/RNP motifs) family protein;	AT5G32450	499800	249900	207432	45.58	42468	0	0	0

DN16947_c0_g3_i2	UNKNOWN;	AT5G47090	541450	270725	203731	44.76	66994	0	0	0
DN16910_c0_g1_i3	phosphoesterase;	AT3G14890	499800	249900	203070	44.62	46830	0	0	0
DN14856_c0_g1_i1	general control non-repressible 3;	AT1G64550	460600	230300	202957	44.59	27343	0	0	0
DN14484_c0_g1_i1	GLU-ADT subunit B;	AT1G48520	541450	270725	202671	44.53	68054	0	0	0
DN12199_c0_g1_i1	FAD/NAD(P)-binding oxidoreductase family protein;AT1G56000	AT1G55980	541450	270725	201458	44.26	69267	0	0	0
DN17071_c0_g1_i15	P-loop containing nucleoside triphosphate hydrolases superfamily protein;	AT3G19720	541450	270725	200707	44.10	70018	0	0	0
DN14670_c0_g1_i3	Major facilitator superfamily protein;	AT5G13400	541450	270725	200089	43.96	70636	0	0	0
DN18571_c1_g3_i3	UDP-D-glucose/UDP-D-galactose 4-epimerase 1;	AT1G63180	460600	230300	200086	43.96	30214	0	0	0
DN13915_c0_g1_i1	Major facilitator superfamily protein;	AT5G54860	499800	249900	199745	43.89	50155	0	0	0
DN11877_c0_g1_i1	plastid transcriptionally active 2;	AT1G74850	460600	230300	199487	43.83	30813	0	0	0
DN12329_c0_g1_i1	Glutathione S-transferase family protein;	AT5G44000	460600	230300	198666	43.65	31634	0	0	0
DN11368_c0_g1_i4	chromatin remodeling factor18;	AT1G48310	499800	249900	194953	42.83	54947	0	0	0
DN14783_c0_g1_i1	TRICHOME BIREFRINGENCE-LIKE 13;	AT2G14530	499800	249900	193686	42.56	56214	0	0	0
DN12637_c0_g2_i1	DNA-binding protein, putative;	AT2G03270	541450	270725	192402	42.27	78323	0	0	0
DN6810_c0_g1_i1	cation exchanger 11;	AT1G08960	499800	249900	192056	42.20	57844	0	0	0
DN18206_c0_g2_i1	cationic amino acid transporter 9;	AT1G05940	632502	316251	190363	41.83	125888	0	0	0
DN18568_c3_g4_i1	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein;	AT2G27590	460600	230300	189343	41.60	40957	0	0	0
DN12042_c0_g1_i2	Translation elongation factor EFG/EF2 protein;AT1G45332	AT2G45030	499800	249900	188648	41.45	61252	0	0	0
DN12988_c0_g1_i1	Transducin/WD40 repeat-like superfamily protein;	AT5G50230	585650	292825	187625	41.22	105200	0	0	0
DN18369_c1_g3_i1	asparagine-linked glycosylation 3;	AT2G47760	423752	211876	187086	41.11	24790	0	0	0
DN12263_c0_g2_i2	Rhodanese/Cell cycle control phosphatase superfamily protein;	AT4G24750	423752	211876	187033	41.09	24843	0	0	0
DN17790_c1_g1_i3	Glyceraldehyde-3-phosphate dehydrogenase-like family protein;	AT5G21060	499800	249900	186625	41.01	63275	0	0	0
DN16443_c0_g1_i7	Phox-associated domain;Phox-like;Sorting nexin, C-terminal;	AT1G15240	499800	249900	186152	40.90	63748	0	0	0
DN18512_c0_g1_i2	purple acid phosphatase 16;	AT3G10150	423752	211876	186133	40.90	25743	0	0	0
DN18909_c4_g1_i2	UNKNOWN;	AT1G65020	389160	194580	183003	40.21	11577	0	0	0
DN15330_c0_g1_i2	pyridoxine biosynthesis 2;	AT5G60540	460600	230300	179880	39.52	50420	0	0	0
DN17160_c0_g1_i2	Radical SAM superfamily protein;	AT2G39670	460600	230300	176429	38.76	53871	0	0	0
DN16070_c1_g1_i1	RNI-like superfamily protein;	AT5G23340	460600	230300	174076	38.25	56224	0	0	0
DN17379_c3_g4_i7	methyltransferases;	AT5G44450	389160	194580	171576	37.70	23004	0	0	0
DN10838_c0_g1_i1	beta-hexosaminidase 1;	AT3G55260	389160	194580	170816	37.53	23764	0	0	0
DN18917_c2_g4_i6	SAC domain-containing protein 8;	AT3G51830	423752	211876	168812	37.09	43064	0	0	0
DN16894_c2_g1_i2	Homocysteine S-methyltransferase family protein;	AT3G25900	356730	178365	165498	36.36	12867	0	0	0
DN13314_c0_g1_i1	RH39;	AT4G09730	423752	211876	164005	36.04	47871	0	0	0
DN12801_c0_g1_i2	Protein of unknown function (DUF1712);AT1G80910	AT1G16020	423752	211876	160994	35.37	50882	0	0	0
DN14916_c0_g1_i2	Dihydroxyacetone kinase;	AT1G48430	423752	211876	158344	34.79	53532	0	0	0
DN13265_c0_g3_i3	oxidoreductase, 2OG-Fe(II) oxygenase family protein;	AT1G11780	389160	194580	156807	34.45	37773	0	0	0
DN7963_c0_g1_i1	ARM repeat superfamily protein;	AT3G05040	356730	178365	150882	33.15	27483	0	0	0
DN18105_c4_g5_i2	electron transfer flavoprotein beta;	AT5G43430	326370	163185	141440	31.08	21745	0	0	0
DN17379_c3_g3_i6	ATP-dependent caseinolytic (Clp) protease/crotonase family protein;	AT4G13360	389160	194580	141333	31.05	53247	0	0	0
DN13513_c0_g1_i5	purple acid phosphatase 28;	AT5G57140	356730	178365	139667	30.69	38698	0	0	0
DN14001_c0_g1_i1	Mitochondrial substrate carrier family protein;	AT1G74240	326370	163185	136321	29.95	26864	0	0	0
DN17464_c0_g2_i1	Protein kinase superfamily protein;	AT1G71810	326370	163185	135569	29.79	27616	0	0	0
DN17136_c2_g2_i1	UNKNOWN;	AT1G62780	326370	163185	134528	29.56	28657	0	0	0
DN13631_c0_g1_i1	sequence-specific DNA binding transcription factors;	AT1G43860	326370	163185	122867	27.00	40318	0	0	0
DN19099_c0_g4_i2	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein;	AT1G35190	271502	135751	110232	24.22	25519	0	0	0
DN16839_c0_g1_i4	Regulator of chromosome condensation (RCC1) family protein;	AT3G02300	246820	123410	103358	22.71	20052	0	0	0
DN14976_c0_g1_i1	signal recognition particle-related / SRP-related;	AT5G61970	81840	40920	28849	6.34	12071	0	0	0

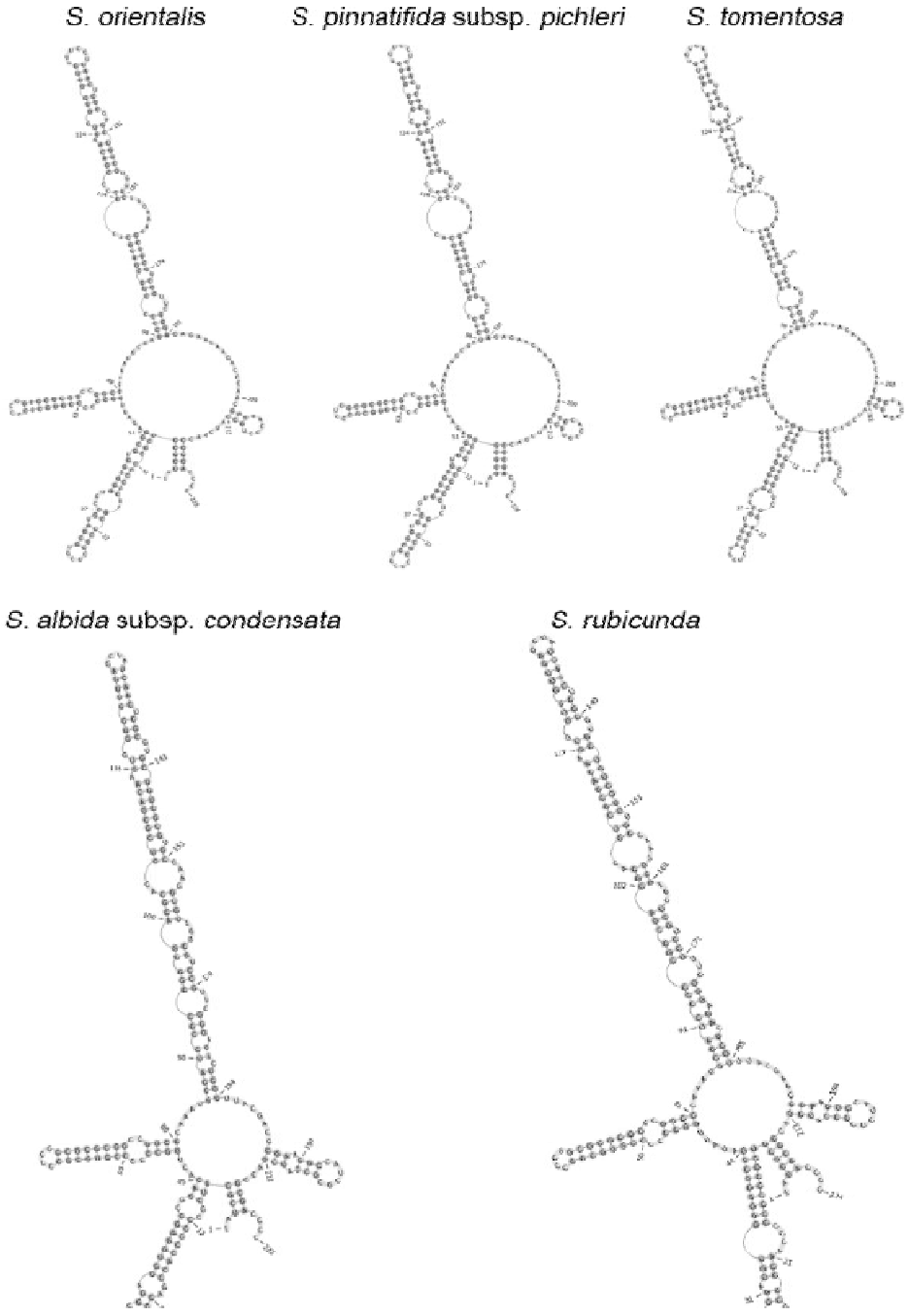
APPENDICES FOR CHAPTER 5

Appendix 18. The voucher information of 51 samples used in nrDNA analyses.

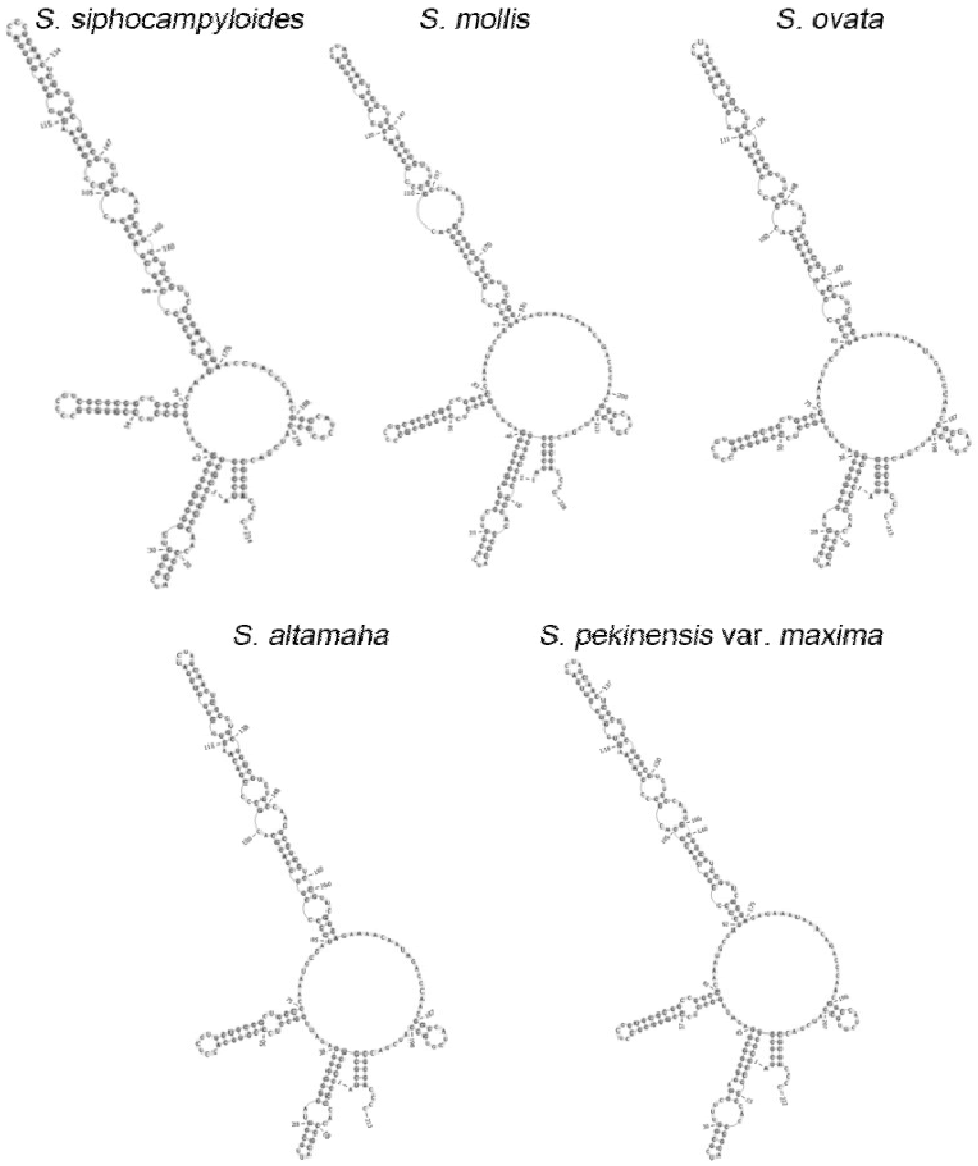
Appendix 19. The variable ITS2 secondary structures of 35 *Scutellaria* taxa.

Appendix 18. The voucher information of 51 samples used in nrDNA analyses.

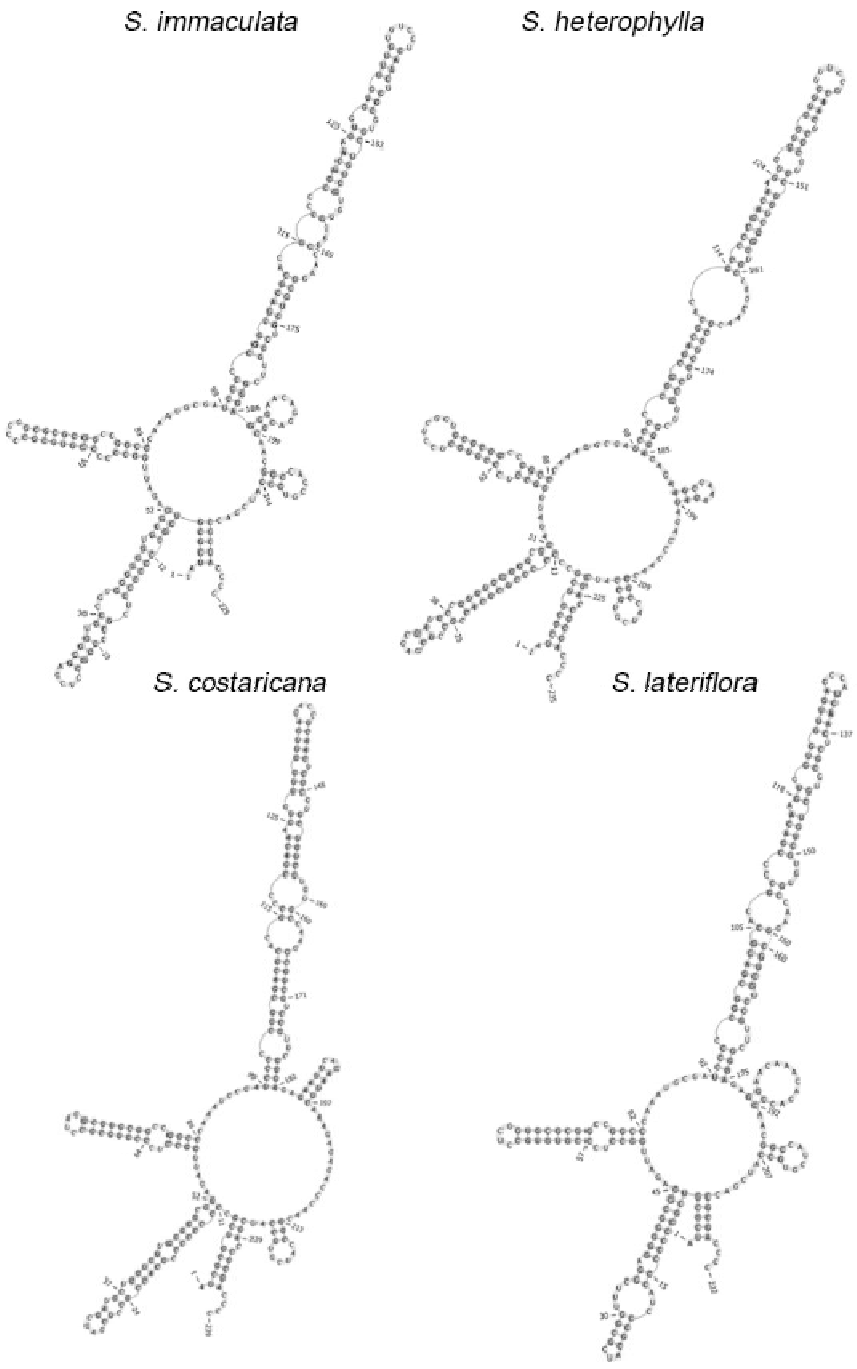
Sci name	Country	Collector and collector number (Herbarium code)
Holmskioldia sanguinea Retz.	Korea	Y. Lee 2019-0001 (SWU)
Scutellaria albida subsp. *condensata* (Rech.f.) J. R. Edm.	Iraq	R Wheeler Haines s.n. (K)
Scutellaria altamaha Small	USA	R. Kral 51180 (MO)
Scutellaria angustifolia Pursh	USA	R. Olmstead 558 (NYBG)
Scutellaria barbata D.Don	Korea	S.Kim 2019-056 (SWU)
Scutellaria bolanderi A. Gray	USA	R. Olmstead 395 (WTU)
Scutellaria costaricana H.Wendl.	USA	R. Olmstead s.n. (WTU)
Scutellaria dependens Maxim.	Japan	K. Yonekura 12880 (KYO)
Scutellaria discolor Colebr.	Myanmar	J. Murata & al. 25593 (KYO)
Scutellaria drummondii Benth.	USA	J. Bacon 2117 (WTU)
Scutellaria farsistanica Rech. f.	China	D. Z. Ji 431 (PE)
Scutellaria forrestii Diels	China	Liu Zhenwen & Zhoujing 1041 (KH)
Scutellaria gaumeri Leonard	Mexico	B. Mogensen 1130 (MO)
Scutellaria guatemalensis Leonard	Mexico	Demetrio Alvarez 11028 (NYBG)
Scutellaria hastifolia L.	Czech Republic	M.Dvorakova, M.Smejkel & J.Vichar 1552 (K)
Scutellaria heterophylla Montbret & Aucher ex Benth.	Turkey	B.Balansa 498 (K)
Scutellaria heydei Hook.f.	India	M.Grace et al. 83 (K)
Scutellaria hintoniana Epling	Mexico	A. de Avila 662 (MO)
Scutellaria hissarica B. Fedtsch.	Tajikistan	L.P.Markova et al. s.n. (K)
Scutellaria immaculata Nevski ex Juz.	Uzbekistan	Vatolkina 4757 (K)
*Scutellaria incana*Spreng.	USA	R. Olmstead 90-10 (WTU)
Scutellaria indica L.	Taiwan	B.J.Conn 4109 (K)
Scutellaria indica var. *coccinea* S. Kim & S. Lee	Republic of Korea	S. Kim 2015-0298 (SWU)
Scutellaria insignis Nakai	Republic of Korea	S. Kim 2015-0110 (SWU)
Scutellaria lateriflora L.	USA	R. Olmstead s. n. (WTU)
Scutellaria leptosiphon Nevski	Tadjikistan	R. Markova et al s.n. (K)
Scutellaria linearis Benth.	India	R.W.Parker 3396 (K)
Scutellaria longifolia Benth.	s. n.	J. J. Castillo & J. M. Vargas 2722 (NYBG)
Scutellaria mexicana (Torr.) A. J. Paton	USA	W. Messinger 257 (WTU)
Scutellaria mollis R. Br.	Australia	R. G. Coveny 9162 (K)
Scutellaria moniliorhiza Kom.	Korea	J. Kim & Y. Lee 2019-002 (SWU)
Scutellaria multicaulis Boiss.	Afghanistan	Rechinger 19308 (K)
Scutellaria nana A.Gray	USA	M. Bourell 6012 (PE)
Scutellaria nummulariifolia Hook. f.	Argentina	T. M. Pedersen 14491 (MO)
Scutellaria orientalis L.	Georgia	P. Davis 33797 (K)
Scutellaria ovata var. *clacarea* (Epling) Gleason	China	R. C. Clark 22239 (PE)
Scutellaria parvula Michx.	USA	R. Dale Thomas & Cecil Slaughter 103883 (NYBG)
Scutellaria pekinensis var. *maxima* S. Kim & S. Lee	Korea	S. Kim 2015-0108 (SWU)
Scutellaria pinnatifida subsp. *mucida* (Stapf) Rech. f.	Persia	K. H. Rechinger 43494 (K)
Scutellaria pinnatifida subsp. *pichleri* (Stapf) Rech. f.	Iraq	A.Rawi 12254 (K)
Scutellaria platensis Speg.	Argentina	B. Sparre 341 (K)
Scutellaria purpurascens Sw.	Bolivia	J. S. Solomo 8558 (MO)
Scutellaria rubicunda Hornem.	Turkey	B. Balansa 499 (K)
Scutellaria scutellarioides (Kunth) Harley	Ecuador	J. A. Hart 1616 (MO)
Scutellaria seleriana Loes.	Mexico	N. Martinez Melendez 1126 1126 (MO)
Scutellaria siphocampyloides Vatke	USA	R. Olmstead 794 (WTU)
Scutellaria splendens Link, Klotzsch & Otto	Mexico	F. J. Rendon Sandoval 571 571 (MO)
Scutellaria strigillosa Hemsl.	Korea	S. Kim 20140705-1 (SWU)
Scutellaria suffrutescens S. Watson	Mexico	James L. Reveal 3390 (NYBG)
Scutellaria tomentosa Bertol.	Iran	Rechinger 46863 (K)
Tinnea aethiopica Kotschy ex Hook.f.	South Africa	s. n. (WTU)



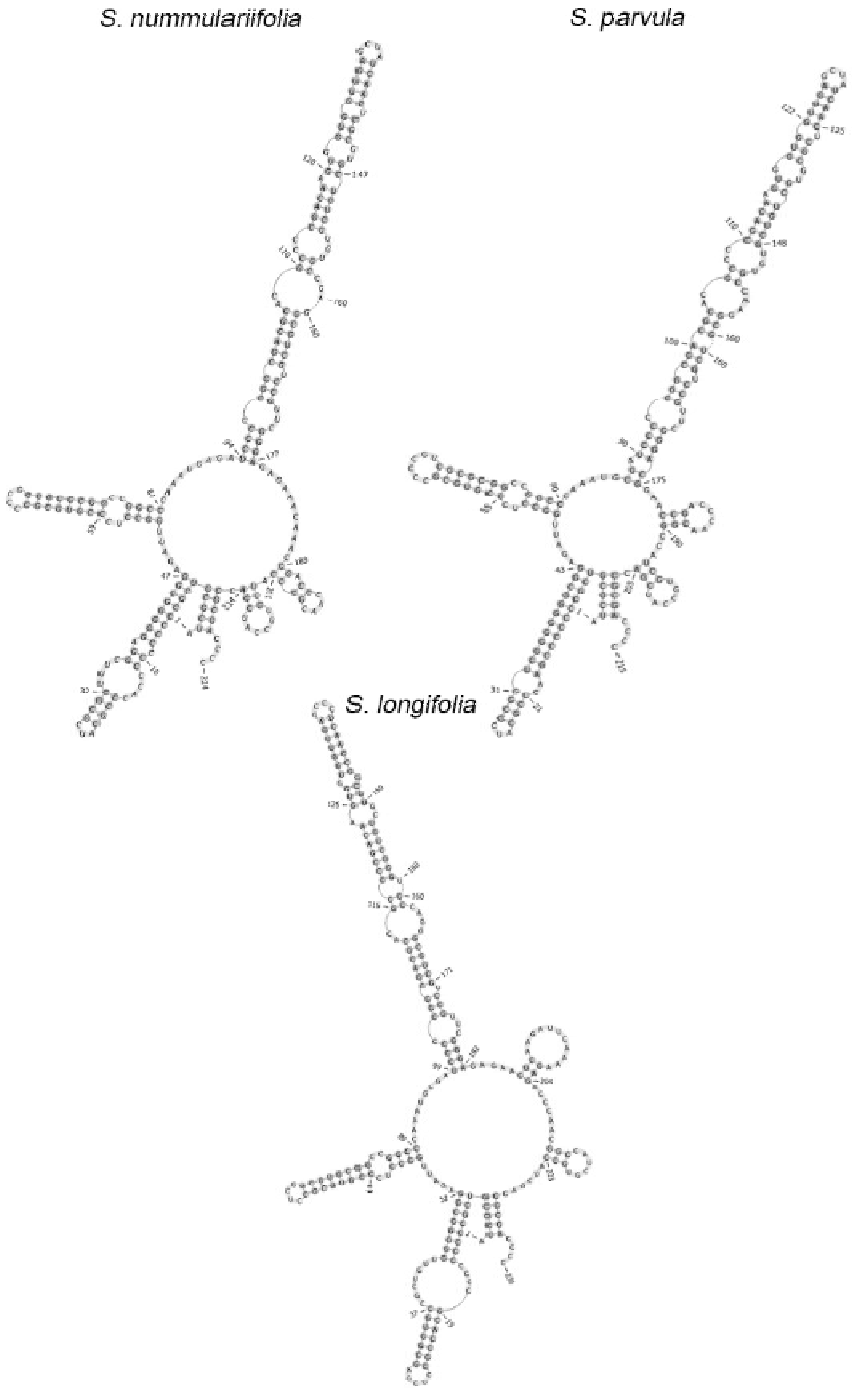
Appendix 19. The variable ITS2 secondary structures of 35 *Scutellaria* taxa. Five taxa included in Type I.



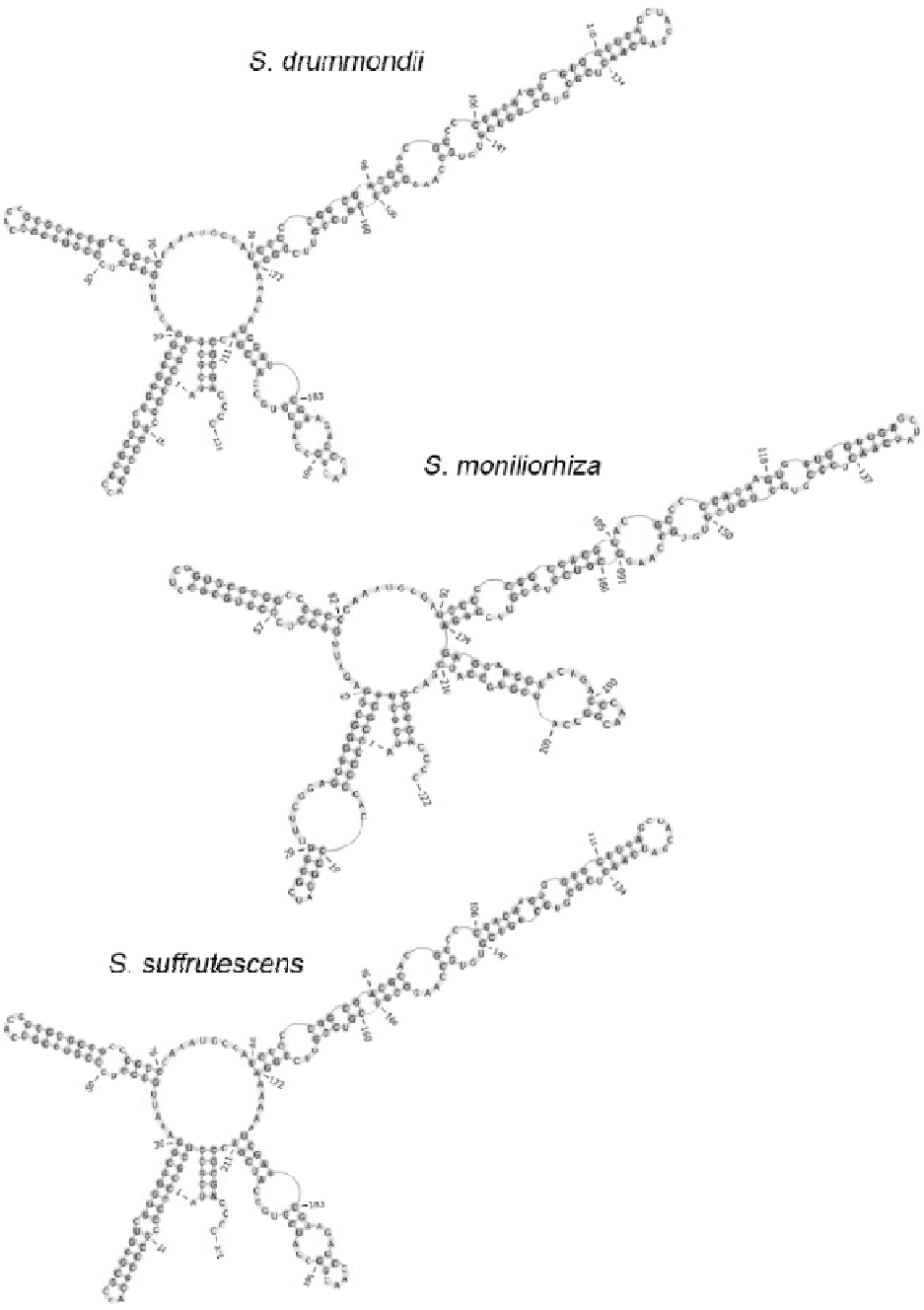
Appendix 19. Continued. Five taxa included in Type I.



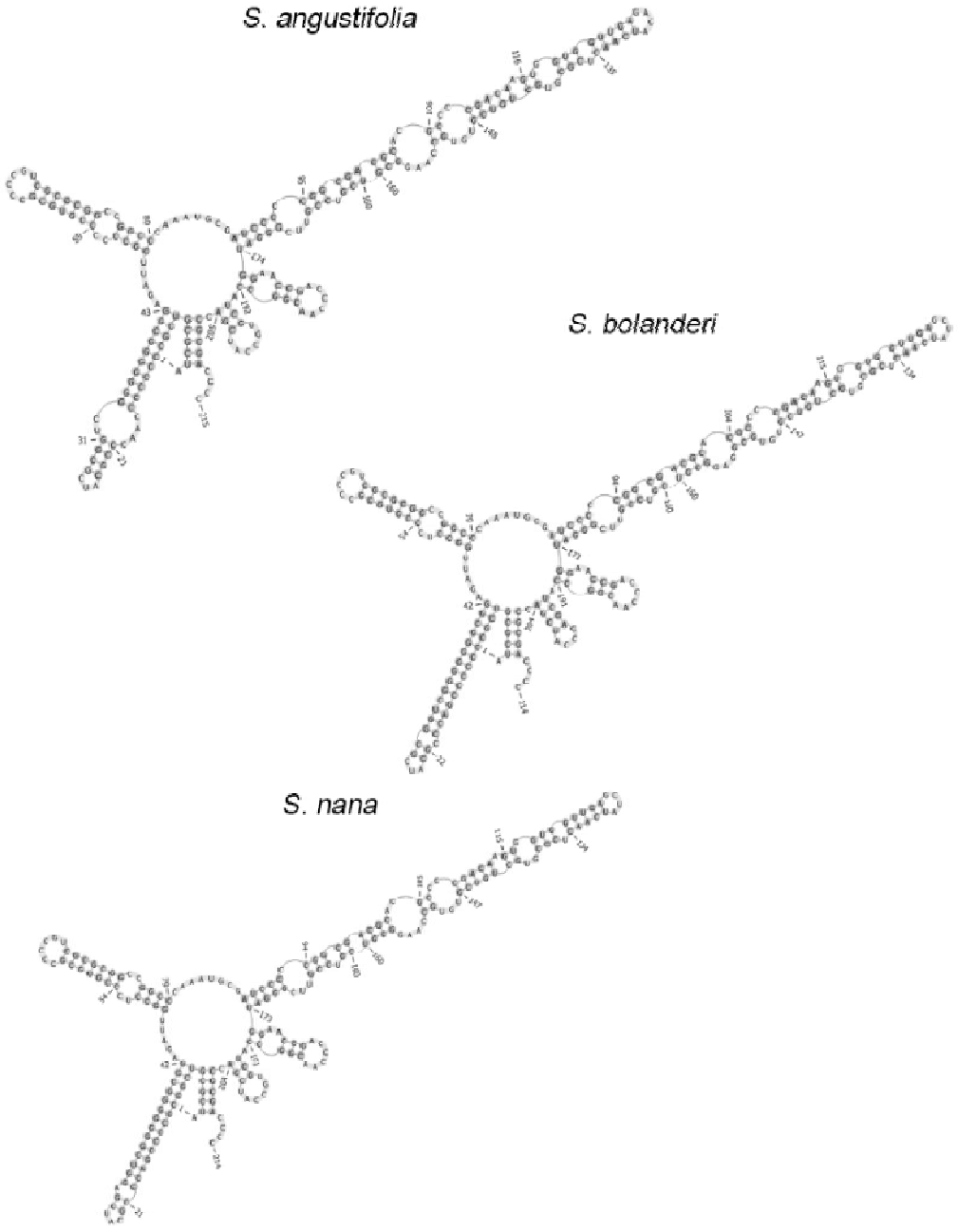
Appendix 19. Continued. Four taxa included in Type II.



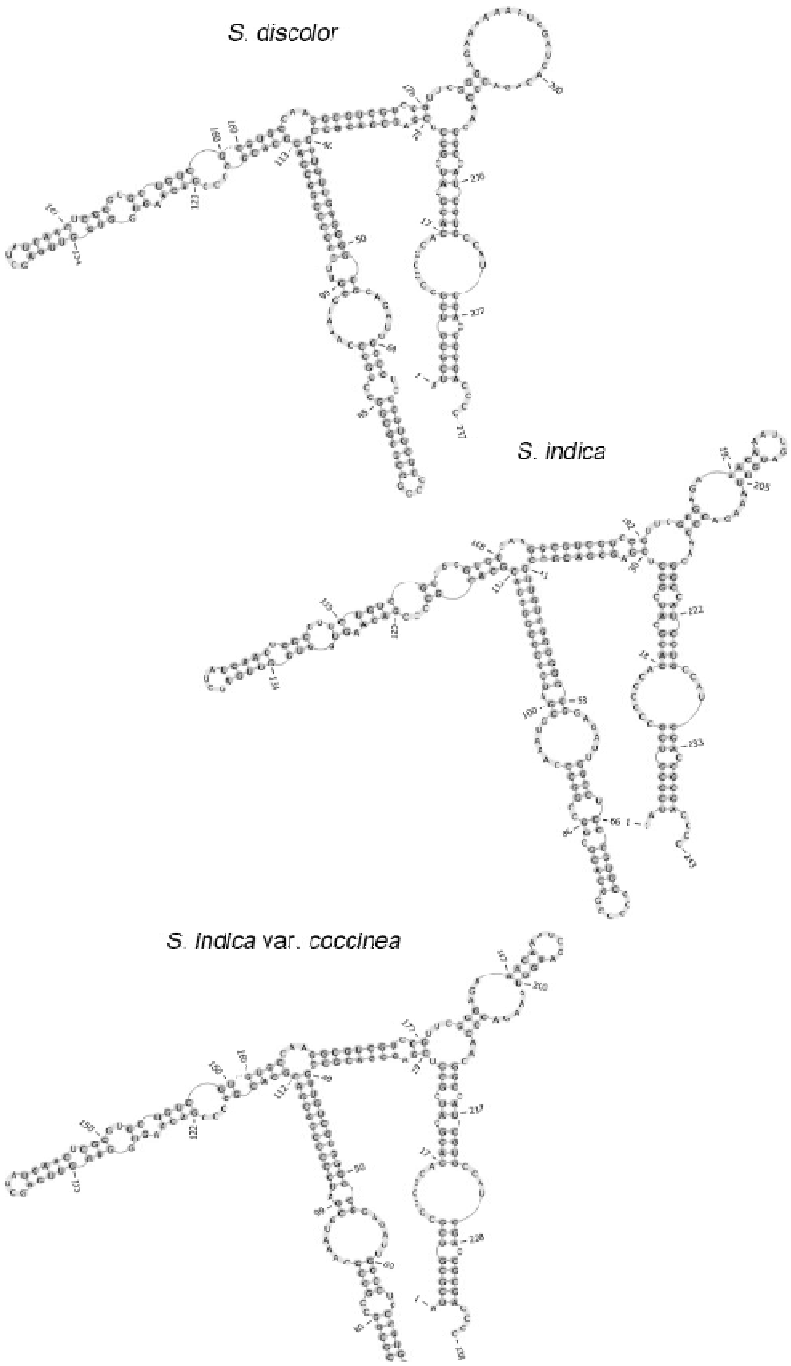
Appendix 19. Continued. Four taxa included in Type II.



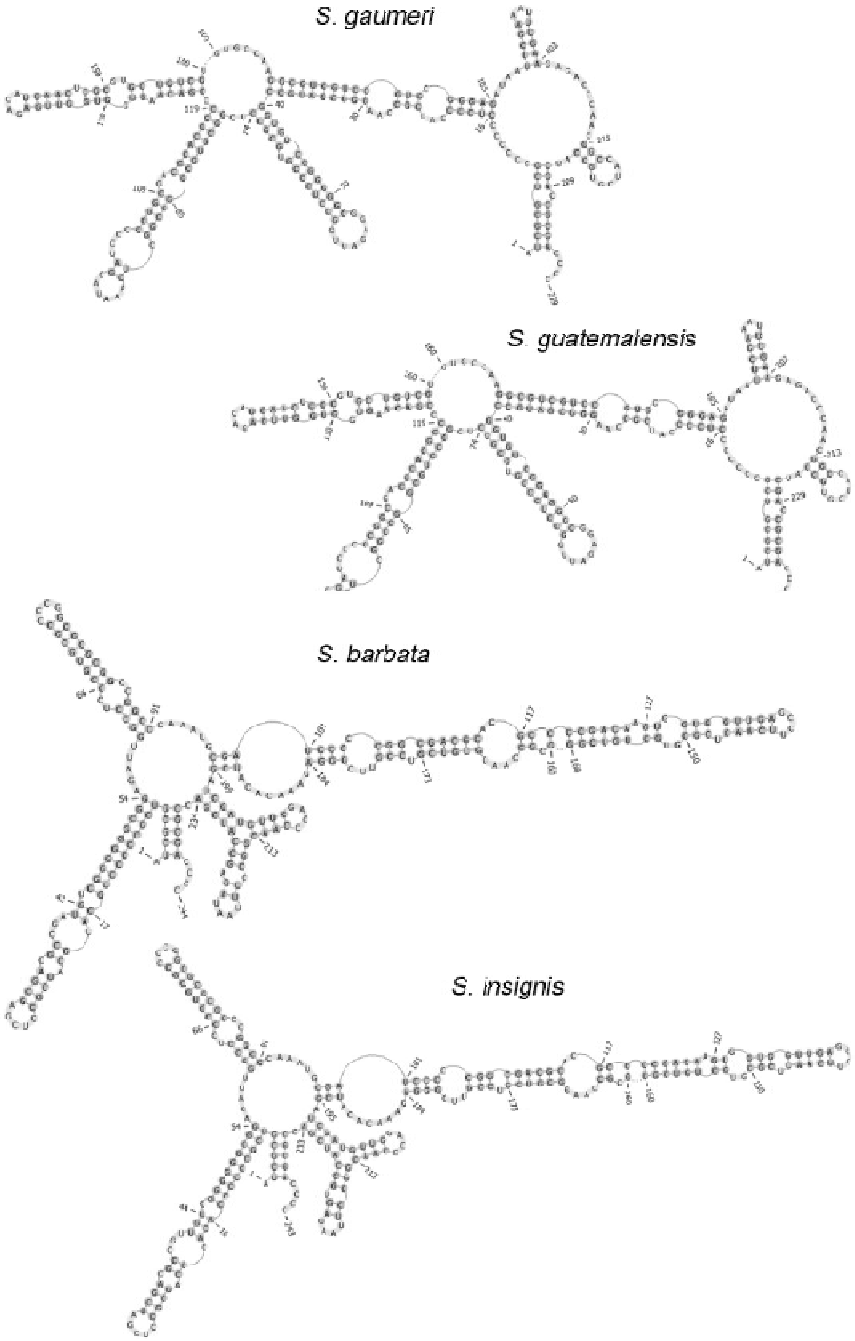
Appendix 19. Continued. Three taxa included in Type III.



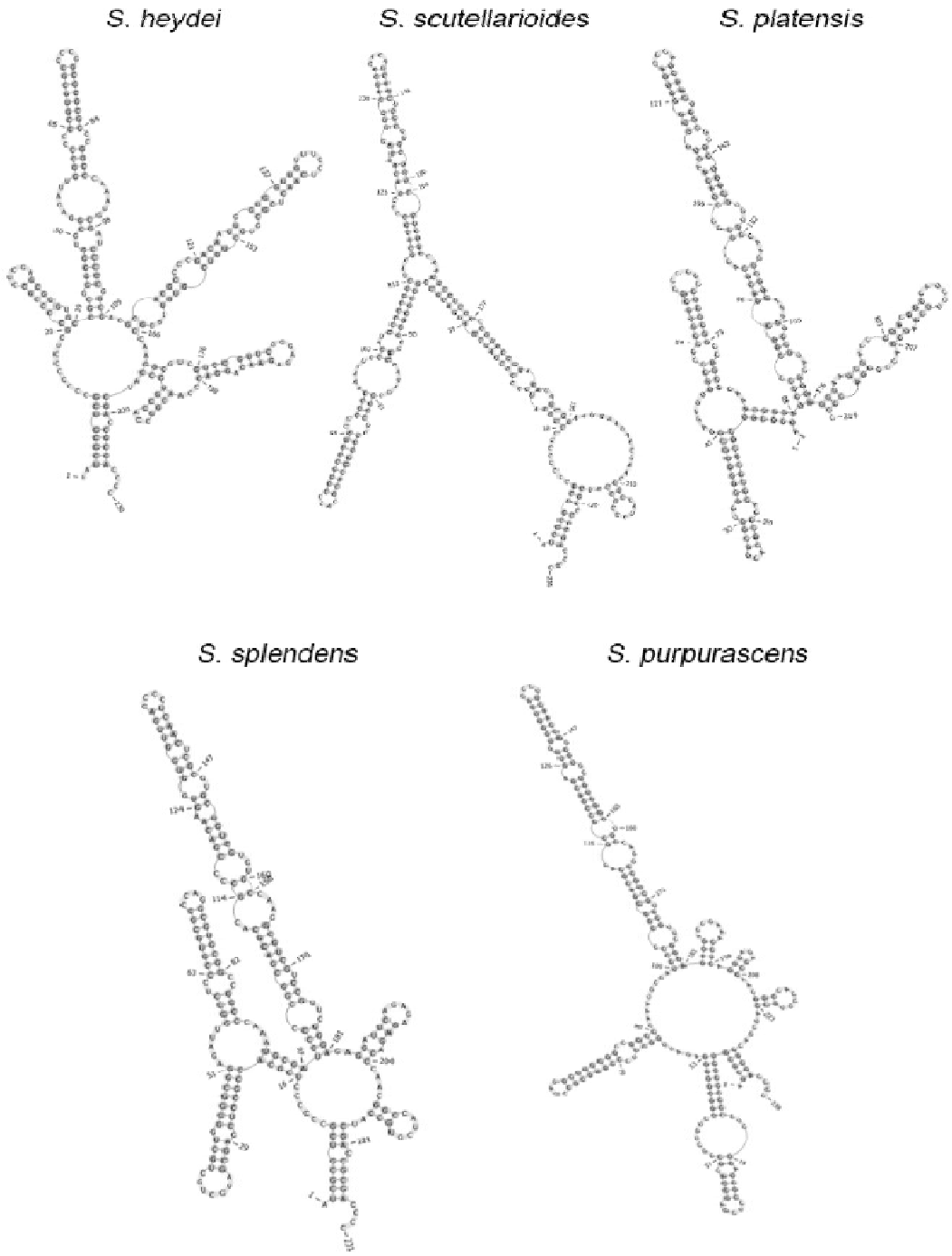
Appendix 19. Continued. Three taxa included in Type IV.



Appendix 19. Continued. Three taxa included in Type V.



Appendix 19. Continued. Four taxa included in Type VI (*S. gaumeri* and *S. guatemalensis*) and VII (*S. barbata* and *S. insignis*)



Appendix 19. Continued. Five taxa with different patterns.