CHAPTER 2

BACKGROUND

2.1 *Cassia*

2.1.1 The genus *Cassia* and the subtribe Cassiinae

The genus *Cassia* and other genera in the subtribe Cassiinae are members of the family Leguminosae. This family is the third largest flowering plant family after Orchidaceae and Asteraceae (Allen and Allen, 1981). Leguminosae was subgrouped by Bentham (1865) into three subfamilies according to their flower structures, i.e. Papillionoideae, Caesalpinioideae and Mimosoideae. The subfamily Papillionoideae has a very irregular flower type with five unequal and overlapped petals whereas flowers of the subfamily Caesalpinioideae are mostly irregular with five slightly unequal but not overlapped petals, usually in showy racemes or panicles. The last of subfamily Leguminosae, Mimosoideae, has regular flowers crowded into globose heads or cylindric spikes and all five petals in each flower are equal.

Among members of the subfamily Caesalpinioideae, the genus *Cassia* L. is the largest genus (Allen and Allen, 1981). This genus are either trees, shrubs or herbs and has over 600 species. Leaves are paripinnate with broad leaflets, some are sensitive to touch. Petal colours are attractively yellow, white, pink or orange. Pods are linear, flat or cylindric (Bentham, 1871 and Irwin & Barneby, 1976 in Allen and Allen, 1980).

The genus *Cassia* has long been recognised as a heterogenous genus because of its variations in shapes of pods, habits, or colouring petals. In 1871, Bentham (in Irwin & Barneby, 1981) divided this large genus into three subgenera:

Cassia, Senna and *Larsiorhegma*. A hundred and ten years later, Irwin and Barneby (1981) split the tribe Cassieae into five subtribes: Coratoniinae, Dialiinae, Dupaquetiinae, Labicheinea and Cassiinae. With their classification, the subtribe Cassiinae was further divided into three genera: *Cassia* L., *Senna* Miller and *Chamaecrista* Moench, using characteristics of filaments and the presence or absence of bracteoles. Three years later, Lock (1987) investigated African members of the subtribe Cassiinae using Irwin and Barneby's classification and re-classified the subtribe into three genera: *Cassia*, *Senna* and *Chamaecrista*. Tucker (1996) studied floral ontogeny of *Cassia* sensu stricto, *Senna* and *Chamaecrista* and confirmed that the three genera can be distinguished from each other by looking through their floral ontogeny.

Among these taxonomic arguments about generic classification in the subtribe Cassiinae, the relationship between the genera *Cassia* and *Senna* is the most confusing one and has raised the case for further experiments. For instance, Perveen and Qaiser (1998) investigated pollen morphology of five genera belonging to subfamily Caesalpinioideae from Pakistan. These five genera, with 11 species, are *Parkinsonia* (*P. aculaeata*), *Senna* (*S. tora, S. holosericea, S. italica* and *S. surattensis*), *Cassia* (*C. pumila* and *C. occidentalis*), *Bauhinia* (*B. variegata*), *Tamarindus* (*T. indica*) and *Caesalpinia* (*C. pulcherrima*). They found that those samples from the genus *Senna* and the genus *Cassia* were in the type III whose pollen grains are non parasyncopate, different apolcopium, recticulate or fossulate-foveolate or regulate tectum. Therefore, their conclusion about this were pollens of these *Cassia* and *Senna* species are relatively uniform.

Apart from flower and pollen morphological studies of the subtribe Cassiinae, Whitty *et al.* (1994) has also performed molecular study for the subtribe and supported Lock's separation (1987) with the three-genus system based on RAPD (Random Amplified Polymorphic DNA) method as a phonetic tool. After that, an even

more complete molecular phylogenetic analysis of the family Leguminosae was done by Käss and Wink in 1996 based on *rbc*L sequences of the chloroplast genome. Fourty-nine species included two *Cassia* species (*C. didymobotrya* Fresen and *C. senna* L.) were studied. The analysis showed that the subfamily Caesalpinioideae is in fact paraphyletic to the other subfamilies whereas the tribe Cassieae in which the genus *Cassia* was put appeared to be a monophyletic group.

In 1997, Doyle *et al.* reexamined a molecular phylogeny of Leguminosae using chloroplast *rbc*L gene sequences. Among 75 species used in this study, three species belong to the subtribe Cassiinae (*C. fistula* L., *Ch. fasciculata* (Michx) Greene and *S. alata* Roxb.). After mapping the 'nodulation' character to the obtained phylogeny, they discovered that the nodulated genus *Chamaecrista* was grouped as a sister clade with the non-nodulate genus *Senna*, and the nodulation character is a parallel gain for these taxa.

Two years later (1999), Ghareeb *et al.* studied systematics of some *Cassia* species with multiple approaches. Ten species were analysed on seed proteins, chromosome numbers and morphological characters. They revealed that the studied taxa should split into two groups. Group I included three *Cassia* species (*C. fistula, C. javanica* and *C. nodusa*) and was designated to be in a subgenus *Fistula* while *C. occidentalis, C. sophera, C. siamea, C. didymobotrya, C. italica, C. senna* and *C. surattensis* represented the group II which then was named as the subgenus *Senna*. Results obtained from this study reinforced the previous taxonomic treatment of the genus *Cassia,* i.e. that of Briton and Rose (1930).

Most recently, Mondal, Mondal and Mandal (2000) have investigated an interspecific variation among eight species of *Cassia* L. based on seed proteins and RFLP (Restriction Fragment Length Polymorphism) of mitochondrial DNA to understand phylogenetic relationships of these species. They grouped the eight species into two clusters : cluster I with *C. occidentalis, C. sophera, C. mimosoides*

and *C. tora* and the cluster II with *C. alata, C. siamea, C. fistula* and *C. renigera*. This grouping was similar to that of Ghareeb *et al.* (1999). In year 2001, Bruneau et. al. has investigated phylogenetic relationships in the Caesalpinioideae from its *trn*L intron sequences for 223 Caesalpinioideae, representing 112 genera. These included 3 genera of Cassiinae and 11 taxa, which are *C. grandis*, *Ch.* sp. 1, *Ch.* sp. 2, *S. alata*, *S. bacillaris*, *S. bauhinioides*, *S. crassiramea*, S. hirsuta, *S. lindheimeriana*, *S. occidentalis* and *S. wislizeni*. From this investigation, they found that genus *Chamaecrista* and genus *Senna* segregated from *Cassia* but they suggested subtribe Cassiinae is not monophyletic because genus *Chamaecrista* was being a basal of an unresolved clade B which composed of members of subfamily Mimosoideae, tribe Caesalpinieae and subtribe Cassiinae (tribe Cassieae; *Cassia* and *Senna*). The comparison between classification systems is in Table 1.

Table 1 comparison between generic classification systems of subtribe Cassiinae.

Year	Author	No. of proposed genus	Generic name	Classification Criteria	No. of studied taxa (*Cassia*, *Senna* or *Chamaecrista*)	No. of taxa also used in this thesis
1871	Bentham	1	*Cassia*	morphological characters	~ 600	17 *Cassia*
1981	Irwin & Barneby	3	*Cassia, Senna, Chamaecrista*	filaments; absence or presence of bracteoles	1023 (222 *Cassia*, 224 *Senna* and 577 *Chamaecrista*)	17 (4 *Cassia*, 11 *Senna* and 2 *Chamaecrista*)
1987	Lock	3	*Cassia, Senna, Chamaecrista*	filaments; absence or presence of bracteoles	87 (14 *Cassia*, 36 *Senna* and 37 *Chamaecrista*)	13 (*3 Cassia* and 10 Senna)

Table 1 continued

Year	Author	No. of proposed genus	Generic name	Classification Criteria	No. of studied taxa (*Cassia*, *Senna* or *Chamaecrista*)	No. of taxa also used in this thesis
1996	Käss & Wink	3	*Cassia*, *Senna*, *Chamaecrista*	*rbc*L sequence	2 (*Cassia didymobotrya* & *Cassia senna*)	-
1996	Tucker	3	*Cassia*, *Senna*, *Chamaecrista*	floral ontogeny	18 (1 *Cassia*, 15 *Senna* and 2 *Chamaecrista*)	5 (1 *Cassia* and 4 *Senna*)
1997	Doyle *et al.*	3	*Cassia*, *Senna*, *Chamaecrista*	*rbc*L sequence	3 (*C. fistula*, *Ch. fasciculata* and *S. alata*)	2 (*C. fistula* and *S. alata*)

Table 1 continued

Year	Author	No. of proposed genus	Generic name	Classification Criteria	No. of studied taxa (*Cassia*, *Senna* or *Chamaecrista*)	No. of taxa also used in this thesis
1998	Perveen & Qaiser	1	*Cassia* or *Senna*	pollen morphology	7 (*S. alaxandria*, *S. holosericia*, *S. italica*, *C. occidentalis*, *S. tora*, *C. pumila* and *S. surattensis*)	4 (*C. occidentalis*, *S. tora*, *C. pumila* and *S. surattensis*)
1999	Ghareeb *et al.*	1 with 2 groups	*Cassia*	seed protein; chromosome numbers; morphological characters	10 -**group I** *C. fistula*, *C. javanica* and *C. nodusa* -**group II** *C. occidentalis*, *C. sophera*, *C. siamea*, *C. didymobotrya*, *C. italica*, *C. senna* and *C. surattensis*	5 -**group I** *C. fistula* and *C. javanica* -**group II** *C. occidentalis*, *C. sophera*, *C. siamea* and *C. surattensis*

Table 1 continued

Year	Authors	No. of genera	Generic name	Characteristics	No. of taxa (*Cassia*, *Senna* or *Chamaecrista*)	Taxa which as same as this experiment
2000	Mondal *et al.*	1 with 2 groups	*Cassia*	seed protein; mtDNA RFLP	8 -**group I** *C. occidentalis*, *C. sophera*, *C. mimosoides* and *C. tora* -**group II** *C. alata*, *C. siamea*, *C. fistula* and *C. renigera*	5 -**group I** *C. occidentalis*, *C. sophera* and *C. tora* -**group II** *C. alata* and *C. fistula*
2001	Bruneau *et al.*	3	*Cassia*, *Senna*, *Chamaecrista*	*trn*L intron sequences	11 (1 *Cassia*, 8 *Senna* and 2 *Chamaecrista*)	4 (*C. grandis*, *S. alata*, *S. hirsuta* and *S. occidentalis*)

2.1.2 *Cassia* in Thailand and its taxonomy

In Thailand, habits of members of the genus *Cassia* are various from trees, shrubs and herbs. Their leaves are paripinnate, often with folia glands. Flowers are simple racemes or panicles, bisexual, yellow or pink to red. Receptacles are very short while sepals are imbricate in buds. Petals are five with 10 stamens. Pods are varying in shape from flat to cylindric, with indehiscent or dehiscent. Twenty-two species with four subspecies were found from all over Thailand.

In 1984, Larsen *et al.* suggested in the Flora of Thailand vol. 4 part I to place all of these members of the subtribe Cassiinae found in Thailand into the genus *Cassia* only. However, their later work in Flora Malesiana series I vol. 12 (Larsen *et al.*, 1996) has made a new recommendation for the genus. Following Irwin & Barneby's concepts in their African Cassiinae monograph, 12 species in the genus *Cassia* have been moved to the genus *Senna*. These species were *Cassia* (*Senna*) *alata*, *C.*(*S.*) *spectabilis*, *C.*(*S.*) *siamea*, *C.*(*S.*) *timoriensis*, *C.*(*S.*) *hirsuta*, *C.*(*S.*) *occidentalis*, *C.*(*S.*) *sophera*, *C.*(*S.*) *fruticosa*, *C.*(*S.*) *surattensis*, *C.*(*S.*) *bicapsularis*, *C.*(*S.*) *tora* and *C.*(*S.*) *obtusifolia*. Not only *Cassia* to *Senna* redesignation, but four other *Cassia* species have also renamed as the genus *Chamaecrista*, i.e. *Cassia* (*Chamaecrista*) *absus*, *C.(Ch.) pumila*, *C.(Ch.) mimosoides* and *C.(Ch.) leschenaultiana*. Moreover, *C. agnes* was also suggested to become a subspecies of *C. javanica* as *C. javanica* is in fact a very polymorphous species with a wide distribution range.

Recently, the genus *Cassia* in Thailand have been particularly investigated by Kidyue (2002) and Pechsri (2003). They collected samples of *Cassia* from all over the country following previous reports of the Flora of Thailand. Only 17 of 22 species could be collected because of changing environments. Some species rarely present, for instance, *C. absus* only found in Khao Tao at Hua Hin. Specimens of *C.*

obtusifolia, a new recorded species for Thailand was also found during sample collections.

In his study, Kidyue (2002) investigated the comparative anatomy of stems, leaves and flowers of 17 *Cassia* sensu lato grew in Thailand. He finally suggested to divide the genus *Cassia* into 4 groups: *Cassia*, *Senna*-1 (tree), *Senna*-2 (shrubs) and *Chamaecrista*. A numerical taxonomy of *Cassia* sensu lato in Thailand has also been investigated by Pechsri in his 2003 work. He separated *Cassia* sensu lato into three groups as same as the classification proposed earlier by Irwin and Barneby using overall canonical discriminant analyses. The three groups are *Cassia* group, *Senna* group and *Chamaecrista* group. However, a 4-group classification (*Cassia* group (included *C. spectabilis*), *Senna* group, *Senna alata* group and *Chamaecrista* group) was also suggested if cluster analyses were performed.

Last but not least, in a cytogenetic point of view, Thai *Cassia* species are very much the same. Umpunjuntara (1990) has investigated 10 *Cassia* species in Chulalongkorn University campus and found that nine out of 10 species (*C. alata, C. bakeriana, C. biflora, C. fistula, C. garrettiana, C. grandis, C. siamea, C. sophera* and *C. spectabilis*) have the same chromosome number equally to 28 (2n=28), while that of *C. surattensis* is 56. This cytogenetic investigation proposed that the basic number of the genus should be seven (x=7) and these nine *Cassia* species are actually allotetraploid while *C. surattensis* with 2n=56 is allooctaploid.

2.1.3 Descriptions and collecting localities of *Cassia* species found in Thailand (Pechsri, 2003 and Larsen *et al.*, 1984) (Fig.1 – Fig. 12, Table 2)

1. *Cassia fistula* L. Sp. Pl. : 377. 1753; Craib in Fl. Siam. En. 1: 509. 1928: K. & S.S. Larsen and Vidal in Fl. C.L.V. 18: 79. 1980; K. & S.S. Larsen and Vidal in Fl. Thailand 4(1): 103. 1984.

Tree, glabrous. *Leaves* large: unpinned; stipules small, caducous, deltoid; pinnate 4-6 pairs, petioles glabrous, rachis teeter; pinnate ovate-oblong, glabrous, apex acute, base cineaste. *Racemes* axillary, pendent; bract caducous *Flower* zygomorphic, pedicel glabrous; sepals ovate-elliptic, petals yellow, ovate, short-clawed; stamens 10, 3 long, 4 shorter; 3 reduced in size with minute anthers; ovary strigulose, style velutinous. *Pods* terete, glabrous, black in colour

Vernacular: Ratchaphruk (ราชพฤกษ์); Khun (คูณ)

Specimen Examined: *S. Pechsri 51*

2. *Cassia garrettiana* Craib, Kew Bull. 1912: 151; Craib in Fl. Siam. En. 2: 510. 1928; K. & S.S. Larsen and Vidal in Fl. C.L.V. 18: 91. 1980; K. & S.S. Larsen and Vidal in Fl. Thailand 4(1): 112. 1984.

Tree. *Leaves* unipinnate; stipules caducous; pinnae 4-11 pairs, petioles 2-3 cm; pinnae lanceolate to ovate, glabrous, apex acuminate, base rounded. *Racemes* leafly on terminal; bracts ovate caducous. *Flower* zygomorphic, pedicels pubescent; sepals elliptic, petals yellow, ovate, short-clawed; stamens 10, filament fattened in the two largest, 5 shorter; 3 reduced in size with minute anthers; ovary and style glabrous. *Pods* flat, glabrous, black in colour

Vernacular: Samae san (แสมสาร); Khi lek khan chang (ขี้เหล็กคันช่าง)

Specimen Examined: *S. Pechsri 59*

Fig. 1 flower characteristics of *C. fistula* (photographed by K. Srisawat)

Fig. 2 flower characteristics of *C. garrettiana* (photographed by K. Srisawat)

3. *Cassia bakeriana* Craib., Kew Bull. 1911: 45; Craib in Fl. Siam. En. 1: 508. 1928; K. & S.S. Larsen and Vidal in Fl. Thailand 4(1): 105. 1984.

Tree, densely hairy on all young parts. *Leaves* unipinnate; stipules lanceolate, attached in the middle; pinnae 7-11 pairs, petioles pubescent, rachis pubescent; pinnae oblong-oblanceolate, hairy, apex round with a small sharp point and base rounded. *Racemes* lateral; bracts lanceolate, apex long-pointed, hairy. *Flower* zygomorphic, pedicels pubescent; sepals ovate-lanceolate, pubescent, petals pinkish, ovate-lanceolate, short-clawed; stamens 10, filaments swollen in the middle in the three largest, 4 shorter; 4 reduced; ovary pubescent, style short. *Pods* terete, pubescent.

Vernacular: Chaiyaphruk (ชัยพฤกษ์); Kalapaphruk (กัลปพฤกษ์)

Specimen Examined: *S. Pechsri 53*

4. *Cassia alata* L., Sp. Pl.: 378. 1753; Craib in Fl. Siam. En. 1: 508. 1928; K. & S.S. Larsen and Vidal in Fl. C.L.V. 18: 86. 1980; K. & S.S. Larsen and Vidal in Fl. Thailand 4(1): 108. 1984.

Shrub, hairy. *Leaves* unipinnate; stipules auriculate, persistent, deltoid; pinnae 7-13 pairs, petioles 2-3 cm; pinnae elliptic-oblong, glabrous, apex and base rounded. *Racemes* axillary, densed; bracts caducous. *Flower* zygomorphic; sepals oblong, petals bright yellow, ovate orbicular to spathulate, short-clawed; stamens 10, filaments thick in the two largest, 4 shorter; 4 reduced in size with minute anthers; ovary and style glabrous. *Pods* winged, thick, glabrous, black in colour

Vernacular: Chum het thet (ชุมเห็ดเทศ); Khi khak (ขี้คาก)

Specimen Examined: *S. Pechsri 55*

Fig. 3 flower characteristics of *C. bakeriana*

(http://home.hiroshima-u.ac.jp/shoyaku/ThaiPP.htm)

Fig. 4 flower characteristics of *C. alata*

(http://yucca.standardout.com/pics/pdb_OlgaN_2003-04-01_1049224322574.jpg)

5. *Cassia grandis* L. f. suppl.: 230. 1718; K. & S.S. Larsen and Vidal in Fl. C.L.V. 18: 80. 1980; K. & S.S. Larsen and Vidal in Fl. Thailand 4(1): 105. 1984.

Deciduous tree trunk with buttress. *Leaves* unipinnate stipules minute; glabrous, apex and base rounded. *Racemes* lateral; bracts caducous. *Flower* zygomorphic; sepals obovate-rounded, pubescent, reflexed, petals first red later pink, finally orange, obovate, short-clawed; stamen 10, filaments recurved in the three largest, 5 shorter; 2 reduced in size with minute anthers; ovary silky tomentose, style short. *Pods* cylindric, woody, rugose, glabrous, black in colour

Vernacular: Kalaphruk (กาฬพฤกษ์)

Specimen Examined: BKF. No. 120511

6. *Cassia javanica* L. var. ***javanica*** Sp. Pl.: 379. 1753, Craib in Fl. Siam. En. 1: 508, 509, 511. 1928; K. & S.S. Larsen and Vidal in Fl. C.L.V. 18: 84. 1980: K. & S.S. Larsen and Vidal in Fl. Thailand 4(1): 107. 1984.

Deciduous tree. *Leaves* unipinnate; stipules elliptic, falcate to pointed, attached in the middle; pinnae 7-16 pairs, petioles glabrous; pinnae elliptic-ovate to oblong, hairy, apex and base rounded. *Racemes* lateral, densed; bracts ovate-acute *Flower* zygomorphic; sepals ovate-acute, dark red to reddish brown, petals first pink later dark red, finally pale, ovate, long-clawed; stamens 10, filaments recurved with a spherical enlargement near the middle in the three largest, 4 shorter; 3 reduced in size with minute anthers; ovary pubescent, style short. *Pods* terete glabrous, black in colour.

Vernacular: Kalapaphruk (กัลปพฤกษ์); Chaiyaphruk (ชัยพฤกษ์)

Specimen Examined: *S. Pechsri 54*

Fig. 5 flower characteristics of *C. grandis*

(http://home.hiroshima-u.ac.jp/shoyakuThaiPP.htm)

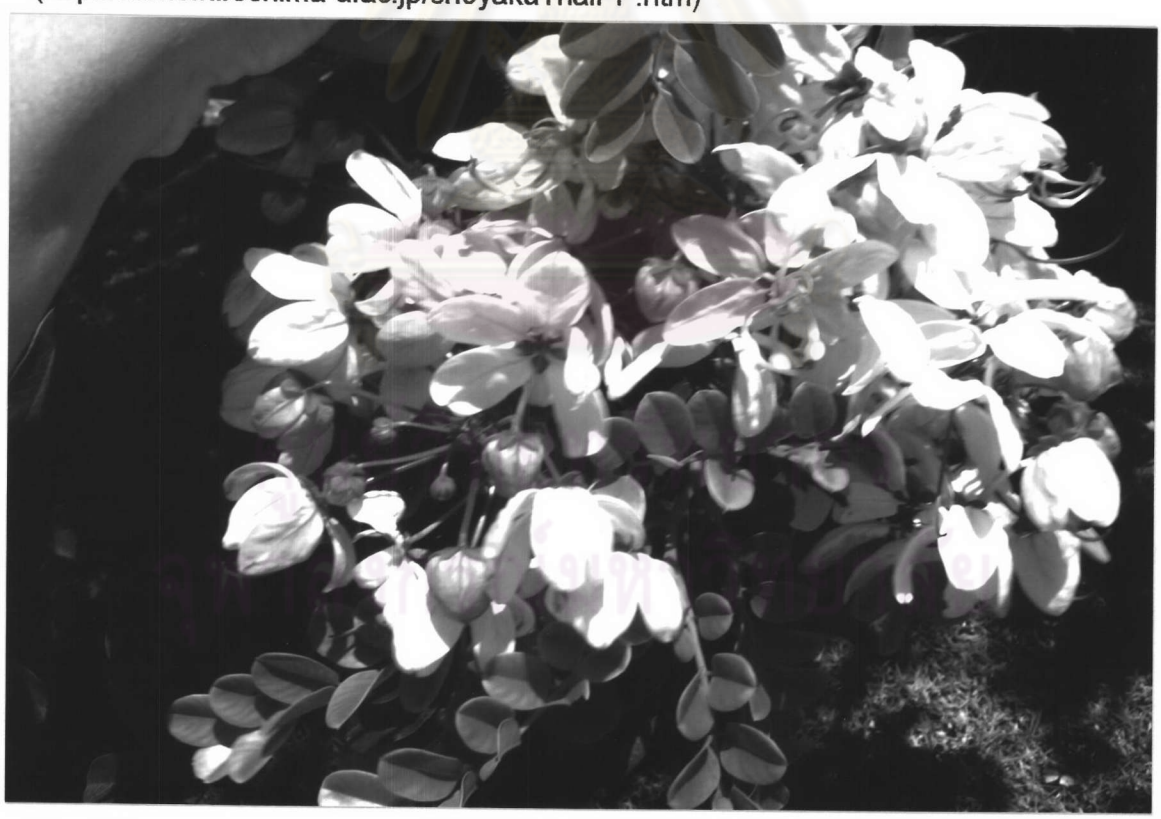

Fig. 6 flower characteristics of *C. javanica* (http://www.hear.org/starr/hiplants/images/

hires/ starr_030702_0027_cassia_javanica.jpg)

7. *Cassia hirsuta* L., Sp. Pl.: 378 1753: K. & S.S. Larsen and Vidal in Fl. C.L.V. 18: 92 1980; K. & S.S. Larsen and Vidal in Fl. Thailand 4(1): 113. 1984.

Herb or undershrub hirsute. *Leaves* unipinnate stipules caducous, hairy; pinnae 3-5 pairs, the upper pairs largest, petioles long villous, sessile; pinnate lanceolate, hirsute, apex acute, base rounded. *Racemes* short, axillary; bracts hirsute. *Flower* zygomorphic, pedicels pubescent; sepals pubescent, petals yellow obovate, glabrous, short-clawed; stamens 10, filaments flat in the two largest, 4 shorter; 4 reduced in size with minute anthers; ovary grayish wooly, style glabrous. *Pod* falcate, hirsuta.

Vernacular: Rang jued ton (รางจืดต้น); Phong pheng (โผงเผง); Dap phit (ดับพิษ)

Specimen Examined: *S. Pechsri 60*

8. *Cassia spectabilis* DC., Cat. Hort. Monsp.: 90. 1813; K. & S.S. Larsen and Vidal in Fl. Thailand 4(1): 110. 1984.

Tree, hairy on young parts. *Leaves* unipinnate; stipules linear, falcate, caducous; pinnae 9-15 pairs, petioles 2-3 cm; pinnae lanceolate, glabrous, apex acute, mucronate, base rounded. *Racemes* large, leafy on terminal; bracts ovate, caducous. *Flower* zygomorphic, pedicels valentinous; sepals unequal, oblong, petals yellow, ovate to spathulate, the lower one larger broad falcate, short-clawed; stamens 10, 7 large, 4 shorter; 3 reduced in size with reuniform minute anthers; ovary and style glabrous. *Pods* terete, glossy, glabrous, black in colour.

Vernacular: Suwanaphruk (สุวรรณพฤกษ์); Khi lek American (ขี้เหล็กอเมริกัน)

Specimen Examined: *S. Pechsri 66*

Fig. 7 flower characteristics of *C. hirsuta*

(http://fireflyforest.com/flowers/yellows/yellow20.html)

Fig. 8 flower characteristics of *C. spectabilis*

(http://www.nparks.gov.sg/nursery/uploadfiles/sennaspectabilis02flofruit.jpg)

9. *Cassia sophera* L., Sp. Pl.: 379. 1753; Craib in Fl. Siam. En. 1: 513.1928; K. & S.S. Larsen and Vidal in Fl. C.L.V. 18: 94. 1980; K. & S.S. Larsen and Vidal in Fl. Thailand 4(1): 115. 1984.

Shrub, glabrous. *Leaves* unipinnate; stipules ovate, caducous; pinnae 7-13 pairs, petioles with gland above the petiole joint; pinnae lanceolate, the upper largest, glabrous, apex acute, base rounded. *Racemes* axillary; bracts ovate, caducous. *Flower* zygomorphic; sepals ovate-rounded, petals yellow obovate, short-clawed; stamens 10, 2 longer, 4 shorter; 4 reduced in size with minute anthers; ovary pubescent, style glabrous. *Pods* cylindric glabrous, brown in colour

Vernacular: Phak khet (ผักเค็ด); Phak wan ban (ผักหวานบ้าน)

Specimen Examined: *S. Pechsri 56*

10. *Cassia surattensis* Burm f. K. & S.S. Larsen, Fl. C.L.V.18: 102. - *C. glauca* Lamk. Craib in Fl. Siam. En. 1: 510. 1928; K. & S.S. Larsen and Vidal in Fl. C.L.V. 18: 102. 1980; K. & S.S. Larsen and Vidal in Fl. Thailand 4(1): 120. 1984.

Shrub, puberulous. *Leaves* unipinnate; linear falcate subpersistent; pinnae 7-9 pairs, petioles 1.5-3 cm; rachis with gland between the 2-3 lower pairs of leaflets, pinnae ovate-oblong, glabrous, apex and base rounded. *Racemes* axillary, densed; bracts ovate-acute. *Flower* zygomorphic; sepals ovate, petals yellow, obovate, short narrow clawed; stamens 10, filaments thick; ovary puberulous, filiform recurved, style glabrous. *Pods* flat, glabrous, dehiscent.

Vernacular: Song badan (ทรงบาดาล)

Specimen Examined: *S. Pechsri 79*

Fig. 9 flower characteristics of *C. sophera* (photographed by K. Srisawat)

Fig. 10 flower characteristics and of *C. surattensis* (photographed by K. Srisawat)

11. *Cassia occidentalis* L. Sp. Pl.: 378 1753; Craib in Fl. Siam. En. 2: 512.1928; K. & S.S. Larsen and Vidal in Fl. C.L.V. 18: 93. 1980; K. & S.S. Larsen and Vidal in Fl. Thailand 4(1): 113. 1984

Undershrub, glabrous. *Leaves* unipinnate; linear to acute; pinnae 4-5 pairs, petioles with large gland above the petiole joint; pinnae unequal-side, ovate to oblong; apex acuminate, base rounded. *Racemes* axillary, densed; bracts linearacute, caducous. *Flower* zygomorphic; sepals ovate, petals yellow with violet veins, ovate, short-clawed; stamens 10, 2 longer, 4 shorter; 4 reduced with minute anthers; ovary tomentose, style glabrous; *Pods* flat, glabrous, brown in colour.

Vernacular: Phak khet (ผักเค็ด); Chumhet lek (ชุมเห็ดเล็ก)

Specimen Examined: *S. Pechsri 67*

12. *Cassia tora* L., Sp. Pl.: 376. 1753: Craib in Fl. Siam. En. 1: 514. 1928: K. & S.S. Larsen and Vidal in Fl. C.L.V. 18: 96. 1980; K. & S.S. Larsen and Vidal in Fl. Thailand 4(1): 117. 1984.

Undershrub, hairy. *Leaves* unipinnate; stipules setaceous, caducous; pinnae 3 pairs, petioles 1-4 cm; rachis with gland between the 2 lower pairs of leaflets, pinnae obovate; apex rounded, base cuneate. *Racemes* axillary densed; bracts linear-acute. *Flower* zygomorphic; sepals ovate, petals yellow obovate, short-clawed; stamens 7, 2 largest, 5 shorter; 3 staminode; ovary pubescent, style glabrous. *Pods* terete.

Vernacular: Chumhet Thai (ชุมเห็ดไทย); Chumhet na (ชุมเห็ดนา)

Specimen Examined: *S. Pechsri 77*

Fig. 11 flower characteristics of *C. occidentalis*

(http://botany.cs.tamu.edu/FLORA/perdeck/afr_016.jpg)

Fig. 12 flower characteristics of *C. tora*

(http://www.tcp-ip.or.jp/~jswc3242/000/ebisugusa.jpg)

13. *Cassia timoriensis* DC., Prod. 2: 499. 1825; Craib in Fl. Siam. En. 1: 514.1928; K. & S.S. Larsen and Vidal in Fl. C.L.V. 18: 88.1980; Larsen and Vidal in Fl. Thailand 4(1): 111. 1984.

Tree with golden hairy throughout. *Leaves* unipinnate; stipules large, auriculate; pinnae 16-22 pairs, petioles 2-3 cm; rachis pubescent; pinnae oblong; apex subacute to mucronate, base rounded; golden pubescent. *Racemes* axillary; pedicels pubescent; bracts ovate, caducous. *Flower* zygomorphic; sepals oblongovate with rounded apex, yellowish pubescent, petals yellow, ovate, short-clawed; stamens 10, 2 largest, 5 shorter; 3 reduced in size with minute anthers; ovary and style glabrous. *Pods* flat, glabrous, dehiscent, brown in colour.

Vernacular: Khi lek luat (ขี้เหล็กเลือด); Khi lek khan chang (ขี้เหล็กคันช่ง)

Specimen Examined: *S. Pechsri 77*

14. *Cassia siamea* Lamk., Enc. 1: 648. 1785; Craib in Fl. Siam. En. 1: 513.1928; K. & S.S. Larsen and Vidal in Fl. C.L.V. 18: 887. 1980: K. & S.S. Larsen and Vidal in Fl. Thailand 4(1): 110. 1984.

Tree, pubescent on young branches. *Leaves* unipinnate; stipules minute, caducous; pinnae 7-11 pairs, petioles 2-3 cm; pinnae ovate-oblong, glabrous, apex and base rounded. *Racemes* terminal, large; bracts obovate with long acute apex. *Flower* zygornorphic pedicels valentinous; sepals thick, oblong, petals yellow, broadly ovate, short-clawed; stamens 10 filaments straight in the two largest, 4-5 shorter; 3 reduced in size with minute anthers; ovary pubescent, style glabrous. *Pods* flat, glabrescent, longitudinally waved with raised sutures

Vernacular: Khi lek ban (ขี้เหล็กบ้าน); Khi lek (ขี้เหล็ก)

Specimen Examined: *S. Pechsri 78*

15. *Cassia obtusifolia* L. Sp. Pl.: 378. 1753.

Herb or undershrub, thinly pubescent. *Leaves* unipinnate; stipules caducous; pinnae 3 pairs, petioles 1-4 cm; rachis with 2 subulate gland between the lowermost pair of leaflets; pinnae obovate, glabrous, apex rounded, base acuminate. *Racemes* axillary; bracts linear. *Flower* zygomorphic; sepals ovate, petals orange-yellow, obovate, short-clawed; stamens 7, 3 longer 4 shorter; 4 staminode; ovary pubescent; style glabrous. *Pods* linear, terete, falcate, glabrous.

Vernacular: Chumhet Thai (ชุมเห็ดไทย)

Specimen Examined: *S. Pechsri 61*

16. *Cassia leschenaultiana* DC., Mem. Soc. Phys. Geneve 2: 132. 1824; Craib in Fl. Siam. En. 1: 511. 1928; K. & S.S. Larsen and Vidal in Fl. C.L.V. 18: 106. 1980; K. & S.S. Larsen and Vidal in Fl. Thailand 4(1): 123. 1984.

Small shrub, densely greyish to yellowish pubescent. *Leaves* unipinnate; stipules linear, persistent; pinnae 35-47 pairs, petioles with discoid gland below the lowest pair of leaflets; pinnae falciform, side unequal, glabrous, apex and base rounded rachis pubescent, canaculate. *Racemes* few, axillary bracts caducous. *Flower* zygornorphic, pedicels pubescent; sepals oblong, yellow, shortly; stamens 9-10, filaments very short; ovary hairy, style recurved. *Pods* flat, dehiscent.

Vernacular: Sa kham khom (ซ่าขามค่อม)

Specimen Examined: *S. Pechsri 50*

17. *Cassia pumila* Lamk. Enc. 1: 651. 1785; Craib in Fl. Siam. En. 1: 513. 1928; K. & S.S. Larsen and Vidal in Fl. C.L.V. 18: 104. 1980; K. & S.S. Larsen and Vidal in Fl. Thailand 4(1): 120. 1984.

Small shrub, pubescent. *Leaves* unipinnate; stipules linear acute, persistent; rachis grooved in side; pinnate 13-17 pairs; petioles pubescent with a long stipitate gland below the lowest pair of leaflets; pinnae narrow elliptic, sessile, hairy along the midrib, upper glabrous, lower pubescent, apex and base rounded. *Racemes* axillary; bracts as the stipule but shorter. *Flower* zygomorphic, pedicels pubescent; sepals lanceolate; petals bright yellow, oblong-obovate, short-clawed; stamens 5-6 ovary tomentose, style glabrous. *Pods* flat, dehiscent, brown in colour

Vernacular: Makham din (มะขามดิน); Makham bia (มะขามเบี้ย)

Specimen Examined: *S. Pechsri 57*

Table 2 collecting localities of *Cassia* in Thailand.

Taxa	Collecting localities
C. alata	Mae Hong Son, Chiang Mai, Song Khla, Ranong, Surat Thani, Yala
C. bakeriana	Chiang Mai, Saraburi, Chiang Rai, Nakhon Ratchasima
C. fistula	Chiang Rai, Kanchanaburi, Phitsanulok, Lampang, Mae Hong Son, Loei, Nakhon Ratchasima, Chiang Mai, Trat, Chon Buri, Phrae, Bangkok
C. garrettiana	Loei, Phrae, Saraburi, Chiang Mai, Chaiyaphum, Chon Buri, Lampang, Buri Ram, Tak, Kanchanaburi, Phetchabun, Kamphaeng phet, Ratchaburi, Khon Kaen
C. grandis	Saraburi, Nakhon Ratchasima, Phetchaburi, Bangkok
C. hirsuta	Kanchanaburi, Mae Hong Son, Chiang Mai, Chumphon
C. javanica	Peninsula, Narathiwat, Kanchanaburi, Chumphon, Loei, Chiang Mai
C. obtusifolia	Bangkok
C. occidentalis	Rayong, Mae Hong Son, Loei, Chiang Rai, Chai Nat, Nakhon Sawan, Chiang Mai, Tak, Kanchanaburi, Nakhon Ratchasima, Prachuap Khiri khan, Songkhla
C. siamea	Prachuap Khiri Kan, Surat Thani, Trat, Nakhon Sawan, Nakhon Si Thammarat, Chiang Mai, Nakhon Nayok, Phrae, Songkhla, Tak, Kalasin, Suphan Buri
C. sophera	Kanchanaburi, Phrae, Kalasin, Phang-nga, Ranong
C. spectabilis	Bangkok, Chiang Mai, Songkhla, Khon Khaen, Phitsanulok
C. surattensis	Phrae, Bangkok, Chiang Mai, Loei

Table 2 (continue)

Taxa	Collecting localities
C. tora	Buri Rum, Chiang Mai, Nakhon Nayok, Loei, Khon Khaen, Phitsanulok, Trat, Yala, Mae Hong Son, Nakhon Ratchasima, Sakon Nakhon
C. timoriensis	Nakhon Ratchasima, Tak, Satun, Lampang, Loei, Saraburi, Khon Khaen, Phrae, Chanthaburi, Kanchanaburi, Nakhon Nayok, Uthai thani, Lop Buri, Yala, Ranong, Mae Hong Son, Surat Thani, Rayong
C. leschenaultiana	Chiang Mai, Loei, Kanchanaburi, Chaiyaphum
C. bicapsularis	Chon buri
C. fruticosa	Saraburi, Bangkok
C. pumila	Phitsanulok, Phang-nga, Kanchanaburi, Chiang Mai, Lampang
C. agnes	Phrae, Surat Thani, Kamphaeng Phet, Trang, Chanthaburi
C. mimosoides	Loei, Kanchanaburi, Satun, Chon Buri, Si Sa Ket, Chiang Mai, Lamphun, Nakhon Ratchasima, Chaiyaphum, Surin
C. absus	Hua Hin (only once collected)

2.2 Molecular phylogenetics

2.2.1 Definition and principle of molecular phylogenetics

In 1950 a German entomologist Willi Hennig published a book named *Grundzüge einer Theorie der Phylogenetischen Systematik*. Hennig' s book contained five basic ideas which began a major revolution in systematics:

1. The relationships lending to the cohesion of living and extinct organisms are genealogical ("blood") relationships.

2. Such relationships exist for individuals within populations, between populations, and between species.

3. All other types of relationship (i.e., phenotypic and genetic) are phenomena correlated with genealogical descent and thus are best understood within the context of genealogical descent with modification (quite literally 'evolution").

4. The genealogical relationships among populations and species may be recovered (discovered) by searching for particular characters which document these relationships.

5. The best general classification of organisms is one that exactly reflects the genealogical relationships among these organisms.

Hennig's ideas were first discussed in an American journal by Kiriakoff (1959). However, wide discussion of phylogenetic methods in English-language came after the publication of Hennig's revised book (1966) and Brundin's (1966) work on chironomid midges. Other early applications of Hennig's methods include those of Koponen (1968: mosses) and Nelson (1969: fishes). More recent discussions and applications of Hennig's methods applied to plants came from works of Bremer (1967, 1978), Bremer and Wanntorp (1978) and Humphries (1979). In fact, Hennig proposed many ideas other than the five basic points listed above. Some of

these ideas are still used (e.g. monoplyly) while others have been discarded. However, these five basic ideas with some modifications continue to provide the major theoretical concepts for 'phylogenetic systematics'.

Phylogenetic systematics or simply 'phylogenetics' is the systematic approach that accomplishes those tasks above. The phylogenetic system can give baseline data from which other comparative biology disciplines can begin their investigations. Phy1ogenetic systematic approaches attempt to recover phylogenetic (genealogical) relationship among groups of organisms and produces classifications that exactly reflect those genealogical relationships. By this concept, taxonomy therefore could be an alternative of phylogenetic systematics because it comprises the theory and practice of describing a diversity of organisms and ordering this diversity into a wording system that conveys information concerning the kind of relationship between organisms (Wiley, 1981).

The meaning of phylogenetics has been extended by Kitching *et al.* (1998) that phylogenetics is a method of classification that utilises the hypothesis of character transformation to hierarchically group taxa into nested sets, and then interprets these relationships as a phylogenetic tree (i.e. an hypothesis of genealogical relationships among a group of taxa with specific ancestry and implied time axis).

There are several types of intrinsic characters can be used for phylogenetic analyses. First, we can utilise morphological characters which are structural attributes of an organism. Morphological characters are the primary source of characters in most groups of organisms. To be a useful set of characters, they must vary between taxa but not depend on their environment. Morphological characters may be observed easily by sight or with simple optical aids such as a microscope. They may be simple or complex and have proven useful in distinguishing taxa at varying levels, from phyla to species. External characters are more predominant

characters than internal characters because they are relatively easy to observe. Some of them can be quantified as a reference such as colour patterns, which may be measured by wavelength.

One of reliable characters is molecular character. DNA is the genetic material that a living descendant was inherited from their ancestor in the lineage. Data retrieved from molecular characters like DNA sequencing always give a robust confidence. This type of characters was introduced to phylogenetic study as DNA sequencing study of primate by Kohne, Chicson and Hoyer (1972). After that, most phylogeneticists have been interested in finding new genes or non-coding regions to study about relationships among organisms. Nowaday, this type of knowledge and approach is known as molecular phylogenetics.

The task of molecular phylogenetics is to convert information in DNA and/or protein sequences into an evolutionary tree. There are two different major methods to build such phylogenetic trees based on how the data are treated. First is a 'distance method', which converts aligned sequences into a pairwise distance matrix, then input that matrix into a tree building method, whereas, a 'discrete method' considers each nucleotide site directly. If the original data are in the form of genetic distances, such as those obtained from DNA hybridisation studies, the distance method then should be used. However, if we have the nucleotide or protein sequences, we should analyse them with a discrete method to avoid the loss information that occurs when sequences are converted into distances. The discrete methods are different from the distance methods as they operate directly on the sequences or on functions derived from the sequences rather than pairwise distances. Neighbour-joining is a popular distance method which seeks the tree whose sum of branch lengths is minimum, while the major discrete method is maximum parsimony which choses the tree (or trees) that require fewest evolutionary changes (Page and Holmes, 1988).

2.2.2 Gene targets

Characters employed in molecular phylogenetics are usually nucleic acid sequences. In the past decade, the number of genes and DNA regions used for phylogeny estimation has grown rapidly (Kitching et al., 1998). Most molecular phylogenetic studies of plants in the early 1990s relied on *rbc*L sequences, and to a much smaller degree on 18S ribosomal DNA (rDNA). However, it is now difficult to review adequately all possible gene candidates for DNA sequencing studies. The field of plant molecular systematics has progressed so rapidly that several genes mentioned recently as new alternatives to *rbc*L are now widely sequenced.

Since the plant and animal kingdoms diverged about 1000 million years ago, their patterns of evolution might have become very different. In fact, plants differ from animals in the organisation of their organelle DNA by having a much larger and structurally more variable mitochondrial genome and by having the third (chloroplast) genome (Palmer, 1985 in MacIntyre.). DNA sequence of higher plants evolves at different rates, depending on wheather they are located in nuclear, chloroplast or mitochondrial genome. In angiosperms mitochondrial DNA evolves at least five times more slowly than nuclear sequences. Moreover, plant and mammalian mitochondrial genomes also differ in that the former frequently undergoes rearrangement and is much larger and more variable in size. Despite containing similar set of genes, the two mitochondrial genomes thus clearly evolve in very different ways (Wolfe, Li and Sharp, 1987).

Therefore, suitable genes for phylogenetic analysis must has 1.) a suitable level of nucleotide substitution or mutation which is higher or slower enough to display relationships of the studied samples, 2.) this gene must occur in all studied samples, and 3.) it must be easy to be amplified and its sequences are usually clear. In this M.Sc. thesis investigation, nucleotide sequences from both chloroplast and nuclear genomes were chosen. The *trn*L intron of the chloroplast genome and the

Internal Transcribe Spacers (ITS) of ribosomal DNA (rDNA) and the 5.8S rDNA subunit of the nuclear genome were the target genes for study. The advantages of these nucleotide sites are that 1.) rates of base substitution in plant nuclear ribosomal DNA and chloroplast DNA are close to each other and faster than that of mitochondrial DNA, especially those of non-coding regions in maize, wheat, soybean, pea, tobacco and petunia (Wolfe *et at.,* 1987), 2.) they have been known to have universal markers (Sang, 2002) which are easy to be PCR amplified, 3.) their PCR primers are simply available and 4.) comparing trees based on nuclear and chloroplast markers can be particulary valuable at lower taxonomic levels, providing a clearer evolutionary process that could not be achieved with only one genome alone (Soltis and Soltis, 1998). Apparently, these genes are suitable to use for tree reconstruction at species level (Clegg, 1994). Moreover, a few comparisons between the evolutionary rates of both chloroplast DNA and nuclear DNA have been done. Wolfe *et al.* (1987) previously showed that most chloroplast DNA evolves as half as nuclear DNA. In the genus *Gentiana* L., ITS1 and ITS2 appears to evolve everagely 2.97 and 1.96 times faster than the *trn*L intron, respectively (Gielly *et al.*, 1996). The application of studies on both non-coding sequences of chloroplast DNA and ITS sequences of nuclear DNA have demonstrated their usefulness in resolving plant phylogenies at inter- and intrageneric levels.

2.2.2.1 Chloroplast *trn*L intron

Chloroplast DNA (cpDNA) sequence variations have been widely used to investigate interspecific relationships among angiosperms and other plants. A chloroplast genome is typically circular molecule characterised by two inverted repeat segments that separate the remainder of the molecule into a large and small single copy region. The single copy genes in the chloroplast genome evolved in intermediate rates compared to those of the mitochondrial and nuclear genes (Wolfe

el al., 1987 and Muse, 2000).

*trn*L is one single copy gene in the chloroplast genome, coding for tRNA which can transfer an amino acid leucine to a constructing polypeptide chain in a translation process. It can be found in all plants, some algae and also eubacteria and cyanobacteria. The last also has been assumed to be the chloroplast ancestor (Kuhsel *et al.,* 1990). This *trn*L gene is separated with an intron that is going to splice before a translation process (Fig.13). The special character of *trn*L intron is its variable size from 200 basepair to 600 basepair among plant species (Kuhsel *et al.*, 1990).

In 1991, six universal primers for three non-coding regions of *trn*F, *trn*L and *trn*T gene were designed by Taberlet *et al.* and tested in algae (*Laminaria digitata* and *Codium tomentosum*), bryophytes (*Acrocladium cuspidatum* and *Lunularia cruciata*), pteridophytes (*Thelypteris palustris* and *Equisetum arvense*), gymnosperms (*Gingko biloba* and *Pinus nigra*) and angiosperms (*Magnolia* sp., *Aconitum* sp., *Salix babylonica, Saponaria officinalis, Rosa canina, Robinia pseudacacia, Bellis perennis, Carex elala* and *Phalaris arundinaceae*). Initial comparisons suggested that these coding regions may evolve from ranging at similar rates to *rbcL* (the most widely used gene for plant phylogenetic analyses) to as much as three times faster than *rbc*L sequences, depending on the study group. These *trn*L coding regions are also easily amplified sequenced by those universal primers.

2.2.2.2 ITS regions of nuclear ribosomal DNA

Nuclear ribosomal RNA gene (nrDNA) codes the RNA component of ribosomes. In a plant genome, nuclear ribosomal DNA (nrDNA) exists in large arrays of tandem repeats of the transcription unit and nontranscribed spacer as a multigene family, ranging variously from 200 repeats in *Linum usitatissimum* to 2200 in *Vicia faba* (Rogers and Bendich, 1987). The long tandem arrays could form a nucleolar

organising region (NOR) at one or a few chromosome loci (Long and David, 1980). Each copy composes of three parts coding for different RNA in ribosome: 18S small-subunit RNA, 5.8S RNA and 28S large-subunit RNA. The 18S small-subunit gene and 5.8S gene are divided with a non-coding region called Internal Transcribed Spacer region 1 (ITS1). The Internal Transcribed Spacer region 2 (ITS2) is placed between 5.8S and 28S large-subunit genes. The ITS regions in nrDNA are shown in Fig.14

Although there are numerous copies of rDNA in the nuclear genome, these rDNA repeat units are highly homogenous, supposedly as a result of concerted evolution (Arnheim *et al.*, 1980). The gene arrays evolve together rapidly through many chromosomal processes such as gene conversion (Hillis *et al.,* 1991) and unequal crossing over (Smith, 1976). These homogenisation processes can repair mutations that spread through a multigene family in a relatively in short time and can make sequence homogeneity of rDNA repeats. Hence, rDNA is more attractive for phylogenetic reconstruction than other nuclear gene region (Arnheim, 1983). While the large and small unit rDNA regions have been used to address phylogenetic questions at the family level or higher taxonomic levels in plants (Zimmer *et al.*, 1989), the ITS sequences appear to be useful for assessing relationships at lower taxonomic levels. ITS regions have rates of sequence substitution useful for evaluating the generic and species level relationships (Hillis and Dixon, 1991). The regions evolve more rapidly than other coding regions in general (Brown *et al.*, 1972; Appels *et al.*, 1986) and are more variable as a result of mutagenic processes such as single-base substitutions and insertions or deletions (indels) (Venkateswarlu and Nazar, 1991).

ITS increasingly became an important locus for molecular systematic studies of a wide diversity of organisms from fungi to flowering plants (White *et al.*, 1990). The ITS regions is even more attractive for molecular phylogenetic studies because it can be amplified easily by a polymerase chain reaction (PCR) for DNA sequencing with universal primers from conserved flanking regions in the I8S, 5.8S

and 28S genes. Moreover, the length conservation of ITS sequences among closely related species helps their sequence alignment and phylogenetic analysis (Baldwin *et al.,* 1992). By combining the ITS sequence data with morphological data, ITS data have been playing a useful role in modern plant systematics and evolution research.

Recently, Varela *et al.* (2004) have studied the relationships in the subtribe Diocleinae of the subfamily Papilionoideae (Leguminosae) using ITS sequences. Sequences of 6 genera from the subtribe Diocleinae revealed that genus *Calopogonium* and genus *Pachyrhizus* do not belong to the Diocleinae clade (group). This result supported previous works and suggested that ITS data could show the phylogenetic differences below generic level in Leguminosae.

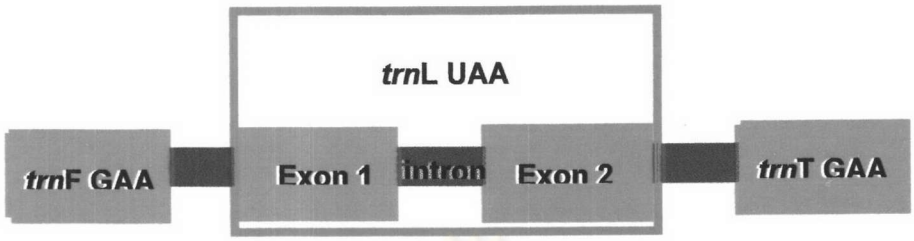

Fig. 13 the location of *trn*L intron (represented in red box) in chloroplast genome. The *trn*L intron is placed between *trn*L gene, exon 1 and exon 2

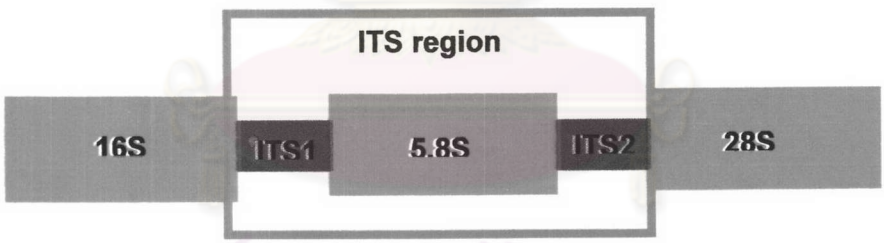

Fig. 14 ribosomal gene copies in nuclear genome. ITS regions (represented in red box) and 5.8S subunit placed between 18S and 28S rRNA genes.

2.2.3 Phylogenetic tree reconstruction methods

2.2.3.1 DNA Data matrix preparation

Phylogenetic analysis of DNA sequence data begins with an alignment of two or more sequences that are hypothesised to be homologous. An alignment involves determining which positions along the DNA or protein sequence are derived from common ancestral positions The alignment probably insert gaps to increase or decrease the nucleotides in each DNA sequences and also to decrease nucleotide mismatching. A character-taxon data matrix is established whenever the sequence alignment is completed. (Page and Holmes, 1998; Doyle and Gaut, 2000).

2.2.3.2 Maximum parsimony and Neighbour-joining

The next step after establishing the character-taxon data matrix is to calculate relationships among taxa. There are several methods (either discreat or distance) to reconstruct a phylogenetic tree (or trees). One most popular method among discrete methods is 'maximum parsimony' which choses the tree (or trees) that requires the fewest evolutionary changes (Page and Holmes, 1998). Optimal criteria for this method is a strategy to search the set of possible trees. Such strategy contain exact searching methods (i.e. exhaustive search and branch-and-bound search) and heuristic searching methods. The 'exhaustive search method' guarantee to find one or all of the shortest cladograms. Every possible cladograms for all included taxa is examined and their lengths are calculated. A simple (phylogenetic trees) algorithm to perform an exhaustive search are that first three taxa are chosen and connected to form an unrooted, fully resolved cladogram for these taxa. The fourth taxon is then selected and joined to each of the tree branches of the first cladogram, yielding three possible networks for four taxa. The fifth taxon is then

added to each of the five branches of the three cladograms. This procedure is continued until all possible cladograms have been constructed. Finally, the lengths of all these cladograms are calculated and the shortest tree is chosen as the most parsimonious phylogenetic trees. This method is however a practical solution for problems with a few number of taxa (approximately 10-12 taxa).

Another exact method but does not require every tree topology to be examined individually is 'branch-and-bound search method'. This method begins with a heuristic calculation of the first cladogram. The length of this cladogram is retained as a reference length (or upper bound) for subsequent cladogram construction. The branch-and-bound method then proceeds in a similar manner to exhaustive search but the lengths of the partial networks are calculated at each step and compared with that of the upper bound. If its length is equal to the upper bound, then this cladogram is retained as one of the set of optimal topologies and the branch-and-bound process continued. However, if the length is less than the upper bound, then this topology is an improvement and its length is substituted as the new upper bound. The attachment of additional taxa can serve only to increase the length further and the number of evaluated cladograms is greatly and quickly reduced. Once all possible paths have been examined, the set of optimal cladograms will have been found. Branch-and-bound applications employ algorithmic devices that reduce computation estimate of the upper bound. However, a branch-and-bound analysis is still time consuming to implement and should not generally be considered for data sets comprising large numbers of taxa.

Approximate or 'heuristic methods' should be adopted in the case of having large members of taxa. This strategy generally uses 'hill climbing' techniques which are essentially trial-and-error and do not guarantee to find all of the minimum-length cladograms but can reduced computational time. Imagine a group of hill hikers aiming to climb to the top of mountain as fast as possible. In order to do that, the 'hikers' best strategy would be to walk up the mountain following the line of steepest

ascent and will eventually reach the summit. However, if there is more than one peak to the mountain, then this approach might yield only a locally optimal result, in that the hikers will simply reach the peak nearest to their starting point. There may be a higher summit elsewhere. Such isolated peaks are referred to as 'islands' If such islands exist then one. This can be translated directly into a search for minimum-length cladograms, the global optimum. The simplest computer algorithms for heuristic search methods make a single pass through the data and construct one tree topology. The resultant cladogram is likely to be only locally optimal unless having good fortune. More complex routines begin with that single topology, then seek to locate the global optimum by rearranging the cladogram in various ways such as 'branch-swapping' algorithm. If multiple islands of tree topology also do exist, then we can run several analyses, which individually start from topologically distinct cladograms.

To use distance methods for phylogenetic tree building, one widely used method is the neighbour-joining. It combines less computational speed with only one resultant tree given. Neighbour-joining technique is a clustering method rather than optimality method, and can not optimise a fitting criterion between tree and data. However, it is a good heuristic method for estimating a minimum evolution tree. One strategy for finding the minimum evolution tree is to first compute the neighbour-joining tree, then see if any local rearrangement of the neighbour-joinging tree produces an even shorter tree. Note that this strategy does not guarantee to find the minimum evolution tree (Page and Holmes, 1998).

2.2.3.3 Tree rooting and consensus tree analyses

When searching for phylogenetic tree, an unrooted tree is first reconstructed, then a rooted tree is further reconstructed . A particular node which is nearest to the common ancestor of analysing taxa will be assigned to be a rooting point of the unrooted tree. Evolutionary polarity will then occur to character states of the ingroups after this process. Briefly, we generally choose one or more taxon to add into the data set as an outgroup for the analysis. Thereafter, most parsimonious tree will be calculated and the node connected to that outgroup will be used as the most proper rooting-point.

- Strict, semi strict and 50% majority rule consensus

After searching with discrete methods, the most parsimonious tree may occur more than one tree. Consensus methods are then usually introduced to solve this problem. They are a convenient means of summarising agreement and disagreement (or congruence and incongruence) between two or more cladograms (phylogenetic trees). All methods of consensus analysis commonly construct a tree from any non contradictory components found among the set of cladograms generated from the initial analysis. Consensus trees can be considered to be indirect methods for resolving character conflicts. They reduce the number of fundamental cladograms produced by parsimony analysis to one tree showing their common components (Kitching *et al.*, 1998). Different consensus methods are suited to different tasks (Nixon and Carpenter, 1996). Here are three consensus methods used in this M.Sc. investigation.

- **Strict consensus**: a strict consemisus tree includes only those groups that occur in all the fundamental cladograms.

- **Semistrict consensus**: a semi-strict consensus tree is formed from

all the uncontradicted components from a set of fundamental cladograms.

- **50%Majority-rule consensus**: a majority-rule consensus tree includes only those components that occur in more than 50% of the fundamental cladograms.

2.2.2.4 Tree evaluation

The simplest supporting measurement for individual clades in the tree is 'branch length'. However, homoplasy (a character that specifies a different and overlapping group of taxa from another to character) makes the interpretation of branch length as support difficult. Other tree support approaches aim to circumvent the problem by assessing the number of extra steps that are required before the clade is lost from the consensus tree of near minimum length cladograms. Here are two tree evaluation methods used in this investigation.

- Bootstrap supporting value

The bootstrap analysis randomly samples characters with replacement to form a pseudoreplicate data-set of the same dimensions as the original. The effect is to delete some characters randomly and to reweight others randomly, with the constraint that the sum of the weights for all characters equals to the number of characters in the matrix. A large number of pseudoreplicates is generated, typically 1000 or more. The most parsimonious cladograms for each pseudoreplicate are then found and the degree of conflict among them assessed by a 50% majority-rule consenseus tree in which a particular group is found might be interpreted as a confidence level associated with that group.

- **Jackknife supporting value**

In contrast to the bootstrap, jackknife sampling is applied without replacement and hence the pseudoreplicate data sets are smaller than the original. Jackknifing aims to achieve better variance estimates from small samples. In first-order jackkniff, pseudoreplicates are constructed by randomly removing one observation (either taxon or character) from the data set. Hence, for a data set of ***T*** observations, ***T*** pseudoreplicates are possible, each comprising ***T-1*** observations from the original sample. The variance of the ***T*** pseudoreplicates are then averaged to give the estimate of the parametric variance.

Current parsimony computer programs also utilise a number of other different statistics to assess the quality of cladograms. Standard measures are consistency index (CI), retention index (RI) and rescale consistency index

- **Consistency index**: a measure of the amount of homoplasy in a character relative to a given cladogram. The consistency index is calculated as the ratio of ***m***, the minimum number of steps a character can exhibit on any cladogram, to ***s***, the minimum number of steps the same character can exhibit on the cladogram in question.

- **Retention index (RI)**: a measure of the amount of similarity in a character that can be interpreted as synapomorphy (shared-derived character) on a given cladogram. The retention index is calculated as the ratio of ***(g-s)*** to ***(g-m)***, where g is the greatest number of steps that a character can exhibit on any cladogram.

- **Rescale consistency index (RC)**: the product of the consistency index and the retention index of a character.