J. Plant Res. 114: 435-442, 2001

Journal of Plant Research

Systematic Status of the *Glycine tomentella* and *G. tabacina* Species Complexes (Fabaceae) Based on ITS Sequences of Nuclear Ribosomal DNA

Yue-Ie C. Hsing[1], Jaw-Shu Hsieh[2], Ching-I Peng[1], Chang-Hung Chou[1] and Tzen-Yuh Chiang[3]*

[1] *Institute of Botany, Academia Sinica, Nankang, Taipei, Taiwan 115*
[2] *Department of Agronomy, National Taiwan University, Taipei, Taiwan 107*
[3] *Department of Biology, National Cheng-Kung University, Tainan, Taiwan 700*

The phylogeny of the subgenus *Glycine* was reconstructed based on nucleotide sequences of the internal transcribed spacer (ITS) region of the nuclear ribosomal DNA to examine the systematic status of the *G. tomentella* and *G. tabacina* species complexes. Rooted at the subgenus *Soja* (2 species, 7 accessions), parsimony analysis was conducted for 17 species (31 accessions) of the subgenus *Glycine*, including 9 and 6 populations of *G. tomentella s.l.* and *G. tabacina s.l.*, respectively. The nrITS phylogeny indicated polyphyly of *G. tomentella s.l.* as well as *G. tabacina s.l.* In the *G. tomentella* species complex, larger legumes, narrower leaflets, and deflexed indumentum hairs differentiated *G. dolichocarpa* from *G. tomentella s.s.* The tetraploid *G. dolichocarpa* ($2n=80$) and aneuploid *G. tomentella* ($2n=38$) represented independent lineages from another clade of the remaining diploid ($2n=40$) and tetraploid species with a DD genome type. Tetraploid *G. tabacina* ($2n=80$) was closely related to *G. dolichocarpa* instead of the diploid *G. tabacina* ($2n=40$) with a BB genome type. The nrITS phylogeny suggested allopolyploidy of *G. dolichocarpa* and of the tetraploid *G. tabacina*, both of which possibly share a common parental species with an AA genome type. Their phylogenetic affinity also indicated biased inheritance of the nrDNA ITS and a possible dominant role of the AA genome. Phylogenetically, *G. dolichocarpa* and allotetraploid *G. tabacina* should be recognized as distinct species.

Key words: *Glycine* — *Glycine tabacina* — *Glycine tomentella* — nrDNA ITS — Phylogeny — Polyploidization

The genus *Glycine* Willd. consists of subgenera *Glycine* and *Soja*, with 16 wild perennial species and two annual species, respectively. The genus is distributed throughout Australia, tropical Asia and East Asia (Newell and Hymowitz 1980). Most species of the subgenus *Glycine* are geographically isolated from the subgenus *Soja* except for *G. tomentella* and *G. tabacina* (Newell and Hymowitz 1978, Ohashi *et al.* 1991). The systematics of *Glycine* has been studied extensively (Hymowitz and Newell 1978, Newell and Hymowitz 1978, 1980, Hymowitz and Singh 1987, Ohashi *et al.* 1991). Over the last two decades, the phylogeny of *Glycine* has also been elucidated based on isozyme data (Broue *et al.* 1977), RFLP patterns of the chloroplast genome (Doyle *et al.* 1990a), urease gene polymorphism (Menancio *et al.* 1990), and nuclear ribosomal DNA variation (Doyle and Beachy 1985, Kollipara *et al.* 1997, Singh *et al.* 1998).

Most *Glycine* species are diploid with a basic chromosome number of $n=20$ (Newell and Hymowitz 1983). Natural tetraploidy ($2n=80$) in *G. tomentella*, *G. tabacina* and *G. hirticaulis*, and aneuploidy [$2n=38, 78$ (fertile) and 79 (sterile)] in *G. tomentella* have been documented (Newell and Hymowitz 1983, Singh and Hymowitz 1985a, Singh *et al.* 1988, 1992, Tindale and Craven 1988, 1993, Kollipara *et al.* 1995). Genome types based on meiotic chromosome behavior were designated and used for interpreting polyploidy (Singh and Hymowitz 1985; Kollipara *et al.* 1997). Multiple origins of both *G. tabacina* and *G. tomentella* have been suggested (Grant *et al.* 1984, Doyle and Brown 1989). Doyle *et al.* (1990b) also suggested that there have been numerous origins involving reticulation in polyploid *G. tabacina*, based on a phylogeny derived from the cpDNA polymorphism. Multiple colonization events in Pacific islands were also suggested on the basis of presence of different plastome types. Cytological and molecular studies of Kollipara *et al.* (1994) indicated frequent allopolyploidization in the evolution of *G. tomentella* complex. A previous study based on RFLP patterns of seed proteins also detected polymorphism within *G. tomentella*, of which one strain with larger legumes shared two specific DNA fragments with *G. tabacina* (Hsing *et al.* 1995). In contrast, the larger-legumed strain shared another two DNA fragments with another *G. tomentella* that had legumes of a smaller size. Another investigation based on SDS-PAGE and Western blotting of seed proteins also revealed genetic differentiation between short-pod, and long-pod *G. tomentella* (Hsieh *et al.* 2001).

Not only did molecular evidence indicate genetic differentiation but morphological features were also diverse in the *Glycine tomentella* complex (Kollipara *et al.* 1994). While studying taxa of Taiwan, Ohashi *et al.* (1991) observed

* Corresponding author: tychiang@mail.ncku.edu.tw

morphological polymorphism in *G. tomentella* (sensu lato) and distinguished *G. dolichocarpa* from *G. tomentella* on the basis of its larger morphs. While fruits of the latter species had two to five seeds, legumes of *G. dolichocarpa* contained five to nine seeds. Upper calyx lobes that were united to 2/3, and the dense, deflexed hairs on the petioles were also distinct in *G. dolichocarpa*. The larger morphs of *G. tomentella* have been illustrated in earlier literature and identified as *G. tomentosa* Benth. (Tang and Lin 1962, Chuang and Huang 1966). Nevertheless, *G. dolichocarpa* has received little attention and has barely been mentioned in recent systematic molecular studies (e.g. Gu *et al*. 1994, Kollipara *et al*. 1997). The systematic status of *G. dolichocarpa* and its phylogenetic relationship with other taxa still remain unclear.

The internal transcribed spacer (ITS) region of the ribosomal DNA, one of the gene families in the nuclear genome, has been frequently used for phylogenetic study at inter- and infrageneric levels (Baldwin *et al*. 1995, Chiang and Schaal 2000) due to its homogeneity among units by concerted evolution as well as fast evolutionary rates (Wojciechowski *et al*. 1993; Li 1997). Kollipara *et al*. (1997) and Singh *et al*. (1998) used the nucleotide sequences of the ITS region of nuclear ribosomal DNA to infer the phylogeny of the entire genus. Nevertheless, multiple samples from each species, including those from the *Glycine tomentella* and *G. tabacina* complex, were not included in the above phylogenetic analyses, in which mostly only a single individual representing each taxon was sampled.

In light of the above developments, this study was performed to examine the monophyly of *Glycine tomentella* sensu lato, and the monophyly of the *G. tabacina* complex with diploid and tetraploid populations. We also examined the genetic status of *G. dolichocarpa* and its relationship to the diploid *G. tomentella*.

Materials and Methods

Plant materials

Both *Glycine tomentella s.l.* and *G. tabacina s.l.* are widespread throughout Australia, tropical Asia, southern China, and Taiwan (Huang and Ohashi 1977). To examine the monophyly of these species complexes, seven populations of *G. tomentella* sensu lato including both strains with large (*G. dolichocarpa* in this study, two populations) and small legumes (*G. tomentella* s.s., five populations), and five populations of *G. tabacina* complexes were sampled. Seeds were collected in natural habitats and germinated in the greenhouse of Academia Sinica in Taipei under uniform conditions (Table 1). Voucher specimens were deposited at the herbarium of Academia Sinica (HAST). To reconstruct a complete phylogeny of subgenus *Glycine*, nucleotide sequences of Kollipara *et al*. (1997) were also included.

Measurement of legume size

Legume sizes, in terms of the number of seeds contained, were measured from 10 randomly selected pods. When fruits were not available from species, whose sequences were taken from Kollipara *et al*. (1997), the information of legume sizes was obtained from the taxonomic literature (Table 1).

DNA isolation, PCR, and nucleotide sequencing

Seedlings of the above materials were ground in liquid nitrogen and stored in a freezer at −70 C. Genomic DNA was extracted from the powdered tissue according to the CTAB procedure (Doyle and Doyle 1987). PCR amplification was performed on a MJ PTC-100 Thermal Cycler using a pair of universal primers, ITS4 (5′-TCCTCCGCTTATTGATATGC-3′) and ITS18 (5′-CGTAACAAGGTTTCCGTAGG-3′) (O'Kane *et al*. 1996). PCR was carried out in 100 μl reaction mixtures containing 10 ng of template DNA, 10 μl of 10X reaction buffer, 10 μl of $MgCl_2$ (25 mM), 10 μl of dNTP mix (8 mM), 10 pmole of each primer, 10 μl of 10% NP-40, and 4 U of *Taq* polymerase (Promega, Madison, USA). PCR amplification conditions were 30 cycles of denaturation at 94 C for 45 s, annealing at 49 C for 1 min 15 s, and extension at 72 C for 1 min 15 s, followed by a final extension step at 72 C for 10 min and storage at 4 C. PCR products were separated on an agarose gel and bands were cut out and purified using a purification kit (Boehringer Mannheim, Mannheim, Germany). Eluted DNA fragments were directly sequenced. Cycle sequencing with *Taq* polymerase was then performed by 32P radioactive labeling using the fmol Sequencing System (Promega). For sequencing, two additional primers ITS3 (5′-GCATCGATGAAGAACGTAGC-3′) and ITS5.8 (5′-ACTC-GATGGTTCACGGGATT-3′) were synthesized. Both strands were sequenced from both ends, using the PCR primers and the two new internal primers.

Sequence alignment and phylogenetic analyses

Nucleotide sequences were aligned with the program CLUSTAL V (Higgins *et al*. 1992). The fixed gap penalty was 35 and the floating penalty was 4. Cladistic analysis of sequence data of the subgenus *Glycine*, rooted at the subgenus *Soja*, was performed with Phylogenetic Analysis Using Parsimony Program (PAUP, Version 3.1.1., Swofford 1993). Heuristic searches were performed with TBR branch swapping, stepwise addition of 10 random replicates, accelerated transformation (ACCTRAN), an unconstrained number of maximum trees, and retention of multiple most parsimonious trees (MULPARS). All characters were unweighted. Confidence values for clades were obtained by bootstrapping (Felsenstein 1985) with 1000 replicates (Hedges 1992). The nodes with bootstrap values greater than 0.70, as a rule of thumb, are significantly supported with $>$ 95% probability (Hillis and Bull 1993). A *g1* test (Huelsenbeck 1991) of skewed tree-length distribution was calculated from 10,000 random trees generated by PAUP to measure the information content of the data. Critical values of the *g1* test were obtained from Hillis and Huelsenbeck (1992). The fit of character data on phylogenetic hypotheses (Swofford 1991) was evaluated using the consistency index (CI; Kluge and Farris 1969) and retention index (RI; Archie 1989). The statistical significance of CI was determined according to the procedure of Klassen *et al*. (1991).

Table 1. The *Glycine* species used in the study, their USDA plant introduction (PI) numbers, seed/pod sizes, and DNA accession numbers

Species	Chromosome Number	PI #	Population #	Collection sites	Seed size (mm)	Seeds/ pod	Pod size (mm)	DNA accession	Ref.
Subgenus *Glycine* Willd.									
G. tomentella Hayata	40	505222	tom 0	Australia. Queensland.	2.5	4-5	18-22	U60542	1
	38	505222	tom 38	Australia. Queensland.	2.4	4-6	19-24	U60544	1
	80	563879	tom 037	Taiwan. Pintung, Haikou.	2.6	3-5	13-18	AJ011345	C
	80	339655	tom 046	Taiwan. Taichung, Chenkunlin.	2.4	3-5	14-20	AJ011343	C
	80	—	tom 047	Taiwan. Kingmen.	2.3	3-5	13-18	AJ011342	C
G. dolichocarpa Tateishi & Ohashi	80	—	tom 019	Australia	3.1	5-7	35-40	AJ011341	C
	80	—	tom 038	Taiwan. Taitung, Chialulan.	2.5	6-8	20-25	AJ224110	C
	80	—	tom 039	Taiwan. Taitung, Tungho.	2.4	6-8	20-25	AJ011340	C
	80	—	tom 040	Taiwan. Taitung, Chialulan.	2.5	6-8	20-25	AJ011339	C
G. tabacina (Labill.) Benth.	40	373990	tab 0	Australia. New S. Wales.	—	1-3	—	U60539	2
	80	320545	tab 003	Taiwan. Penghu, Makung.	2.3	6-8	30-35	AJ009788	C
	80	—	tab 005	Taiwan. Penghu, Shitakubau.	2.3	6-8	30-35	AJ224111	C
	80	—	tab 010	Taiwan. Penghu, Wanan.	2.5	6-8	30-35	AJ011346	C
	80	—	tab 016	Taiwan. Penghu, Gibay.	2.6	6-8	30-35	AJ009789	C
	80	—	tab 064	Taiwan. Penghu, Chimay.	2.4	6-8	30-35	AJ011347	C
G. albicans Tind. & Graven	40	—	alb 0	W. Australia	4.6	2-3	25-35	U60541	3
G. arenaria Tind.	40	505204	are 0	W. Australia	2.8	2-3	12-18	U60543	4
G. argyrea Tind.	40	505151	arg 0	Australia. Queensland.	3.1	8-11	50-60	U60535	5
G. canescens Herm.	40	440932	can 0	S. Australia	4.5	6-8	45-55	U60533	6
	40	—	can 001	Australia	3.6	6-8	23-28	AJ011348	6
G. clandestine Wendl.	40	440958	cla 0	Australia. New S. Wales.	2.2	6-7	26-30	U60534	4
G. curvata Tind.	40	505166	cur 0	Australia. Queensland.	4.1	4-5	35-45	U60547	7
G. cyrtoloba Tind.	40	440962	cyr 0	Australia. Queensland.	2.6	4-5	23-30	U60548	5
G. falcate Benth.	40	505179	fal 0	Australia. Queensland.	4.3	1-2	18-25	U60549	6
G. hirticaulis Tind. & Graven	40	—	hir 0	Australia. New Territory.	3.9	1-2	15-25	U60545	3
G. lactovirens Tind. & Graven	40	—	lac 0	W. Australia	3.9	2-3	20-25	U60540	3
G. latifolia (Benth.) Newell & Hymowitz	40	—	lat 0	Australia. New S. Wales.	2.9	2-3	15-25	U60538	1
	40	—	lat 001	Australia	3.2	2-3	15-25	AJ009786	1
G. latrobeana (Meissn.) Benth.	40	—	lar 0	Australia. Victoria.	2.4	-	-	U60536	1
G. microphylla (Benth.) Tind.	40	—	mic 0	Australia. Queensland.	2.	4-5	15-20	U60537	4
G. pindanica Tind. & Craven	40	—	pin 0	W. Australia	3	2-3	22-27	U60546	8
Subgenus Soja (Moench) F.J. Herm.									
G. max (L.) Merr.	40	—	max 0	(cultivar)	7.7	2-3	40-50	U60551	C
cv. Shi-shi	40	—	max 001	(cultivar)	7.6	2-3	40-50	AJ009787	C
cv. Williams' 82	40	—	max W' 82	(cultivar)	7.7	2-3	45-60	AJ011337	C
G. soja Sieb.& Zucc.	40	81762	soj 0	Russia	3.9	2-3	22-30	U60550	C
[=*G. max spp. formosana*]	40	—	soj 001	Taiwan. Taoyuan. Shimen.	3.1	2-3	15-20	AJ009790	C
	40	07217	soj 039	South Korea	3.9	2-3	22-30	AJ009791	C
	40	407303	soj 043	China. Nanjing.	4.0	2-3	22-30	AJ224109	C

1. Yeh and Cheng (1991), 2. Tateishi and Ohashi (1992), 3. Tindale and Craven (1988), 4. Tindale (1986a), 5. Tindale (1987), 6. Newell and Hymowitz (1978), 7. Tindale (1986b), 8. Tindale and Craven (1993).
"—" indicates no information available, and "C" indicates the present study.

Results

Sequence variation

The alignment of 38 sequences from of 20 accessions consisted of 241 bp in the ITS1 region (ranging from 212 to 233 bp), 174 bp of 5.8 S (ranging from 156 to 173 bp), and 233 bp in the ITS2 region (ranging from 202 to 221 bp) (alignment available from authors on request). Both ITS regions, ranging from 583 bp (*Glycine tomentella* 037) to 611 bp (*G. falcata* 0), are C+G rich with 56.8%. The ITS2 region had 41.1% polymorphic sites (88 out of 214), ITS1 region 31.2%, and the 5.8S rRNA gene 28.7%. Indels involving one to three base pairs were commonly observed. For instance, 385 single-base indels occurred among 38 sequences. Species from the subgenus *Soja* shared two deletions, an eleven-base deletion between positions 80 and 90 and a deletion of eight bases between positions 134 and 141. Five nucleotide changes, positions 48 (C), 101 (C), 115 (C), 254 (T), and 574 (C), characterized the subgenus *Soja*. A deletion of thirteen base pairs between positions 324 and 336 occurred in one population of *G. tomentella* (037). An insertion of TGCCGT-GTC (positions 466-474) occurred in *G. falcata*. All species of subgenus *Glycine*, except for *G. curvata*, *G. cyrtoloba* and *G. falcata*, shared a deletion of CGC at positions 444-446.

Phylogenetic analyses

A strict consensus of nine equally parsimonious trees with 459 steps, a CI of 0.763 ($p<0.05$) and RI of 0.848, was obtained by cladistic analyses (Fig. 1). A *g1* statistic of −0.859 indicated significant phylogenetic information ($p<0.05$) of the data matrix. The subgenera *Soja* and *Glycine* were supported with bootstrap values of 100%. Within the subgenus *Glycine*, the following five clades were recognized: **I**. [*G. cyrtoloba, G. curvata*] (with bootstrap of 100%); **II**. {[(*G. albicans, G. lactovirens*), (*G. tomentella*, ((*G. arenaria, G. hirticaulis*), *G. pindanica*))], *G. tomentella* (*tom38*)} (with bootstrap of 80%); **III**. [(*G. tabacina, G. dolichocarpa*), (*G. argyrea, G. canescens, G. clandestina*), *G. latrobeana*] (with bootstrap of 74%); **IV**. [(*G. latifolia, G. tabacina* (clone *tab0*), *G. microphylla*] (with bootstrap of 90%); and **V**. [*G. falcata*] (Fig. 1). In contrast, within the subgenus *Soja*, neither *G. soja* nor *G. max* formed a monophyletic group.

Variation in fruit size

Number of seeds in each pod and seed size at mature stage were measured from the *Glycine* plants used for nucleotide sequencing in this study (Table 1). We classified legume size into three classes: small with one to three seeds, medium with four to five seeds, and large with six to nine seeds. In contrast, only two sizes, small (two to five)

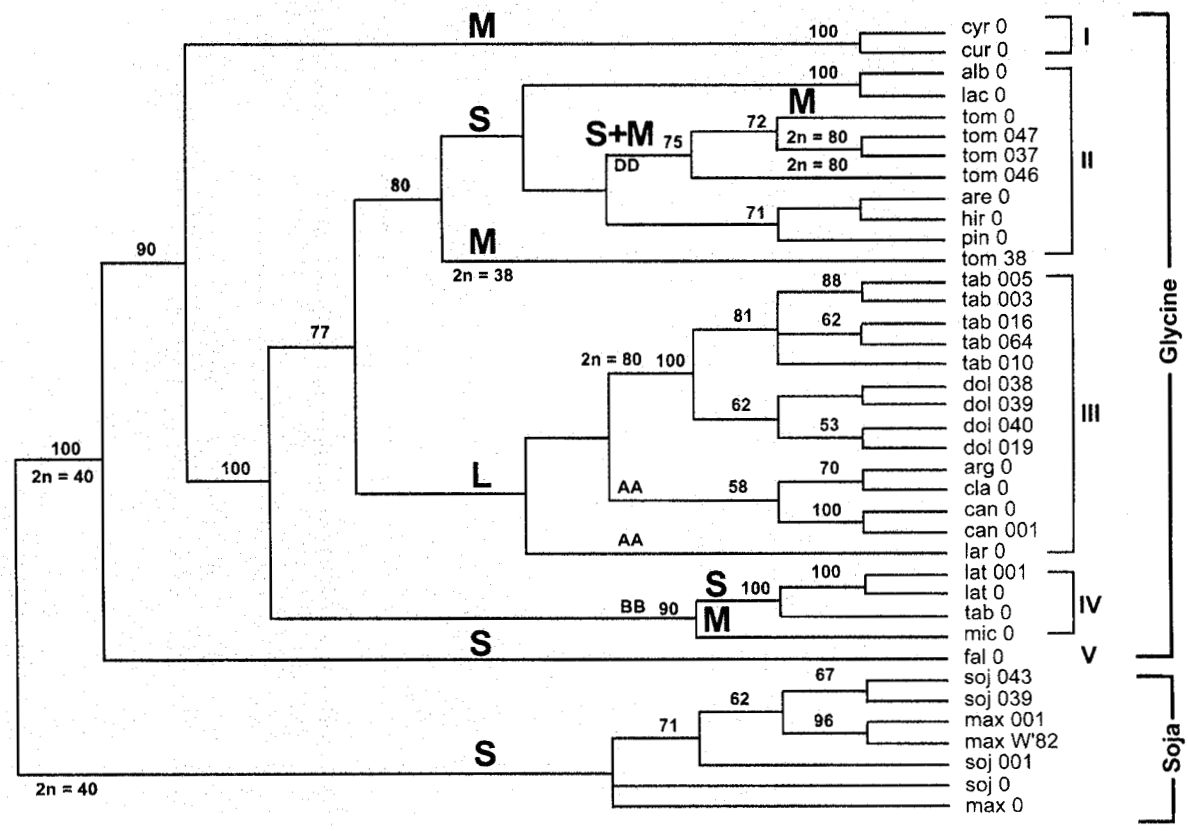

Fig. 1. The strict consensus of nine equally parsimonious trees (CI=0.763, RI=0.848) with a length of 459 steps reconstructed by PAUP based on ITS sequences of *Glycine*. The numbers above branches indicate the bootstrap values (%) derived from 1000 replicates. S (small), M (medium) and L (large) indicate estimated legume size. AA, BB, and DD represent genome types.

and large (five to nine), were used by Tateishi and Ohashi (1992). Legume size was indicated on the most parsimonious tree (Fig. 1). According to our classification, subgenus *Soja* and many taxa of *Glycine* have small legumes. Medium legumes are found in clade I, *G. tomentella* 38, and *G. microphylla*. Polymorphism involving small and medium-sized legumes occurred within and between individuals of *G. tomentella* (both diploids and tetraploids in clade II) (Table 1, Fig. 1). Large legumes are distributed in taxa of clade III.

Discussion

Phylogeny and evolution of legume size of Glycine

The topology of the reconstructed phylogeny of *Glycine* in this study was consistent with the trees estimated by Kollipara *et al.* (1997), in which the two subgenera, *Soja* and *Glycine* consisting of one and five clades, respectively, and were significantly supported. The parsimony trees based on ITS sequences were also congruent with those reconstructed from the chloroplast genome (Doyle *et al.* 1990a) as well as A-199a sequences from the nuclear genome (Zhu *et al.* 1995).

Most taxa in *Glycine* are diploid with $2n=40$ (Table 1, Fig. 1). Polyploidy is distributed in *G. tomentella*, with a tetraploid and one aneuploid ($2n=38$) population in clade II and in *G. tabacina* and *G. dolichocarpa* of clade III. Unexpectedly, *G. dolichocarpa* formed a monophyletic group with the tetraploid *G. tabacina* instead of the diploid or tetraploid *G. tomentella* (clade II) (Fig. 1). This monophyletic group in turn was a sister group to the clade of *G. argyrea, G. canescens* and *G. clandestina* and also to *G. latrobeana*. No single, exclusive clade of tetraploid species could be identified in *Glycine* (Fig. 1). That is, polyploidy in *Glycine* has evolved several times.

Small and medium-sized legumes also appear to have evolved several times (Fig. 1). Nevertheless, within clades, each character state was homologous based on the definition of Patterson (1982) in which a homolog should pass similarity, conjunction, and congruence tests and form a nested relationship with its character states. According to the distribution of states on the parsimony tree, small legumes represent the primitive state. They were found in subgenus *Soja* and in clades II, IV and V (the basal clade) of subgenus *Glycine* (Fig. 1). Taxa with large legumes all occur in clade III. However, our subjective classification may influence this interpretation of legume evolution. Polymorphism occurring in *G. tomentella* further indicated the unnatural division of legume size. When medium and small legumes were merged into a single category by Tateishi and Ohashi (1992), the evolution of legume size in *Glycine* appeared to be congruent with the overall phylogeny and a seemingly nested hierarchical relationship between the character states small and large appeared (Fig. 1). Legume sizes in *Glycine* now appear to have evolved from small to large.

Multiple origins of Glycine tomentella sensu lato *and the differentiation between* G. tomentella *and* G. dolichocarpa

Kollipara *et al.* (1994) proposed the notion of multiple origins of *G. tomentella* based on RFLP evidence of seed protein profiles, chymotrypsin inhibitor, and anti-KTI immunocross-reactive banding patterns. Based on our phylogenetic analyses, using multiple samples of each taxon in question, *G. tomentella* sensu lato has evolved at least three times, twice in clade II and once in clade III (*G. dolichocarpa*) (Fig. 1). Nevertheless, paralogy of ITS sequences, that is duplication prior to speciation, could also result in such a "scattered" distribution (cf. Hillis 1994; Doyle and Doyle 1998). In this study, data congruence in genus Glycine, similar to earlier cytological and molecular evidence (Kollipara *et al.* 1994) and RFLP patterns of seed proteins (Hsing *et al.* 1995), agreed with the allopolyploidy hypothesis, although paralogy of nrITS cannot be ruled out. The topology of the parsimony tree supports the monophyly of *G. dolichocarpa* found by Tateishi and Ohashi (1992). Morphologically, *G. dolichocarpa* can be distinguished from other *G. tomentella* by larger legumes, narrower lanceolate to ovate leaflets, and deflexed indumentum hairs (Tateishi and Ohashi 1992).

In contrast to the maintenance of parental ITS sequences in allotetraploid *Arabidopsis suecica* (O'Kane *et al.* 1996) and the possession of multiple and novel alleles in *Begonia* × *taipeiensis* (Chiang *et al.* 2001), but like many other allopolyploid species such as cotton (Wendel *et al.* 1995), *Phlox* (Ferguson *et al.* 1999), wild soybeans (Doyle and Brown 1989) and *Begonia formosana* (Liu 1999), *Glycine dolichocarpa*, with an allopolyploid origin, possessed only a single allele of the ITS loci. As previously shown in a number of plants, homogeneity among copies of ribosomal DNA can occur very rapidly (Roger and Bendich 1987) and an infrequent variant may become predominant over a few generations (Saghai-Maroof *et al.* 1984, Fuertes Aguilar *et al.* 1999). Apparently, cConcerted evolution via unequal crossing-over contributed to ITS between-copy homogenization after mating and historical hybridization events (Wendel *et al.* 1995, Roelofs *et al.* 1997). In *Glycine*, hypothetical allotetraploid hybrids (*G. dolichocarpa*) apparently attained their own lineages independent from the parental species.

According to the parsimony tree, *Glycine dolichocarpa* is nested in a clade with taxa of the AA genome (cf. Kollipara *et al.* 1997). The other possible parental species for *G. dolichocarpa* would be one of the taxa such as *G. canescens*. Reciprocal artificial hybridization between *G. tomentella* ($2n=40$) and *G. canescens* ($2n=40$) conducted by Singh and Hymowitz (1985b), however, yielded a hybrid F1 with all seeds sterile. In contrast, plants of *G. dolichocarpa* collected from Taiwan were largely fertile. Polyploidization in addition to hybridization may be critical in restoring fertility in *G. dolichocarpa*. It is noticeable that three Taiwanese populations occurred in the clade with the accession *G. tomentella* 019 from Australia. This phylogenetic relationship indicates that the allotetraploids distributed in Taiwan may have their origins in Australia or nearby areas, where hybridization and polyploidization occurred before long-distance dispersal (Hymowitz *et al.* 1990). According to Doyle *et al.* (1990b) the origin of polyploid *G. tabacina* is likely to be a recent event. The recency should also hold for *G. dolichocarpa*, which is

sister to tetraploid *G. tabacina*. Analysis of chloroplast DNA, which has a slower rate of evolution and maternal inheritance, may be able to identify the maternal origin of the hybrids (Palmer 1987).

According to the artificial hybridization data, barriers to sexual reproduction have evolved between *Glycine dolichocarpa* and other tetraploid (Kollipara *et al*. 1994) or aneuploid *G. tomentella* (Newell and Hymowitz 1983). Based on its reproductive isolation and its close phylogenetic relationship to tetraploid *G. tabacina*, as well as the presence of diagnosable morphological characters (cf. phylogenetic species concept; Cracraft 1983), *G. dolichocarpa* should be recognized as a species that is evolving as an independent lineage.

Multiple origins of Glycine tabacina

The tetraploid accessions of *Glycine tabacina* were located in clade III, while a diploid accession of *G. tabacina* 0 was placed in clade IV (Fig. 1). This observation supports the hypothesis of multiple origins of *G. tabacina* (Doyle *et al*. 1990b). Reproductive isolation has evolved and persists among polyploid strains of *G. tabacina*, which may have gained their chloroplast genome from different parental species (Doyle *et al*. 1990c). A previous investigation thus suggested that the tetraploid plants, such as our Taiwan samples, evolved via allopolyploidization (Doyle *et al*. 1990b). The reticulate evolution in both tetraploid *G. tabacina* and *G. dolichocarpa* via independent hybridization events may have occurred between their diploid ancestors, diploid *G. tabacina* (BB genome) and *G. tomentella* (DD genome), respectively, and another taxon with the AA genome from clade III. Such hypothetical hybridization would also agree with our findings of genetic polymorphism based on seed maturation protein profiles (Hsing *et al*. 1995). Interestingly, although diploid *G. tomentella* and *G. tabacina* contributed their genomes to the allotetraploids, the ITS region of the AA genome seems to be "dominant" and to suppress that of the other genomes. Although homogenization among rDNA copies within an individual via concerted evolution has frequently been emphasized (*cf*. Arnheim 1983), "unequal" inheritance of the ITS gene families seems to have occurred in these hybrids. The same has been found in the molecular evolution of 5S ribosomal DNA in allopolyploid cottons (Cronn *et al*. 1996).

Morphological features were also diverse in the *Glycine tabacina* complex. The Taiwanese tetraploids, which were designated as the A plastome group (Doyle *et al*. 1990b), can be distinguished from the diploid and other tetraploid plants with B plastome by the dimorphic leaflets and larger legumes (cf. Tateishi and Ohashi 1992). Based on its genetic and morphological differentiation, tetraploid *G. tabacina* must be recognized as a taxon that represents a novel lineage. According to Doyle *et al*. (1990b), the tetraploid *G. tabacina*, containing the A plastome type, originated in Australia from hybridization involving an AA genome species (clade III) and a diploid BB genome (clade IV) *G. tabacina*. This hybrid then dispersed northward (also see Hymowitz *et al*. 1990). Correspondingly, Taiwan and the Marianas Islands represented marginal populations (Newell 1981). Genetic variation of the ITS nucleotide sequences existing among populations of the Penghu islands indicates that multiple invasions may have occurred if the evolution of allotetraploidy was more recent as suggested by Doyle *et al*. (1990b).

In conclusion, nucleotide sequences of nuclear ribosomal DNA provide further insight into the systematics and evolution of *Glycine*. In this study, phylogenetic analyses supported the hypothesis of multiple origins of *G. tomentella* as well as that of *G. tabacina*. Based on the inferred phylogeny, we recommend recognizing *G. dolichocarpa* and tetraploid *G. tabacina* with A plastomes as distinct and well-defined taxa. In addition, future studies should include tetraploid and aneuploid populations of *G. tabacina* and *G. tomentella* (Doyle *et al*. 1990c, Kollipara *et al*. 1994).

The work was funded by the Academia Sinica, Taiwan. We thank W.C. Tsai and C.F. Chiang for assistance with lab work and library research. We are indebted to Dr. P. Mick Richardson for his critical comments on the manuscript, and to Prof. T. Hymowitz for his instructive comments on the evolution of *Glycine tomentella* species complex and many other comments on molecular techniques.

References

Archie, J.W. 1989. Homoplasy excess ratios: new indices for measuring levels of homoplasy in phylogenetic systematics and a critique of the consistency index. Syst. Zool. **38**: 253-269.

Arnheim, N. 1983. Concerted evolution of multigene families. *In* M. Nei and R.K. Koehn, eds., Evolution of Genes and Proteins. Sinauer, Sunderland. pp. 38-61

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. Ann. Missouri Bot. Gard. **82**: 247-277.

Broue, P., Marshall, D.R. and Muller, W.J. 1977. Biosystematics of subgenus Glycine (Verdc.): isoenzymatic data. Austral. J. Bot. **25**: 555-566.

Chiang, T.Y. and Schaal, B.A. 2000. The internal transcribed spacer 2 region of the nuclear ribosomal DNA and the phylogeny of the moss family Hylocomiaceae. Plant Syst. Evol. **224**: 127-137.

Chiang, T.Y., Hong, K.H. and Peng C.I 2001. Experimental hybridization reveals biased inheritance of the internal transcribed spacer of the nuclear ribosomal DNA in *Begonia* × *taipeiensis*. J. Plant Res. **114**: 343-351.

Chuang, C.C. and Huang, C. 1966. The Leguminosae of Taiwan for pasture and soil improvement. Animal Industry Series No. 7, Department of Botany, National Taiwan University, Taipei, Taiwan.

Cracraft, J. 1983. Species concepts and speciation analysis. Curr. Ornithol. **1**: 159-187.

Cronn, R.C., Zhao, X., Paterson, A.H. and Wendel, J.F. 1996. Polymorphism and concerted evolution in a tandemly repeated gene family. 5S ribosomal DNA in diploid and allopolyploid cottons. J. Molec. Evol. **42**: 685-705.

Doyle, J.J. and Beachy, R.N. 1985. Ribosomal gene variation in soybean (*Glycine*) and its relatives. Theoret.

Appl. Genet. **70**: 369-376.

Doyle, J.J. and Brown, A.H.D. 1989. 5S nuclear ribosomal gene variation in the *Glycine tomentella* polyploid complex (Leguminosae). Syst. Bot. **14**: 398-407.

Doyle, J.J. and Doyle, J.L. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull. **19**: 11-15.

Doyle J.J. and Davis J.I. 1998. Homology in molecular phylogenetics: a parsimony perspective. *In* D.E. Soltis, P.S. Soltis, and J.J. Doyle, eds., Molecular Systematics of Plants II: DNA Sequences. Kluwer Academic Publishers, Boston. pp. 101-131.

Doyle, J.J., Doyle, J.L. and Brown, A.H.D. 1990a. A chloroplast-DNA phylogeny of the wild perennial relatives of soybean (*Glycine* subgenus *Glycine*): congruence with morphological and crossing groups. Evolution **44**: 371-389.

Doyle, J.J., Doyle, J.L., Brown, A.H.D. and Grace, J.P. 1990b. Multiple origins of polyploids in the *Glycine tabacina* complex inferred from chloroplast DNA polymorphism. Proc. Natl. Acad. Sci. U.S.A. **87**: 714-717.

Doyle, J.J., Doyle, J.L., Grace, J.P. and Brown, A.D.H. 1990c. Reproductively isolated polyploid races of *Glycine tabacina* (Leguminosae) had different chloroplast genome donors. Syst. Bot. **15**: 173-181.

Felsenstein, J. 1985. Confidence limits on phylogenies, an approach using the bootstrap. Evolution **39**: 783-791.

Ferguson, C.J., Levin, D.A. and Jansen, R.K. 1999. Natural hybridization between an outcrossing and a selfing *Phlox* (Polemoniaceae): the maternal species of F1 hybrids. Plant Syst. Evol. **218**: 153-158.

Fuertes Aguilar. J., Rossello, J.A. and Feliner, G.N. 1999. Nuclear ribosomal DNA (nrDNA) concerted evolution in natural and artificial hybrids of *Armeria* (Plumbaginaceae). Molec. Ecol. **8**: 1341-1346.

Grant, J.E., Brown, A.D.H. and Grace, J.P. 1984. Cytological and isozyme diversity in *Glycine tomentella* Hayata (Leguminosae). Austral. J. Bot. **32**: 665-677.

Gu, J., Hui, D.W., Zhuang, B.C., Song, W.Y., Xu, B. and Chen, S.Y. 1994. Studies on rDNA ITS1 regions of soybean and its wild relatives. Acta Bot. Sin. **36**: 759-764.

Hedges, S.B. 1992. The number of replications needed for accurate estimation of the bootstrap *P* value in phylogenetic studies. Molec. Biol. Evol. **9**: 366-369.

Higgins, D.G., Bleasby, A.J. and Fuchs, R. 1992. CLUSTAL V. improved software for multiple sequence alignment. Comp. Appl. Biosci. **8**: 189-191.

Hillis, D.M. 1994. Homology in molecular biology. *In* B.K. Hall, ed., Homology: the hierarchical basis of comparative biology. Academic Press, San Diego, pp. 339-369.

Hillis, D.M. and Bull, J.J. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biol. **42**: 182-192.

Hillis, D.M. and Huelsenbeck, J.P. 1992. Signal, noise, and reliability in molecular phylogenetic analyses. J. Hered. **83**: 189-195.

Hsieh, J.S., Hsieh, K.L., Tsai, Y.C. and Hsing, Y.I. 2001. Each species of *Glycine* collected in Taiwan has a unique seed protein pattern. Euphytica **118**: 67-73.

Hsing, Y.I.C., Hsieh, K.L., Huang, Y.C. and Hsieh, J.S. 1995. The relationships of cultivated soybeans and their wild relatives collected from Taiwan revealed by seed proteins. Bot. Bull. Acad. Sin. **36**: 65-72.

Huang, T.C. and Ohashi, H. 1977. *Glycine*. *In* H.L. Li, ed., Flora of Taiwan, 1st edn. Vol. 3. Epoch Publishing Co.: Taipei, pp. 293-298.

Huelsenbeck, J.P. 1991. Tree-length distribution skewness. an indicator of phylogenetic information. Syst. Zool. **40**: 257-270.

Hymowitz, T. and Newell, C.A. 1978. Taxonomy, speciation, domestication, dissemination, germplasm resources and variation in genus *Glycine*. *In* R.J. Summerfield and A.H. Bunting, eds., Advances in Legume Science. Royal Botanical Gardens, Kew, pp. 251-264.

Hymowitz, T. and Singh, R.J. 1987. Taxonomy and speciation. *In* J.R. Wilcox, ed., Soybeans: Improvement, Production, and Uses-Revision. American Society of Agronomy, Madison, pp. 23-48.

Hymowitz, T., Singh, R.J. and Larkin, B.P. 1990. Long-distance dispersal. the case for the allopolyploid *Glycine tabacina* (Labill.) Benth. and *G. tomentella* Hayata in the West-Central Pacific. Micronesica **23**: 5-13.

Klaseen, G.J., Mooi, R.D. and Locke, A. 1991. Consistency indices and random data. Syst. Zool. **40**: 446-457.

Kluge, A.G. and Farris, J.S. 1969. Quantitative phyletics and the evolution of anurans. Syst. Zool. **18**: 1-32.

Kollipara, K.P., Singh, R.J. and Hymowitz, T. 1994. Genomic diversity and multiple origins of tetraploid ($2n=78, 80$) *Glycine tomentella* Hayata. Genome **37**: 448-459.

Kollipara, K.P., Singh, R.J. and Hymowitz, T. 1995. Genomic relationships in the genus *Glycine* (Fabaceae. Phaseoleae). use of a monoclonal antibody to the soybean Bowman-Birk inhibitor as a genome marker. Amer. J. Bot. **82**: 1104-1111.

Kollipara, K.P., Singh, R.J. and Hymowitz, T. 1997. Phylogenetic and genomic relationships in the genus *Glycine* Willd. based on sequences from the ITS region of nuclear rDNA. Genome **40**: 57-68.

Li, W.H. 1997. Molecular Evolution. Sinauer Publishers, Sunderland.

Liu, S.L. 1999. Multiple origin of *Begonia formosana* based on nucleotide variation of chloroplast DNA *atp*B-*rbc*L spacer and nrDNA ITS. Master thesis, National Cheng-Kung University, Tainan, Taiwan.

Menancio, D.I., Hepburn, A.G. and Hymowitz, T. 1990. Restriction fragment length polymorphism (RFLP) of wild perennial relatives of soybean. Theoret. Appl. Genet. **79**: 235-240.

Newell, C.A. 1981. Distribution of *Glycine tabacina* (Labill.) Benth. in the West-Central Pacific. Micronesica **17**: 59-65.

Newell, C.A. and Hymowitz, T. 1978. A reappraisal of the subgenus *Glycine*. Amer. J. Bot. **65**: 168-179

Newell, C.A. and Hymowitz, T. 1980. A taxonomic revision in the genus *Glycine* subgenus *Glycine* Willd. (Leguminosae). Brittonia **32**: 63-69.

Newell, C.A. and Hymowitz, T. 1983. Hybridization in the genus *Glycine* subgenus *Glycine* Willd. (Leguminosae, Papilionoideae). Amer. J. Bot. **70**: 334-348.

Ohashi, H., Tateishi, Y., Nemoto, T. and Hoshi, H. 1991.

Taxonomic studies on the Leguminosae of Taiwan IV. Sci. Rep. Tohoku Imp. Univ., Ser. 4, Biol. **40**: 1-37.

O'Kane, Jr. S.L., Schaal, B.A. and Al-Shehbaz, I.A. 1996. The origin of *Arabidopsis suecica* (Brassicaceae) as indicated by nuclear rDNA sequences. Syst. Bot. **21**: 559-566.

Palmer, J.D. 1987. Chloroplast DNA evolution and biosystematic uses of chloroplast DNA variation. Amer. Naturalist **130**: S6-S29

Patterson, C. 1982. Morphological characters and homology. *In* K.A. Joysey and A.E. Fridays, eds., Problems of Phylogenetic Reconstruction. Academic Press, New York, pp. 21-74.

Roelofs, D., van Velzen, J., Kuperus, P. and Bachmann, K. 1997. Molecular evidence for an extinct parent of tetraploid species *Microseris acuminata* and *M. campestris* (Asteraceae, Lactuceae). Molec. Ecol. **6**: 641-649.

Rogers, S.O. and Bendich, A.J. 1987. Ribosomal RNA genes in plants: variability in copy number and in the intergenic spacer. Plant Molec. Biol. **9**: 509-520.

Saghai-Maroof, M.A., Soliman, K.M., Jorgensen, R.A. and Allard, R.W. 1984. Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. Proc. Natl. Acad. Sci. U.S.A. **81**: 8014-8018.

Singh, R.J. and Hymowitz, T. 1985a. Intra- and interspecific hybridization in the genus *Glycine*, subgenus *Glycine* Willd. Chromosome pairing and genome relationships. Zeit. Pflanzenz. **95**, 289-310.

Singh, R.J. and Hymowitz, T. 1985b. The genomic relationships among six wild perennial species of the genus *Glycine* subgenus *Glycine* Willd. Theoret. Appl. Genet. **71**: 221-230.

Singh, R.J., Kollipara, K.P. and Hymowitz, T. 1987. Intersubgeneric hybridization of soybeans with a wild perennial species, *Glycine clandestina* Wendl. Theoret. Appl. Genet. **74**: 391-396.

Singh, R.J., Kollipara, K.P. and Hymowitz, T. 1988. Further data on the genomic relationships among wild perennial species ($2n=40$) of the genus *Glycine* Willd. Genome **302**: 166-176

Singh, R.J., Kollipara, K.P. and Hymowitz, T. 1992. Genomic relationships among diploid wild perennial species of the genus *Glycine* Willd. subgenus *Glycine* revealed by crossability, meiotic chromosome pairing and seed protein electrophoresis. Theoret. Appl. Genet. **85**: 276-282.

Singh, R.J., Kollipara, K.P. and Hymowitz, T. 1998. The genomes of *Glycine canescens* F.J. Herm. and *G. tomentella* Hayata of western Australia and their phylogenetic relationships in the genus *Glycine* Willd. Genome **41**: 669-679.

Swofford, D.L. 1991. When are phylogeny estimates from molecular and morphological data incongruent? *In* M. Miyamoto and J. Cracraft, eds., Phylogenetic Analysis of DNA Sequences. Oxford University Press, New York, pp. 295-333.

Swofford, D.L. 1993. PAUP. Phylogenetic Analysis Using Parsimony, Version 3.1.1. Computer program distributed by the Illinois Natural History Survey. Champaign, IL.

Tateishi, Y. and Ohashi, H. 1992. Taxonomic studies on Glycine of Taiwan. J. Jap. Bot. **67**: 127-147.

Tang, W.T. and Lin, C.C. 1962. Characterization of wild soybeans in Taiwan. Sci. Agric. **10**: 75-78 (In Chinese).

Tindale, M.D. 1986a. A new north Queensland species of *Glycine* Willd. (Fabaceae). Brunonia **9**: 99-103.

Tindale, M.D. 1986b. Taxonomic notes on three Australian and Norfolk Island species of *Glycine* Willd. (Fabaceae: Phaseolae) including the choice of a neotype for *G. clandestine* Wendl. Brunonia **9**: 179-191.

Tindale, M.D. 1987. Two new eastern Australia species of *Glycine* Willd. (Fabaceae). Brunonia **7**: 207-213.

Tindale, M.D. and Craven, L.A. 1988. Three new species of *Glycine* (Fabaceae, Phaseolae) from NorthWestern Australia with notes on amphicarpy in the genus. Austral. Syst. Bot. **1**: 399-410.

Tindale, M.D. and Craven, L.A. 1993. *Glycine pindanica* (Fabaceae, Phaseolae), a new species from West Kimberley, Western Australia. Austral. Syst. Bot. **6**: 371-376.

Wendel, J.F., Schnabel, A. and Seelanan, T. 1995. An unusual ribosomal DNA sequence from *Gossypium gossypioides* reveals ancient, cryptic, intergenomic introgression. Molec. Phylogenet. Evol. **4**: 298-313.

Wojciechowski, M.F., Sanderson, M.J., Baldwin, B.G. and Donoghue, M.J. 1993. Monophyly of aneuploid *Astragalus* (Fabaceae). evidence from nuclear ribosomal DNA internal transcribed spacer sequences. Amer. J. Bot. **80**: 711-722.

Yeh, M.S. and Cheng, S.H. 1991. Colored Illustration of Leguminous Plant Resources in Taiwan. Council of Agriculture, Taipei.

Zhu, T., Shi, L., Doyle, J.J. and Keim, P. 1995. A single nuclear locus phylogeny of soybean based on DNA sequence. Theoret. Appl. Genet. **90**: 991-999.

(Received February 12, 2001; accepted August 17, 2001)