Plant Syst Evol (2010) 285:23–32
DOI 10.1007/s00606-009-0252-y

ORIGINAL ARTICLE

A molecular phylogenetic study of *Hemsleya* (Cucurbitaceae) based on ITS, *rpl*16, *trn*H-*psb*A, and *trn*L DNA sequences

Hong-Tao Li · Jun-Bo Yang · De-Zhu Li · Michael Möller · Amin Shah

Received: 19 September 2007 / Accepted: 19 November 2009 / Published online: 22 December 2009

Abstract This paper presents the first molecular phylogeny of the genus *Hemsleya* using nuclear ITS and plastid *trn*H-*psb*A, *rpl*16, and *trn*L DNA sequences to examine the relationships among *Hemsleya* species. Phylogenetic relationships were elucidated using a combined analysis of all four datasets, however, the number of parsimony-informative characters was still insufficient to resolve all relationships. Parsimony and Bayesian trees were highly congruent. Twenty-three species of *Hemsleya* split into two major clades corresponding to two subgenera, i.e., subg. *Graciliflorae* and subg. *Hemsleya*. These results are partly in agreement with Li's sectional classification. However, the molecular data are inconsistent with Li's classification at the subsectional level. The molecular phylogeny revealed a striking overall correlation between the phylogenetic relationships of the species and their geographical distribution. The Kangdian ancient landmass could be the center of origin of the genus.

Keywords *Hemsleya* · Molecular phylogeny · Biogeography · ITS · *trn*H-*psb*A · *rpl*16 · *trn*L

H.-T. Li · J.-B. Yang · D.-Z. Li (✉) · A. Shah
Key Laboratory of Biodiversity and Biogeography,
Kunming Institute of Botany, Chinese Academy of Sciences,
650204 Kunming, Yunnan, China
e-mail: dzl@mail.kib.ac.cn

H.-T. Li
e-mail: lihongtao@mail.kib.ac.cn

J.-B. Yang
e-mail: jbyang@mail.kib.ac.cn

A. Shah
e-mail: amin@mail.kib.ac.cn

H.-T. Li
The Graduate School of Chinese Academy of Sciences,
100049 Beijing, China

M. Möller
Royal Botanic Garden Edinburgh, 20A Inverleith Row,
Edinburgh EH3 5LR, Scotland, UK
e-mail: m.moller@rbge.org.uk

Introduction

The Cucurbitaceae is among the most extensively studied of all angiosperm families (Jeffrey 1980a, b, 1990, 2005; Wu and Chen 1985, 1986; Decker-Walters et al. 1990; Wilson et al. 1992; King et al. 1993; Ng 1993; Montes-Hernández and Eguiarte 2002; Kocyan et al. 2007). In spite of extensive phylogenetic research, many relationships are still unknown, especially at the infrageneric level. The genus *Hemsleya* Cogn. includes 24 species according to Li (1993) and is a member of tribe Zanonieae of the subfamily Nhandiroboideae (Jeffrey 2005). Most *Hemsleya* species are distributed in the Hengduan Mountains of the Sino-Himalayan Forest subkingdom, with a few species occurring in the Sino-Japanese Forest subkingdom (Lu 1982; Wu and Chen 1985, 1986; Li 1993). The genus has been studied by Jeffrey (1980a, b, 2005) and Wu and Chen (1985, 1986) using a morphological-phytogeographic approach, while Li (1993) published a monograph on the systematics and evolution of *Hemsleya* employing an integrative approach involving anatomical, embryological, palynological, and cytological methodologies. However, various morphological-based studies resulted in conflicting relationships among *Hemsleya* species and the infrageneric classifications have never been tested using molecular data.

Most members of the genus *Hemsleya* occupy moist and shady understory habitats in montane moist evergreen broadleaved forests. The flowers are dioecious and occur in

an inflorescence of seriate dichasia. The male flowers possess five stamens, while female flowers bear three stigmas that bifurcate at the apex. All fruits are nine- to ten-ribbed capsules, and seed are with or without seed-wings. Five densely imbricate petals form radiate, bow-like, and spherical, umbraculiform, or patelliform corollas. The annual herbaceous vines grow from perennial rhizomatous tubers (persistent roots in *H. graciliflora*), with alternate palmate compound leaves comprising 5–11 leaflets. The plants are well known for their medicinal uses (Nie et al. 1984; Nie and Chen 1986; Dinan et al. 1997; Hano et al. 1997; Yang et al. 2000; Chiu and Gao 2003; Chen et al. 2003; Lin et al. 2003), and consequently several species are cultivated for medicinal purposes, such as *H. amabilis*, *H. chinensis*, and *H. macrosperma*. The medical properties of *Hemsleya* are exploited in the pharmaceutical industry. Anticancer activities of *H. amabilis* have been reported (Wu et al. 2002).

The genus *Hemsleya* was established in 1888 by Cogniaux ex Forbes et Hemsley (Forbes and Hemsley 1888; Jeffrey 1980a, b; Wu and Chen 1986; Li 1993) with only one species. In the following decades, few new species were reported and as a result, many botanists thought that this genus was a monotypic or an oligotypic genus until the 1980s. Many new species have been described since then. It is now considered a medium-sized genus in the Cucurbitaceae with a principal distribution in southern and southwestern China. The first complete infrageneric classification encompassed 31 species and ten varieties in four sections based on morphological characters of flowers, fruits, and seed (Wu and Chen 1985, 1986). The genus was chemotaxonomically studied by Nie and Chen (1986) and Qiu et al. (2005). Li (1993) proposed a new infrageneric classification with 24 species and 12 varieties in two subgenera: subg. *Graciliflorae* with only one species, and subg. *Hemsleya*, which was further divided into three sections: sect. *Delavayanae*, sect. *Amabiles* (including subsect. *Macrocarpae*, *Dipterygiae*, *Amabiles*, and *Sphaerocarpae*), and sect. *Hemsleya* (including subsect. *Carnosiflorae* and *Chinenses*).

The main objective of this paper is to reconstruct a molecular phylogeny of the genus *Hemsleya* and to test the infrageneric classifications of Wu and Chen (1985, 1986) and Li (1993). In addition, we would like to address the cause of the biogeographic pattern of this Sino-Himalayan genus and to get a better understanding of the origin and evolution of the genus in relation to the uplift of the Himalayas. In this study, we sequenced the nuclear ribosomal internal transcribed spacers (nrITS) as well as three chloroplast regions: the *rpl*16 intron, the *trn*H-*psb*A intergenic spacer, and the *trn*L region. Both ITS and *trn*H-*psb*A have been useful in clarifying interspecific and intergeneric relationships in Cucurbitaceae (Jobst et al. 1998; Jarret and Newman 2000; Garcia-Mas et al. 2004), particularly in the subfamily Cucurbitoideae.

Materials and methods

Plant samples

All *Hemsleya* species recognized by Li (1993), except *H. heterosperma*, were sampled to represent all subsections (Wu and Chen 1986; Li 1993). *Hemsleya heterosperma*, distributed in Myanmar, was not sampled because of inaccessibility. We used one species of *Gomphogyne* and one species of *Gynostemma* as outgroups, as both genera are members of the subtribe Gomphogyninae and are closely related to *Hemsleya* (Jeffrey 1990, 2005; Li 1993; Chen 1995; Kocyan et al. 2007). Samples were collected in the wild, except for *H. endecaphylla,* which was collected from locally cultivated material (Table 1). Healthy, clean leaves were quickly dried in silica gel, and voucher herbarium specimens were deposited at the Herbarium of the Kunming Institute of Botany of the Chinese Academy of Sciences (KUN).

DNA extraction, amplification, and sequencing

For each sample, total genomic DNA was isolated from 0.2 g silica gel-dried or 0.4 g fresh leaf material ground in liquid nitrogen using a modified CTAB method (Doyle and Doyle 1987), using 4% CTAB and the addition of approximately 1% polyvinyl polypyrrolidone (PVP) and 2% β-mercaptoethanol instead of 2% CTAB.

Following extraction, DNA was amplified by PCR (Saiki et al. 1988). The nrITS region includes the ITS1, 5.8S, and ITS2 sequences and was amplified with the reverse primer ITS-4 (5′-TCC TCC GCT TAT TGA TAT GC-3′) and the forward primer ITS-5 (5′-GGA AGG AGA AGT CGT AAC AAG G-3′) (White et al. 1990). The *trn*H-*psb*A region was amplified using the method described by Shaw et al. (2005) with the forward primer *trn*H^{GUG} (5′-CGC GCA TGG TGG ATT CAC AAT CC-3′) (Tate and Simpson 2003) and the reverse primer *psb*A (5′-GTT ATG CAT GAA CGT AAT GCT C-3′) (Sang et al. 1997). The *rpl*16 intron region was amplified following Jordan et al. (1996) with the forward primer *rpl*16-F71 (5′-GCT ATG CTT AGT GTG TGA CTC GTT G-3′) and the reverse primer *rpl*16-R1516 (5′-CCC TTC ATT CTT CCT CTA TGT TG-3′). The *trn*L region (including the *trn*L intron and the *trn*L-F spacer) was amplified following Taberlet et al. (1991) with the forward primer ucp-c (5′-CGA AAT CGG TAG ACG CTA CG-3′) and the reverse primer ucp-f (5′-ATT TGA ACT GGT GAC ACG AG-3′).

Table 1 Taxa and voucher specimens of *Hemsleya* and outgroups used in this study

Species	Section	Subsection	Locality information	GenBank no.				Voucher
				ITS	*trn*H-*psb*A	*rpl*16	*trn*L	
Subg. *Graciliflorae*								
Hemsleya graciliflora	Graciliflorae		Widespread (SH, SJ)	EF621654	EF621700	EF621640	EU305242	H. T. Li 069
Subg. *Hemsleya*								
H. delavayi	*Delavayanae*		Songming, Yunnan (SH)	EF424063	EF424068	EF424072	EU305236	H. T. Li 048
H. macrocarpa	*Amabiles*	*Macrocarpae*	Yongde, Yunnan (SH)	EF621652	EF621698	EF621632	EU305246	H. T. Li 003
H. mitrata	*Amabiles*	*Macrocarpae*	Yongde, Yunnan (SH)	EF621650	EF621696	EF621630	EU305248	H. T. Li 052
H. panlongqi	*Amabiles*	*Macrocarpae*	Gulin, Sichuan (SJ)	EF621647	EF621693	EF621624	EU305251	H. T. Li 040
H. dipterygia	*Amabiles*	*Dipterygiae*	Jinxiu, Guangxi (SJ)	EF621659	EF621704	EF621623	EU305237	H. T. Li 064
H. lijiangensis	*Amabiles*	*Amabiles*	Lijiang, Yunnan (SH)	EF424065	EF424070	EF424075	EU305244	H. T. Li 047
H. amabilis	*Amabiles*	*Amabiles*	Songming, Yunnan (SH)	EF424066	EF424071	EF424074	EU305232	H. T. Li 060
H. sphaerocarpa	*Amabiles*	*Sphaerocarpae*	Kaili, Guizhou (SJ)	EF621645	EF621691	EF621634	EU305253	H. T. Li 004
H. panacis-scandens	*Hemsleya*	*Carnosiflorae*	Gejiu, Yunnan (SH)	EF621648	EF621694	EF621636	EU305250	H. T. Li 002
H. omeiensis	*Hemsleya*	*Carnosiflorae*	Omei, Sichuan (SH)	EF621649	EF621695	EF621627	EU305249	H. T. Li 035
H. pengxianensis	*Hemsleya*	*Carnosiflorae*	Pengxian, Sichuan (SJ)	EF621646	EF621692	EF621626	EU305252	H. T. Li 033
H. turbinata	*Hemsleya*	*Carnosiflorae*	Yongde, Yunnan (SH)	EF621644	EF621690	EF621633	EU305254	H. T. Li 836
H. ellipsoidea	*Hemsleya*	*Carnosiflorae*	Omei, Sichuan (SH)	EF621658	EF621703	EF621631	EU305239	H. T. Li 014
H. carnosiflora	*Hemsleya*	*Carnosiflorae*	Luoping, Yunnan (SJ)	EF621661	EF621685	EF621635	EU305233	H. T. Li 001
H. chengyihana	*Hemsleya*	*Carnosiflorae*	Yongde, Yunnan (SH)	EF621660	EF621705	EF621637	EU305234	H. T. Li 055
H. zhejiangensis	*Hemsleya*	*Carnosiflorae*	Taishun, Zhejiang (SJ)	EF621643	EF621689	EF621639	EU305255	H. T. Li 012
H. chinensis	*Hemsleya*	*Chinenses*	Omei, Sichuan (SH)	EF424064	EF424069	EF424073	EU305235	H. T. Li 023
H. gigantha	*Hemsleya*	*Chinenses*	Shimian, Sichuan (SH)	EF621656	EF621701	EF621629	EU305241	H. T. Li 028
H. longicarpa	*Hemsleya*	*Chinenses*	Pengxian, Sichuan (SJ)	EF621653	EF621699	EF621625	EU305245	H. T. Li 034
H. macrosperma	*Hemsleya*	*Chinenses*	Zhaotong, Yunnan (SH)	EF621651	EF621697	EF621628	EU305247	H. T. Li 070
H. dulongjiangensis	*Hemsleya*	*Chinenses*	Gongshan, Yunnan (SH)	EF621657	EF621686	EF621621	EU305238	YDJ 20394
H. endecaphylla	*Hemsleya*	*Chinenses*	Lijiang, Yunnan (SH)	EF621655	EF621702	EF621638	EU305240	H. T. Li 049
Outgroups								
Gomphogyne cissiformis			Yongde, Yunnan	EF621663	EF621706	EF621641	EU305231	H. T. Li 835
Gynostemma pentaphyllum			Widespread	EF621662	EF621687	EF621622	EU305230	H. T. Li 005

SH Sino-Himalayan Forest subkingdom, *SJ* Sino-Japanese Forest subkingdom

PCR amplifications were performed in 25 µl volume reactions with final concentrations of 1× buffer, 0.2 U of *Taq* polymerase, 2.5 mM of $MgCl_2$, 250 µM of dNTPs, 0.2 µM of each forward and reverse primer, and approximately 50 ng of template DNA. The thermal cycler (PE9600 or PE9700, Perkin Elmer) was programmed with an initial denaturation step of 4 min at 94°C; followed by 36 cycles of 1 min at 94°C, 45 s at 50°C, 1 min at 72°C; and a final extension step of 7 min at 72°C. PCR products were visualized by ethidium bromide and agarose gel electrophoresis, cleaned with Wizard PCR preps DNA Purification system (Promega, Madison, WI, USA), and both strands sequenced for sequence confirmation using the same primers as used for PCR amplifications.

In order to check for the concerted evolution of ITS, PCR products of the species *H. amabilis*, *H. chinensis*, *H. lijiangensis*, *H. endecaphylla*, and *H. dulongjiangensis* were cloned into Promega's pGEM-T System I vector. Ligation, transformation, and plating were carried out following the recommendations of the manufacturer with some modifications.

Cycle sequencing was carried out on PE9600 or PE9700 thermal cyclers. Each 5 µl reaction contained 1 µl of purified PCR product, 1 µl of 0.2 µM of primer, 1 µl Big Dye, and 2 µl ddH_2O and was run through a program of 33 cycles of 96°C for 10 s, 50°C for 5 s, and 60°C for 4 min. Reactions were cleaned using CleanSEQ (Perkin-Elmer, Norfolk, CT) and run on an ABI PRISM 3700 DNA capillary automated sequence analyzer.

Phylogenetic analyses

DNA sequences were assembled using Lasergene analysis software (DNASTAR, Madison, WI, USA; Burland 2000). Clustal X (Thompson et al. 1997) was used to produce a basic sequence alignment, which was further corrected manually using BioEdit (Hall 1999). Gaps were coded using GapCoder (Simmons and Ochoterena 2000).

Maximum parsimony (MP) analyses were conducted using PAUP 4.0b10 (Swofford 2001). Characters were treated as unordered and unweighted. Heuristic searches were conducted with tree-bisection reconnection (TBR) branch swapping, MulTrees ON, and 10,000 random taxon addition replicates holding 20 trees at each step. Bootstrap support (BS) values for individual clades were calculated by running 1,000 bootstrap replicates of the data, with starting trees acquired by a single replicate of random stepwise addition of taxa, under TBR branch swapping, and MulTrees ON. The consistency index (CI), retention index (RI), rescaled consistency index (RC), and pairwise distances were obtained through PAUP 4.0b10 as the actual number of site differences excluding indels.

Bayesian analyses (BA) were conducted using MrBayes 3.12 (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003). Model parameters were determined by the Akaike information criterion (AIC) in ModelTest v3.6 (Posada and Crandall 1998) for the best fitting models of molecular evolution for each gene. The results based on the AIC test showed that the best-fit models were the TIM + G model for the ITS sequences [Base = (0.1934 0.3680 0.2742), Nst = 6, Rmat = (1.0000 1.2784 0.3202 0.3202 2.7168), Rates = gamma, Shape = 0.6543, Pinvar = 0], the TMV + G model for the *trn*H-*psb*A spacer [Base = (0.4131 0.1335 0.1393), Nst = 6, Rmat = (2.3874 1.5769 0.4531 0.6652 1.5769), Rates = gamma, Shape = 0.3713, Pinvar = 0], the K81uf model for the *rpl*16 intron [Base = (0.2791 0.1671 0.1362), Nst = 6, Rmat = (1.0000 1.2242 0.6386 0.6386 1.2242), Rates = gamma, Shape = 0.4051, Pinvar = 0], and the TMV + G model for the *trn*L region [Base = (0.3591 0.1759 0.1628), Nst = 6, Rmat = (0.6098 1.2978 0.2622 1.1866 1.2978), Rates = gamma, Shape = 0.2978, Pinvar = 0]. Four independent Markov Chain Monte Carlo chains were run simultaneously and sampled every 100 generations for a total of 1,000,000 generations. To establish the "burn-in" phase, i.e., log probability values stationarity, a plot of generations against log likelihood scores was performed using Excel 2003 (Microsoft, Redmond, WA, USA); these burn-in trees were discarded from the analysis. The majority rule consensus tree was calculated using PAUP 4.0b10 (Swofford 2001).

Congruence among the data sets (ITS, *trn*H-*psb*A, *rpl*16 intron, and *trn*L region) was evaluated with the partition homogeneity test of Farris et al. (1994, 1995) implemented in PAUP 4.0b10 (Swofford 2001), using 1,000 resampling replicates under maximum parsimony, and all characters equally weighted.

We conducted MP and BA analyses on individual datasets (ITS, *trn*H-*psb*A, *rpl*16 intron, and *trn*L region) and the combined dataset. Phylogenetic inferences were based on comparisons of both the MP strict consensus tree and Bayesian majority rule consensus tree.

Results

ITS clone sequences

Five independent clones of the ITS regions in each of five species of *Hemsleya* were sequenced. Among *H. amabilis*, *H. lijiangensis*, *H. endecaphylla,* and *H. dulongjiangensis*, the five sequences of each species were identical. In *H. chinensis*, there was a nucleotide difference among the five sequences. The cloned ITS sequences were used to construct a MP dendrogram, and sequences clearly grouped into subclades by species.

Sequence comparisons

Sequence characteristics for the five datasets (ITS, *trn*H-*psb*A, *rpl*16, *trn*L, and the combined ITS + *trn*H-*psb*A + *rpl*16 + *trn*L matrix) are shown in Table 2. The aligned combined matrix, including coded gaps, consisted of 3,087 characters, of which 354 (11.5%) were variable and 130 (4.9%) were potentially parsimony informative. The ITS sequences alone showed variation at 159 of 552 sites (29%), and of these 48 (8.7%) were potentially informative. The *trn*H-*psb*A spacer sequences possessed variation at 55 out of 495 sites (11.1%), of which 24 (4.8%) were potentially parsimony informative. The *rpl*16 intron sequence showed variation at 84 out of 1,056 sites (8%), and 34 (3.2%) were potentially informative. The *trn*L region sequences possessed variation at 56 out of 984 sites (5.7%), of which 24 (2.5%) were potentially parsimony informative.

The entire ITS region, including both spacers and the 5.8S gene, was 631 base pairs (bp) for all taxa, except for *H. graciliflora*, which had a sequence length of 643 bp. Alignment of this species required a total of three indels. The first two were a 10-bp and a 1-bp insertion in ITS1, and another was a 1-bp insertion in ITS2. The ITS1 region alone was 228 bp (except for *H. graciliflora* with 239 bp), while the ITS2 region was 239 bp (except for *H. graciliflora*, which was 240 bp). The 5.8S gene was the most conserved region and was 164 bp long in all species sequenced. There were no variable sites in the 5.8S gene, so it was excluded from the analyses. The *trn*H-*psb*A spacer region was approximately 500 bp long, with nine indels ranging from 1 to 22 bp in length. The 22-bp insertion was an autapomorphy for *H. graciliflora*. The *rpl*16 intron region was approximately 1,050 bp, with a poly-A region ranging from 10 to 12 bp among all species included, as well as an 8-bp insertion unique to *H. macrosperma*. The *trn*L region was about 1 kb long including eight 1-bp indels.

Pairwise divergence

The calculated pairwise divergence values of nucleotide substitutions between all pairs of 25 taxa showed an average of 1.59%, ranging from 0.97 to 7.39%. *Hemsleya* species had an average pairwise divergence of 0.57%, ranging from 0.44 to 0.92%. Subgenus *Graciliflorae* showed a pairwise divergence from subg. *Hemsleya* of 0.92%, and subg. *Hemsleya* had an average pairwise divergence of only 0.55%, ranging from 0.41 to 0.68%.

Phylogenetic analyses

Separate analyses using MP and BA of the ITS, *trn*H-*psb*A, *rpl*16, and *trn*L datasets produced trees with relatively low resolution for many clades. The partition homogeneity test suggested that the four datasets were not significantly incongruent ($P = 0.05$). Furthermore, previous studies have questioned the validity of the incongruence length test (Reeves et al. 2001; Yoder et al. 2001), and it may be more reasonable to identify strongly supported topological conflict as a measure of incongruence (Wiens 1998). Therefore, the four datasets were combined for phylogenetic analyses.

In our study, the method of data analysis (MP or BA) had no effect, and the resulting topologies were highly similar. MP and BA recovered well-resolved topologies with moderately supported branches. The MP analysis of the combined data yielded two most parsimonious trees of 630 steps (CI = 0.854, RI = 0.708, RC = 0.605). The strict consensus tree was highly resolved with mostly moderate branch support (Fig. 1), as shown by the bootstrap values. The BA reached stationarity after around 80,000 generations. Thus the first 801 trees were eliminated, leaving a total of 9,200 trees in the majority rule consensus tree. The topologies of the MP and BA analyses strongly supported the monophyly of the genus *Hemsleya* (BS = 100%; PP = 100%). In both analyses (Figs. 1, 2), the basal-most group was found to be *H. graciliflora* of the monotypic subg. *Graciliflorae*, which was sister to the monophyletic subg. *Hemsleya* (BS = 86%; PP = 100%). Among the remainder of subgenus *Hemsleya*, three moderately supported clades were resolved. Clades B and C formed sister clades in relation to clade A in the analyses (BS = 75%, PP = 100%).

Table 2 DNA site variation and tree statistics for the five data sets used in the phylogenetic analyses of 23 *Hemsleya* species and two outgroups

	Characters (*n*)	Variable sites (*n*)	Informative sites (*n*)	Informative sites (%)	Trees (*n*)	Tree length	CI	RI	RC
ITS	552	159	48	8.7	3	270	0.925	0.730	0.675
*trn*H-*psb*A	495	55	24	4.8	16	101	0.917	0.837	0.767
*rpl*16	1,056	84	34	3.2	1	138	0.950	0.882	0.839
*trn*L	984	56	24	2.5	4	95	0.853	0.868	0.740
ITS + *trn*H-*psb*A + *rpl*16 + *trn*L	3,087	354	130	4.9	2	630	0.854	0.708	0.605

CI Consistency index, *RI* retention index, *RC* rescaled consistency index

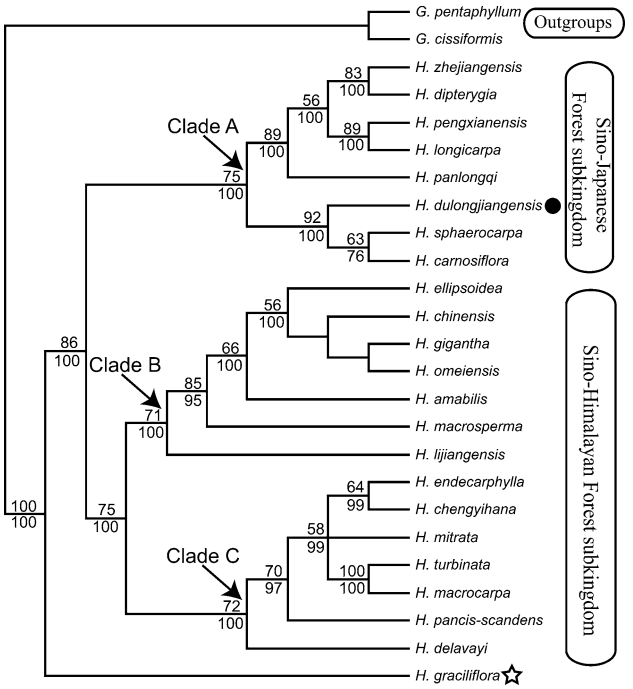

Fig. 1 Strict consensus tree (CI = 0.854, RI = 0.708, RC = 0.605) of the two most parsimonious trees based on combined data of four DNA regions (ITS, *trn*H-*psb*A, *rpl*16, and *trn*L) for 25 taxa. Bootstrap values are shown *above the branches*; posterior probability values are shown *below the branches*. *Filled circle* indicates the geographical distribution belonging to Sino-Himalayan Forest subkingdom. *Star* indicates the only widespread species with its geographical distribution both in the Sino-Himalayan Forest subkingdom and the Sino-Japanese Forest subkingdom

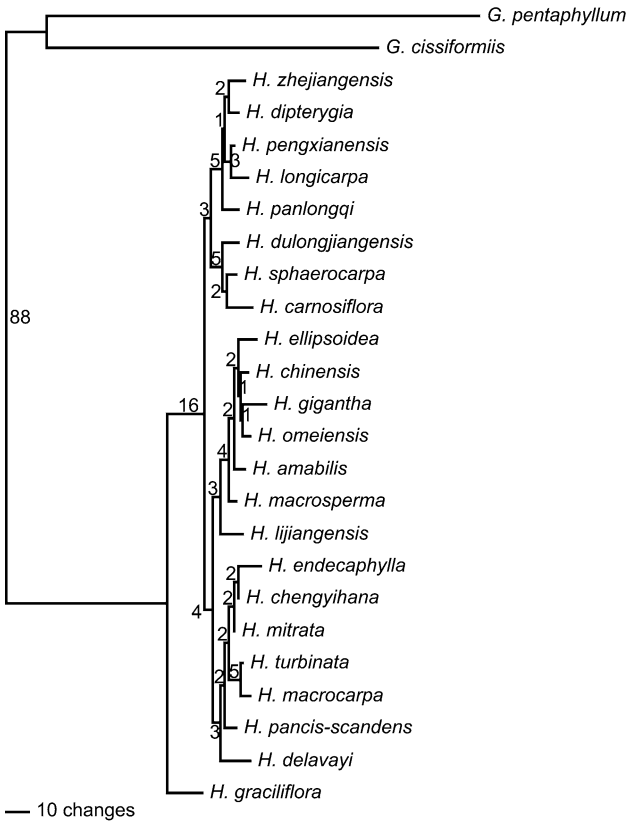

Fig. 2 One of the two most parsimonious trees as a phylogram. Branch lengths are shown *above the branches*

Within the clades, some sister pair relationships were well supported in both analyses.

Clade A, including *Hemsleya panlongqi*, *H. zhejiangensis*, *H. dipterygia*, *H. pengxiansis*, and *H. longicarpa*, as well as *H. dulongjiangensis*, *H. sphaerocarpa*, and *H. carnosiflora*, represents the Sino-Japanese Forest subkingdom, except for *H. dulongjiangensis*, which is from northwestern Yunnan, part of the Sino-Himalayan Forest subkingdom. Clade B comprises *H. ellipsoidea*, *H. chinensis*, *H. gigantha*, *H. omeiensis*, *H. amabilis*, *H. macrosperma*, and *H. lijiangensis.* The first four species are from southern Sichuan, and the last three species are from northeastern Yunnan, except *H. lijiangensis* from northwestern Yunnan. In fact, the first six species are continuous in their geographical distribution. Clade C consists of *H. endecaphylla*, *H. chengyihana*, *H. mitrata*, *H. turbinata*, *H. macrocarpa*, *H. panics-scandens*, and *H. delavayi*. These species are geographically restricted to southwestern Yunnan (excluding *H. endecaphylla*, collected from locally cultivated material in Lijiang, northwestern Yunnan). Species of clades B and C are all from the Sino-Himalayan Forest subkingdom.

Discussion

DNA sequences from several different genes, gene regions, and genomes, including nuclear and chloroplast sequences, have been extensively employed to determine relationships at both higher and lower taxonomic levels in Cucurbitaceae (Ganal and Hemleben 1986; Jobst et al. 1998; Zhang et al. 2006). Comprehensive molecular phylogenetic studies have been carried out on many genera within the family (Sanjur et al. 2002; Chung et al. 2003, 2006; Clarke et al. 2006), but mostly in subfamily Cucurbitoideae and a few in subfamily Nhandiroboideae, and none involving the genus *Hemsleya*.

The combined analyses of nuclear ribosomal ITS and chloroplast *trn*H-*psb*A spacer, *rpl*16 intron, and *trn*L region sequences proved to be very useful in resolving species relationships within *Hemsleya*. In spite of the poorly resolved trees produced by analyses of individual data sets, combined analyses of the four DNA regions recovered highly resolved topologies. Several studies have previously shown that better resolution and support for phylogenetic topologies is achieved by increasing the number of characters (Soltis et al. 1998; Graybeal 1998; Hillis 1998;

Bremer et al. 1999; Wortley et al. 2005). The current study is in agreement with those studies.

Although the present sequence data did not resolve all relationships among *Hemsleya* species, the different phylogenetic analyses conducted (MP, BA) resulted in congruent trees that supported the monophyly of the genus *Hemsleya* with respect to the two outgroup genera chosen (Jeffrey 1990, 2005; Li 1993). The trees were well resolved and had moderately supported topologies. In the Bayesian analysis, the topologies received stronger support values. Some studies (Rannala and Yang 1996; Suzuki et al. 2002; Wilcox et al. 2002) have shown that Bayesian posterior probability values are almost always higher than parsimony bootstrap values for a given node, as the posterior probability value is dependent on the model, the data, and the priors. Therefore, the Bayesian topology should be interpreted with more caution. Similarly, the bootstrap analysis also has the limitation of resampling to consistently underestimate true nodal support (Hillis and Bull 1993; Wilcox et al. 2002). Thus, in this study, we interpreted our results in the light of both bootstrap and posterior probability values.

In general, the ITS region and the *trn*H-*psb*A spacer can provide enough informative sites for phylogenetic analyses. The ITS region was able to resolve phylogenetic relationships among species of the subfamily Cucurbitoideae (Jobst et al. 1998). Furthermore, *trn*H-*psb*A spacer has been discussed as one of the barcoding regions (Kress et al. 2005; Vischi et al. 2006; Kress and Erickson 2007). However, neither provided enough informative sites in *Hemsleya*. Separate ITS, *trn*H-*psb*A, *rpl*16, and *trn*L datasets did not provide enough informative sites among *Hemsleya* species. Despite using a combined analysis of four datasets, the number of parsimony-informative characters was still insufficient to resolve all relationships among species of *Hemsleya*. Some subclades did not receive bootstrap support higher than 80%.

There may be two reasons for the unresolved relationships in our phylogenetic study of *Hemsleya*. First, the genus may be of relatively recent origin. The genus *Hemsleya* may have originated at the transition of the Paleogene and Neogene (Li 1993). The rather low values in pairwise divergence (about 0.55%) among subg. *Hemsleya* and the short branches in one of the two most parsimonious trees (Fig. 2) also suggests that subg. *Hemsleya* may be a more recent split from subg. *Graciliflora*.

The other possible explanation is interspecific hybridization and polyploidy. Hybridization could lead to chloroplast capture among taxa. Consequently, this would be detectable as incongruence between nuclear and chloroplast phylogenies. However, the *trn*H-*psb*A, *rpl*16 intron, and *trn*L region had few informative sites and produced poorly resolved trees; thus they are unable to provide clear evidence of hybrid origins. *Hemsleya* can be easily hybridized in cultivation (Li 1993), however, there is no empirical evidence for the hybrid origin of any of the species in the wild. Although the current geographic distribution and isolation of most *Hemsleya* species tends to discourage hybridization (with the exception of the widespread *H. graciliflora*), dioecism and xenogamy of all species tends to favor outcrossing; thus the chance for hybridization may exist. Moreover, plants of *H. turbinata*, *H. macrocarpa*, *H. chengyihana*, and *H. mitrata* are distributed in adjacent areas, increasing the potential for hybridization among those species. The variable somatic chromosome numbers, ranging from $4n = 28$ (tetraploid) to $6n = 42$ (hexaploid) (Li 1993), could be indicative of hybridization and polyploidy. However, polyploidy has not been evaluated in *Hemsleya* species by experimental studies.

In both the MP and BA analysis of the combined datasets, we recovered topologies in which *H. graciliflora* was the basal-most group of *Hemsleya* (Fig. 1). These topologies are in agreement with the subgeneric division of the genus by Li (1993). Li (1993) defined *Hemsleya* with two subgenera, *Hemsleya* and *Graciliflorae*, mainly on the basis of the presence or absence of tubers, respectively. The single species of subg. *Graciliflorae* (*H. graciliflora*), widespread at altitudes below 800 m, was found to be sister to all other species of subg. *Hemsleya*. Based on the morphology of flowers and seed, subgenus *Hemsleya* is divided into three sections: sect. *Delavayanae* (with patelliform corollas and suberose seed-wings), sect. *Amabiles* (with radiate corollas), and sect. *Hemsleya* (with patelliform, umbraculiferous, or spherical corollas and no seed-wings). At the sectional level, the topologies partly corresponded to the classification of Li (1993). However, species of sect. *Amabiles* and sect. *Hemsleya* are nested in clades A, B, and C, implying that both sections are polyphyletic. For instance, in clade C, *H. turbinata* (sect. *Hemsleya*) was resolved as sister to *H. macrocarpa* (sect. *Amabiles*) with BS of 100%. In clade A, *H. zhejiangensis* (sect. *Hemsleya*) and *H. dipterygia* (Sect. *Amabiles*) were resolved as sister species with BS of 83%. The type species of sect. *Amabiles*, *H. amabilis*, is clustered with most species of sect. *Hemlseya*, including the type species *H. chinensis* in clade B, but with low BS support. At the subsectional levels, molecular data only weakly supported the traditional classification of Li (1993). Clades B and C reflected subsect. *Chinenses* and subsect. *Carnosiflorae* respectively, however, the members of subsections were not the same as Li's classification.

However, while the molecular tree does not completely agree with the current morphological classification, the topology correlates well with the geographical distribution of the species. Floristic divisions of Wu and Wu (1998)

suggest that the forest area in the East Asiatic Floristic Kingdom can be divided into three subkingdoms: the Sino-Japanese Forest subkingdom, the Sino-Himalayan Forest subkingdom, and the Qinghai-Xizang Plateau subkingdom. Each clade in our analysis occupies a distinct geographical distribution, isolated from other clades (Fig. 3). All species of clades B and C are distributed in the Sino-Himalayan Forest subkingdom, while those in clade A occur in the Sino-Japanese Forest subkingdom, with the exception of *H. dulongjiangensis* in northwestern Yunnan (part of the Sino-Himalayan Forest subkingdom).

All of the basal clades occur in the Sino-Himalayan Forest subkingdom, including the basal-most species, *H. graciliflora* (which is distributed across the two subkingdoms). It is therefore likely that the eastern part of the Sino-Himalayan Forest subkingdom, in which more than 70% of the genus is distributed (Li 1993), is the center of diversification and speciation of *Hemsleya*. This area was located in the Kangdian ancient landmass (Zhou 1984). Furthermore, there has been no great geologic and climatic change in the Kangdian ancient landmass since the Tertiary (Zhou 1984; Wu 1984, unpublished), so this could also be the center of origin of *Hemsleya*. This conclusion is in agreement with Li (1993) and Chen (1995). With a southeast to northeast Yunnan and south Sichuan continuum within clades B and C, respectively, along with several distribution zones of the Sino-Japanese Forest subkingdom, a bidirectional expansion from the Kangdian ancient landmass may have occurred. If clades B and C truly are the sister group of clade A, this may indicate that diversification in the Sino-Himalayan Forest subkingdom is parallel to diversification in the Sino-Japanese Forest subkingdom. However, diversification in the Sino-Himalayas is higher as a result of the much stronger influence of the uplift of the Himalayas. It seems, therefore, that geography may be a better indicator of phylogenetic relationships within the genus than habit, flower morphology, or seed morphology.

The analyses of ITS, *trn*H-*psb*A, *rpl*16, and *trn*L sequences provide the first molecular phylogeny of the genus *Hemsleya* and offer insights into the relationships among its species. However, the gene regions used were unable to resolve all phylogenetic relationships within *Hemsleya*. Thus, additional studies utilizing gene regions with higher rates of nucleotide substitutions are needed to fully resolve and support the species relationships. Nevertheless, this study provides new interpretations of the systematics and evolutionary relationships within the genus

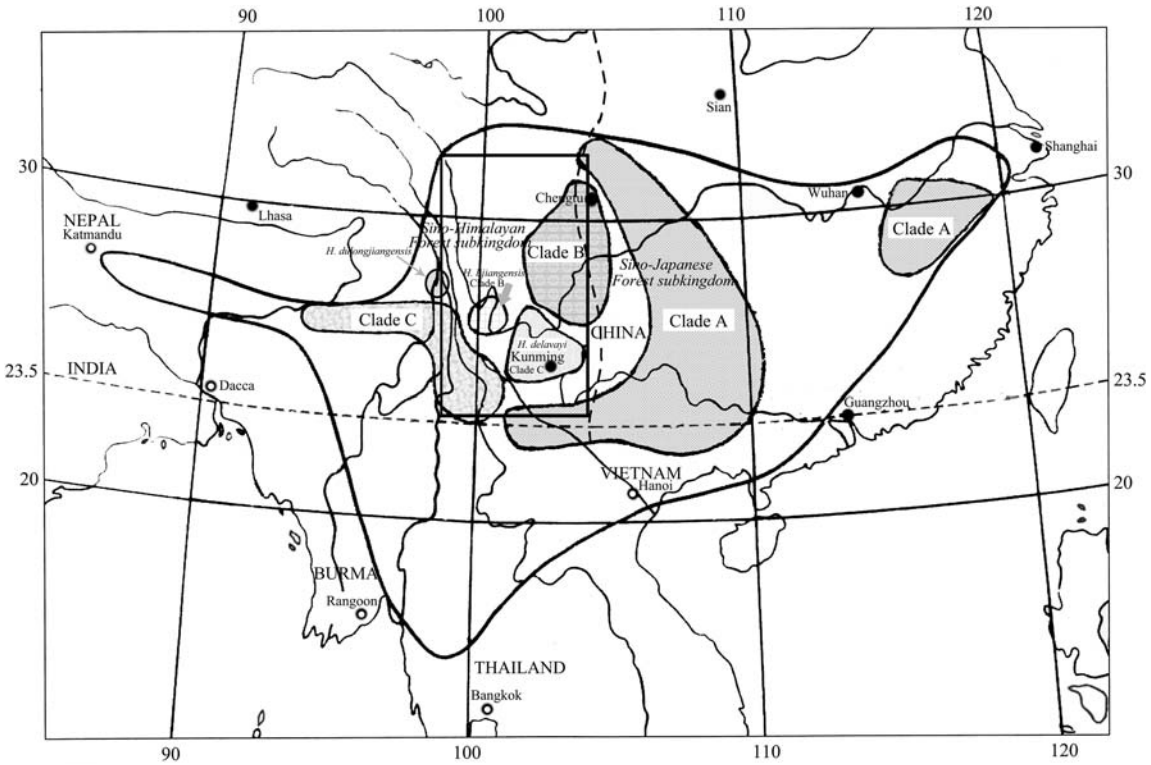

Fig. 3 The geographical distribution of *Hemsleya*. Map showing the distribution of clades. *Black dashed line* divides the Sino-Himalayan Forest subkingdom and Sino-Japanese Forest subkingdom. The *rectangle* represents the diversity and speciation centers of *Hemsleya* and also the main body of the Kangdian ancient landmass

Hemsleya and demonstrates that geography, habitat, and other ecological factors may predominantly act as vital driving forces in the evolution of *Hemsleya*.

Acknowledgments The research was supported by the National Basic Research Program of China (973 program, grant no. 2007CB411601) and the National Natural Science Foundation of China (grant 40830209). The authors are extremely grateful to J.K. Triplett (Iowa State University), Dr. Lian-Ming Gao, Dr. Ting-Shuang Yi, and Dr. Jin-Mei Lu (KIB) for their valuable comments to help improve the manuscript.

References

Bremer B, Jansen RK, Oxelman B, Backlund M, Lantz H, Kim KJ (1999) More characters or more taxa for a robust phylogeny: case study from the coffee family (Rubiaceae). Syst Biol 48:413–435

Burland TG (2000) DNASTAR's Lasergene sequence analysis software. Methods Mol Biol 132:71–91

Chen S-K (1995) A classificatory system and geographical distribution of the genus *Gynostemma* Bl (Cucurbitaceae). Acta Phytotax Sin 33:403–410

Chen Y, Chiu M-H, Gu K, Li Z-R (2003) Cucurbitacin and triterpenoid glycosides from *Hemsleya gigantha*. Chin Chem Lett 14:475–478

Chiu M-H, Gao J (2003) Three new cucurbitacins from *Hemsleya lijiangensis*. Chin Chem Lett 14:389–392

Chung SM, Decker-Walters DS, Staub JE (2003) Genetic relationships within the Cucurbitaceae as assessed by ccSSR marker and sequence analyses. Can J Bot 81:814–832

Chung SM, Staub JE, Chen JF (2006) Molecular phylogeny of *Cucumis* species as revealed by consensus chloroplast SSR marker length and sequence variation. Genome 49:219–229

Clarke AC, Burtenshaw MK, McLenachan PA, Erickson DL, Penny D (2006) Reconstructing the origins and dispersal of the polynesian bottle gourd (*Lagenaria siceraria*). Mol Biol Evol 23:893–900

Decker-Walters DS, Walters TW, Posluszny U, Kevan PG (1990) Genealogy and gene flow among annual domesticated species of *Cucurbita*. Can J Bot 68:782–789

Dinan L, Whiting P, Sarker SD, Kasai R, Yamasaki K (1997) Cucurbitane-type compounds from *Hemsleya carnosiflora* antagonize ecdysteroid action in the *Drosophila melanogaster* BII cell line. Cell Mol Life Sci 53:271–274

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:11–15

Farris JS, Källersjö M, Kluge AG, Bult C (1994) Testing significance of incongruence. Cladistics 10:315–319

Farris JS, Källersjö M, Kluge AG, Bult C (1995) Constructing a significance test for incongruence. Syst Biol 44:570–572

Forbes FB, Hemsley WB (1888) Index florae sinensis. J Linn Soc Lond Bot 23:490

Ganal M, Hemleben V (1986) Comparison of the ribosomal RNA genes in four closely related *Cucurbitaceae*. Pl Syst Evol 154:63–77

Garcia-Mas J, Monforte AJ, Arús P (2004) Phylogenetic relationships among *Cucumis* species based on the ribosomal internal transcribed spacer sequence and microsatellite markers. Pl Syst Evol 248:191–203

Graybeal A (1998) Is it better to add taxa or characters to a difficult phylogenetic problem? Syst Biol 47:9–17

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser 41:95–98

Hano Y, Shi Y-Q, Taro Nomura, Yang P-Q, Chang W-J (1997) Two acetogenins from *Hemsleya ellipsoidea*. Phytochemistry 46:1447–1449

Hillis DM (1998) Taxonomic sampling, phylogenetic accuracy, and investigator bias. Syst Biol 47:3–8

Hillis DM, Bull JJ (1993) An empirical-test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst Biol 42:182–192

Huelsenbeck JP, Ronquist F (2001) MrBayes: Bayesian inference of phylogeny. Bioinformatics 17:754–755

Jarret RL, Newman M (2000) Phylogenetic relationships among species of *Citrullus* and the placement of *C. rehmii* De Winter as determined by internal transcribed spacer (ITS) sequence heterogeneity. Genet Resour Crop Evol 47:215–222

Jeffrey C (1980a) A review of the Cucurbitaceae. Bot J Linn Soc 81:233–247

Jeffrey C (1980b) The Cucurbitaceae of Eastern Asia. Royal Botanic Gardens, Kew

Jeffrey C (1990) Systematics of the Cucurbitaceae: an overview. In: Bates DM, Robinson RW, Jeffrey C (eds) Biology and utilization of the Cucurbitaceae. Cornell University Press, Ithaca, p 3–28

Jeffrey C (2005) A new system of Cucurbitaceae. Bot Zhurn 90:332–335

Jobst J, King K, Hemleben V (1998) Molecular evolution of the internal transcribed spacers (ITS1 and ITS2) and phylogenetic relationships among species of the family Cucurbitaceae. Mol Phylogenet Evol 9:204–219

Jordan WC, Courtney MW, Neigel JE (1996) Low levels of intraspecific genetic variation at a rapidly evolving chloroplast DNA locus in North American duckweeds (Lemnaceae). Am J Bot 83:430–439

King K, Torres RA, Zentgraf U, Hemleben V (1993) Molecular evolution of the intergenic spacer in the nuclear ribosomal RNA genes of Cucurbitaceae. J Mol Evol 36:144–152

Kocyan A, Zhang L-B, Schaefer H, Renner SS (2007) A multi-locus chloroplast phylogeny for the Cucurbitaceae and its implications for character evolution and classification. Mol Phylogenet Evol 44:553–577

Kress WJ, Erickson DL (2007) A two-locus global DNA barcode for land plants: the coding rbcL gene complements the non-coding trnH-psbA spacer region. PLoS ONE 2(6): e508. doi: 10.1371/journal.pone.0000508

Kress JW, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH (2005) Use of DNA barcodes to identify flowering plants. Proc Natl Acad Sci USA 102:8369–8374

Li D-Z (1993) Systematics and evolution of *Hemsleya* (Cucurbitaceae). Yunnan Science and Technology Press, Kunming

Lin Y-P, Chiu M-H, Li Z-R (2003) Two new compounds from *Hemsleya penxianensis* var. *gulinensis*. Chin Chem Lett 14:169–172

Lu A-M (1982) Material for the genus *Hemsleya* Cogn. Acta Phytotax Sin 20:87–90

Montes-Hernández S, Eguiarte LE (2002) Genetic structure and indirect estimates of gene flow in three taxa of *Cucurbita* (Cucurbitaceae) in western Mexico. Am J Bot 89:1156–1163

Ng TJ (1993) New opportunities in the Cucurbitaceae. In: Janick J, Simon JE (eds) New crops. Wiley, New York, pp 538–546

Nie R-L, Chen Z-L (1986) The research history and present status on the chemical components of genus *Hemsleya* (Cucurbitaceae). Acta Bot Yunnan 8:115–124

Nie R-L, Kasai R, Kasai R (1984) Saponins from Chinese medical plants. I. Isolation and structures of hemsloisides. Planta Med 50:322–327

Posada D, Crandall KA (1998) Modeltest: testing the model of DNA substitution. Bioinformatics 14:817–818

Qiu M-H, Chen S-K, Chen J-C, Zhou L, Li Z-R, Nie R-L (2005) Chemotaxonomy of Cucurbitaceae. Chin J Appl Environ Biol 11:673–685

Rannala B, Yang ZH (1996) Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. J Mol Evol 43:304–311

Reeves G, Chase M, Goldblatt P, Rudall P, Fay M, Cox A, Lejeune B, Chies TDE (2001) Molecular systematics of Iridaceae: evidence from four plastid DNA regions. Am J Bot 88:2074–2087

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA-polymerase. Science 239:487–491

Sang T, Crawford DJ, Stuessy TF (1997) Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). Am J Bot 84:1120–1136

Sanjur OJ, Piperno DR, Andres TC, Wessel-Beaver L (2002) Phylogenetic relationships among domesticated and wild species of *Cucurbita* (Cucurbitaceae) inferred from a mitochondrial gene: implications for crop plant evolution and areas of origin. Proc Natl Acad Sci USA 99:535–540

Shaw J, Lickey EB, Beck JT, Farmer SB, Liu W, Miller J, Siripun KC, Winder CT, Schilling EE, Small RL (2005) The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. Am J Bot 92:142–166

Simmons MP, Ochoterena H (2000) Gaps as characters in sequence based phylogenetic analyses. Syst Biol 49:369–381

Soltis DE, Soltis PS, Mort ME, Chase MW, Savolainen V, Hoot SB, Morton M (1998) Inferring complex phylogenies using parsimony: an empirical approach using three large DNA data sets for angiosperms. Syst Biol 47:32–42

Suzuki Y, Glazko GV, Nei M (2002) Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. Proc Natl Acad Sci USA 99:16138–16143

Swofford DL (2001) PAUP*. Phylogenetic analysis using parsimony and other methods 4.0b10. Sinauer Associates, Sunderland

Taberlet P, Gielly L, Patou IG, Bouvet J (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol Biol 17:1105–1109

Tate JA, Simpson BB (2003) Paraphyly of *Tarasa* (Malvaceae) and diverse origins of the polyploid species. Syst Bot 28:723–737

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The Clustal_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882

Vischi M, Arzenton F, De Paoli E, Paselli S, Tomat E, Olivieri AM (2006) Identification of wild species of sunflower by a specific plastid DNA sequence. Helia 45:11–18

White TJ, Bruns T, Lee S, Taylor J (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis M, Gelfand D, Sninsky J, White T (eds) PCR protocols: a guide to methods and application. Academic Press, San Diego, pp 315–322

Wiens JJ (1998) Combining data sets with different phylogenetic histories. Syst Biol 47:568–581

Wilcox TP, Zwickl DJ, Heath TA, Hillis DM (2002) Phylogenetic relationships of the dwarf boas and a comparison of Bayesian and bootstrap measures of phylogenetic support. Mol Phylogenet Evol 25:361–371

Wilson HD, Doebley J, Duvall M (1992) Chloroplast DNA diversity among wild and cultivated members of *Cucurbita* (Cucurbitaceae). Theor Appl Genet 84:859–865

Wortley AH, Rudall PJ, Harris DJ, Scotland RW (2005) How much data are needed to resolve a difficult phylogeny? Case study in Lamiales. Syst Biol 54:697–709

Wu C-Y, Chen Z-L (1985) Materia and floram Cucurbitacearum Sinensium–*Hemsleya* Cogn. Acta Phytotax Sin 23:121–143

Wu C-Y, Chen Z-L (1986) Flora reipublicae popularis sinicae, vol 73. Science Press, Beijing, pp 102–129

Wu C-Y, Wu S-G (1998) A proposal for a new floristic kingdom (realm): the E. Asiatic kingdom, its delimitation and characteristics. In: Zhang AL, Wu SG (eds) Proceedings of the first international symposium on floristic characteristics and diversity of East Asian plants. China Higher Education Press/Springer, Beijing/Berlin, pp 3–42

Wu J, Wu Y, Yang B-B (2002) Anticancer activity of *Hemsleya amabilis* extract. Life Sci 71:2161–2170

Yang Y-K, Chiu M-H, Gao C-W, Nie R-L, Lu Y, Zheng Q-T (2000) The novel structure of a disaccharide derivative from *Hemsleya amabilis*. Tetrahedron 56:7433–7435

Yoder AD, Irwi JA, Payseur BA (2001) Failure of the ILD to determine data combinability for slow loris phylogeny. Syst Biol 50:408–424

Zhang L-B, Simmons MP, Kocyan A, Renner SS (2006) Phylogeny of the Cucurbitales based on DNA sequences of nine loci from three genomes: implications for morphological and sexual system evolution. Mol Phylogenet Evol 39:305–322

Zhou T-R (1984) Physico-geography of China: palaeogeography, vol 1. Science Press, Beijing