

Phylogeny and biogeography of Balsaminaceae inferred from ITS sequences

Yong-Ming Yuan^{1,5}, Yi Song¹, Koen Geuten², Elisette Rahelivololona³, Sébastien Wohlhauser¹, Eberhard Fischer⁴, Erik Smets² & Philippe Küpfer¹

¹ Institut de Botanique, Université de Neuchâtel, Emile-Argand 11, CH-2007 Neuchâtel, Switzerland. yong-ming.yuan@unine.ch (author for correspondence)

² Laboratory of Plant Systematics, K.U. Leuven, B-3001 Leuven, Belgium. koen.geuten@bio.kuleuven.ac.be

³ Parc Botanique et Zoologique de Tsimbazaza, B.P. 4096, Antananarivo, Madagascar. prota.madagascar@dts.mg

⁴ Institut für Biologie, Universität Koblenz-Landau, Universitätsstr. 1, D-56070 Koblenz, Germany. efischer@uni-koblenz.de

⁵ South China Institute of Botany, Chinese Academy of Sciences, Guangzhou, P. R. China. yong-ming.yuan@univie.ch

Sequences of the internal transcribed spacers (ITS) of nuclear ribosomal DNA were acquired for 112 species of Balsaminaceae worldwide and five species of its closest relatives Marcgraviaceae and Tetrameristaceae. Phylogenetic analyses applying parsimony and distance estimates confirmed the monophyly of Balsaminaceae and suggest the monophyly of *Impatiens*. Within *Impatiens*, a few clades are recognized with strong support. Two of the most important clades are the spurless Madagascan endemic group, and the one comprising species with broadly fusiform fruits and the basic chromosome number $x = 8$, that shows a Southeast Asia, southern India, Africa, and Madagascar connection. Despite recognition of several strongly supported small lineages, ITS data alone could not resolve relationships among most of the lineages with confident support values. ITS phylogenies are therefore of limited taxonomic value for *Impatiens*. However, ITS phylogenies do reveal that extant *Impatiens* species are of Southeast Asian origin, from where dispersals to boreal Eurasia and North America, to central Asia and eastern Europe via the Himalayas, and to India and Africa have occurred. The Madagascan *Impatiens* show an African origin. Molecular phylogenies suggest the ancestral basic chromosome number to be $x = 10$, and the spurred flowers and elongated linear fruits to be plesiomorphic states in *Impatiens*. A predominantly descending dysploid chromosome evolution, following dispersal of the clade with broadly fusiform fruits from Southeast Asia to India, Africa, and Madagascar, is also suggested.

KEYWORDS: Balsaminaceae, biogeography, chromosome evolution, ITS, phylogeny.

INTRODUCTION

Balsaminaceae are a medium-sized family with two genera, *Hydrocera* and *Impatiens*. Other generic names published for Balsaminaceae, such as *Petalonema* Peter, *Semeiocardium* Zoll. and *Impatientella* H. Perrier, are confirmed to be synonyms of *Impatiens* (Grey-Wilson, 1989a; Rao & al., 1986). *Hydrocera* has only one species, *H. triflora*, a semi-aquatic herb native to the Indo-Malesian countries. Morphologically, it can be easily distinguished from *Impatiens* by its five free petals and the indehiscent berry-like fruit. *Impatiens* is a highly diversified genus with over 900 species distributed primarily in the highlands and mountains of the Old World tropics and subtropics with five conspicuous diversity hotspots: tropical Africa, (ca. 109 spp.; Grey-Wilson, 1980b), Madagascar (ca. 120 spp.; Fischer & Rahelivololona, 2002), southern India and Sri Lanka (ca. 150 spp.), the eastern Himalayas (ca. 120 spp.), and Southeast Asia in its broad sense (including Burma,

Thailand, southwest China, the Indochina peninsula, and the Malesian archipelagos, ca. 250 spp.). Many new species are still being described from these regions (e.g., Chen, 2000; Shimizu, 2000; Fischer & Rahelivololona, 2002; Fischer & al., 2003; Huang & al., 2003). High proportions of local endemism are associated with these hotspots, for example, as many as 91% of the southern Indian species are endemic (Rao & al., 1986), and almost all the native species of Madagascar are endemic. Contrary to paleotropical areas, only a few *Impatiens* species are found in temperate areas of the northern hemisphere. There are no native species in South America or Australia. Some *Impatiens* species are of horticultural importance, e.g., the popularly grown flowers, the “Balsam”, *I. balsamina*, the “Busy Lizzie”, *I. walleriana*, and the “New Guinea Hybrid *Impatiens*”, *I. hawk-eri*.

Impatiens is a well-known example of a taxonomically difficult group (Hooker & Thompson, 1859; Grey-Wilson, 1980b). The delicate yet hypervariable structure

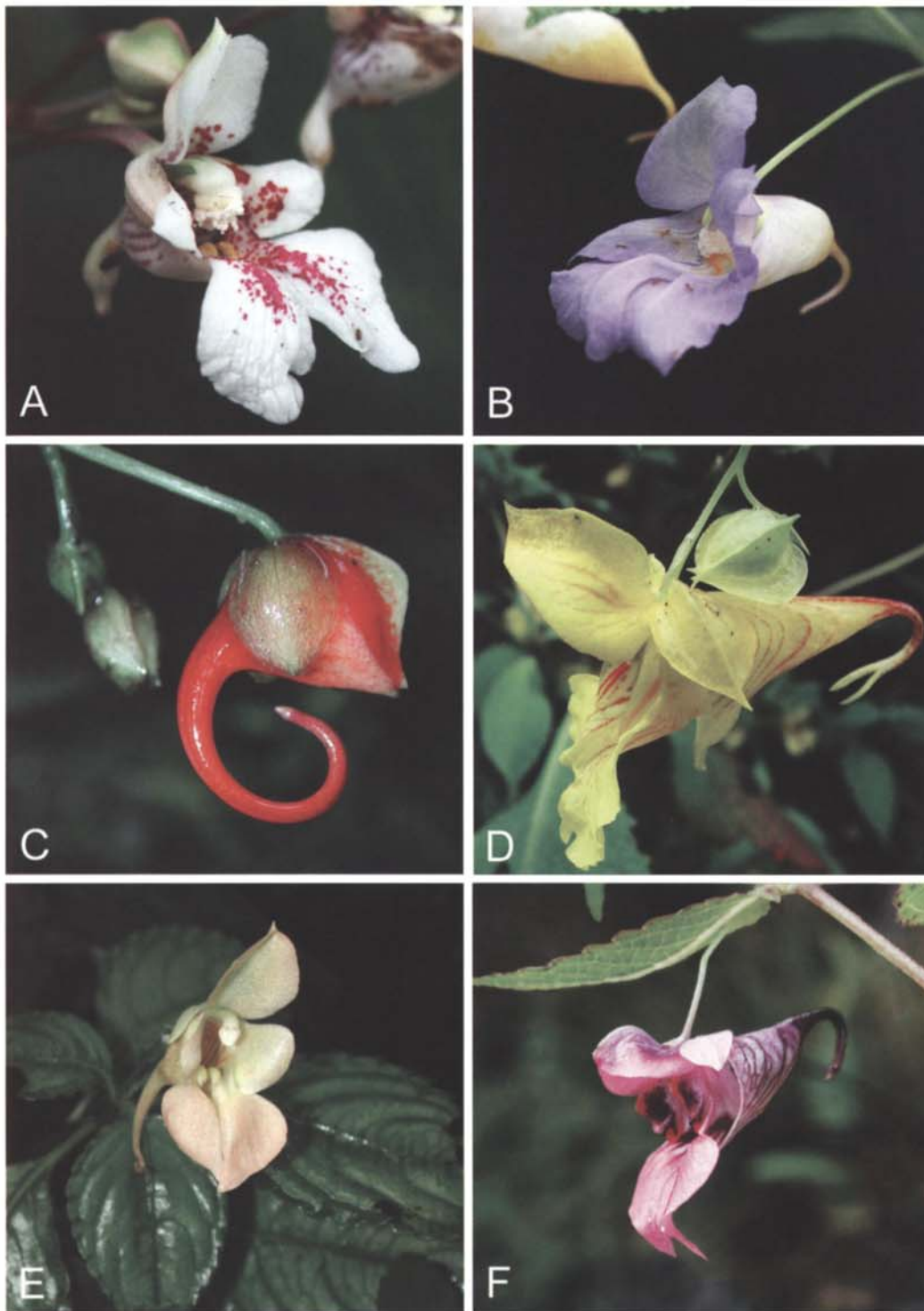


Fig. 1. *Impatiens* shows tremendous diversity in floral colour and morphology, and can be considered as the “dicots counterpart of orchid”. Its flower is zygomorphic and resupinate through twisting of the pedicel. Each flower has three sepals, two of which are reduced lateral sepals and one petal-like lower sepal that is modified into a nectary-tipped spur that exhibits a wide range of variation in form and size, from shallowly navicular to bucciniform or deeply saccate, short or long filiform, straight, twisted or curved. Each flower has five petals. One upper dorsal petal is usually hood-like. The other four lower petals are united into two lateral pairs, yet each retains two unequal lobes. The united lateral petals are also extremely variable in shape and size, associated with different pollinators as they provide a suitable landing platform and entrance guide for pollinators to the spur and nectar. The five stamens are united by the upper part of the filaments and completely cap the gynoecium, which has five fused carpels. Shown here are examples of flowers of *Impatiens* of different colours. A, *I. campanulata*; B, *I. arguta*; C, *I. yingjiangensis*; D, *I. soulieana*; E, *I. mengtseana*; F, *I. delavayi*.

and fragile nature of its flowers (Fig. 1) make it almost impossible to determine a species when specimens are pressed conventionally without detailed descriptions or drawings of its floral morphology (Grey-Wilson, 1980c). Many early publications based on herbarium specimens are therefore incomplete or misleading. The only global infrageneric classification of *Impatiens* was the early work of Warburg & Reiche (1895), which recognized 14 sections. However, this classification was considered neither natural nor practical (Grey-Wilson, 1980b). Thus a reliable infrageneric classification is still missing. Although aggregates of closely allied species can be recognized, it is hard to unravel the phylogenetic relationships among the species aggregates on the basis of gross morphology. These difficulties have necessarily limited most studies on *Impatiens* so far to be regional and purely descriptive taxonomic treatments (e.g., Perrier de la Bathie, 1934, 1948; Humbert, 1956; Chen, 1978; Grey-Wilson, 1980a, 1985, 1989a; Akiyama & al., 1992a, 1995, 1996; Akiyama & Ohba, 2000; Fischer & Rahelivololona, 2002).

Impatiens shows a wide range of chromosome number variation of $2n = 6, 8, 10, 12, 14, 15, 16, 17, 18, 19, 20, 24, 26, 28, 30, 32, 34, 36, 40, 44, 48, 50, 54,$ and 66 (Song & al., 2003). The most frequent numbers are $2n = 14, 16, 18,$ and 20 , that take the major part (78%) of the species observed. Different hypotheses have been proposed regarding the evolution of the basic numbers through dysploidy in conjunction with polyploidy in *Impatiens*. Jones & Smith (1966) and Akiyama & al. (1992b) suggested $x = 7$ to be the ancestral type, from which the other numbers were derived mainly by ascending dysploidy, whereas Rao & al. (1986) suggested evolution of $x = 7, 9$ and 10 from the basic number $x = 8$ through both descending and ascending dysploidy. Phylogenetic studies may offer independent test on the evolution of the basic chromosome numbers.

Balsaminaceae show an interesting distributional pattern. While *Hydrocera* is an Indo-Malesian genus, *Impatiens* basically exhibits prominent African-Malagasy/Indo-Himalayan-Southeast Asian phytogeographic connection, since its diversity centers are all located along this chain. Whereas the species in northern temperate regions probably represent recent radiations from Southeastern Asia and the adjacent Sino-Himalayan areas, it is still questionable whether the disjunctive distribution of the main diversity centers across the paleotropical regions represents ancient vicariant events resulting from the fragmentation of Gondwana, or more recent dispersals. Jones & Smith (1966) suggested that *Impatiens* originated in the Himalayan region and dispersed to other areas based on species diversity and karyological data. On the contrary, Grey-Wilson (1980b) formulated an hypothesis suggesting that Balsaminaceae

originated in western Gondwana in the Paleogene ca. 50 million years ago (Ma), and subsequently spread to Southeast Asia through Madagascar and India at the time of or after the Indian plate collided with Laurasia (ca. 45 Ma). He rejected the possibility of an overland migration between Africa and India. These conclusions were based on the pronounced similarity of species among Africa, Madagascar, and southern India, as well as on the assumed sister relationship between Balsaminaceae and Tropaeolaceae. The distributional patterns of extant Balsaminaceae, however, suggest that the diversification of the family probably started in South or Southeast Asia instead of Africa, as *Hydrocera* occurs only in these regions. Assuming a Southeast Asian origin of *Impatiens* and a subsequent radiation to Africa and Madagascar through India can equally result in the pronounced similarity of species among India, Africa, and Madagascar that was correctly recognized by Grey-Wilson (1980b). Obviously, the competing biogeographic hypotheses can be tested by using the phylogenies of the species from the main diversification centers.

Despite the conspicuous diversity and the interesting distributional patterns shown by Balsaminaceae, few molecular phylogenetic studies have been made on this family. Several studies aimed at addressing phylogenetic relationships at the family level or above have included limited representative species (not more than five) of Balsaminaceae. These studies have confirmed Balsaminaceae as a member of the order Ericales at the base of the asterids (Morton & al., 1996, 1997; Soltis & al., 2000; Albach & al., 2001; Anderberg & al., 2002; Bremer & al., 2002; Geuten & al., 2004), instead of the order Geraniales of former Rosidae as traditionally considered (Cronquist, 1981). Recently, 25 species from the eastern Himalayan area have been subjected to a molecular phylogenetic study using chloroplast *rbcL* and *trnL-F* sequences (Fujihashi & al., 2002). Besides its inappropriate analysis (e.g., too distant outgroups from rosids were used), this study is sketchy due to its limited sampling, and the resulted phylogenies are incapable of addressing the questions about overall phylogeny and biogeography of the whole family as mentioned above. Global phylogenetic relationships within the family remain unknown, and the more general problems such as diversification patterns of floral and vegetative morphology and historical biogeography of the family need to be addressed through more comprehensive global studies of the entire family. Here we conducted a comprehensive molecular phylogenetic study on Balsaminaceae by using nucleotide sequence data of internal transcribed spacer regions of nuclear ribosomal DNA. Through the molecular phylogenies we intended to further examine the morphological and karyological evolution, as well as the historical biogeography, of the family.

MATERIALS AND METHODS

Ingroup sampling and outgroup choice. — We sampled both genera of Balsaminaceae, *Hydrocera* and *Impatiens*. As far as possible, samples were selected to maximize representation of the whole distribution range and the diversity of *Impatiens*. Representatives from all the hot-spots of species diversity of the genus were sampled. Representatives of the closely related families, Marcgraviaceae and Tetrameristaceae sensu APG (2003), were sampled as outgroups, following the results revealed by the recent molecular phylogenetic studies on large scope (Soltis & al., 2000; Albach & al., 2001; Anderberg & al., 2002; Geuten & al., 2004). The species, origin of samples, voucher information, and GenBank sequence accessions are listed in the Appendix (see online version of *Taxon*). Our data matrix includes 117 taxa in total, of which five taxa of Tetrameristaceae and Marcgraviaceae were used as outgroups. A reduced analysis limited to Balsaminaceae (112 taxa in total) was also conducted, to reveal possible consequences caused by high divergence among the three families. In the latter case, *Hydrocera triflora* was considered as an outgroup.

DNA extraction, PCR amplification and sequencing. — Total DNA was extracted from fresh, silica-gel dried, or herbarium leaves with the CTAB method of Doyle & Doyle (1987) or the DNeasy Plant Mini Kit (QIAGEN AG, Basel). The ITS fragment was amplified via standard PCR in 25 µl reaction volume as described in Yuan & al. (2003). Successfully amplified DNA fragments were purified prior to sequencing using the QIAquick™ PCR purification kit (QIAGEN AG, Basel) following the manufacturer's protocol. Cycle sequencing reactions were performed using the dye-terminator chemistry as implemented in the ABI PRISM® BigDye™ Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems) in a Biometra thermal cycler. The sequencing products were cleaned using the Ethanol/Sodium Acetate precipitation method and then analyzed on an ABI310 automated sequencer (Applied Biosystems). Automation-generated base-calls were edited manually against the electropherograms using the software Sequence Navigator (Applied Biosystems, Foster City, U.S.A.). Alternatively, some sequences were obtained via the methods described in Geuten & al. (2004).

Sequence alignment. — The obtained ITS sequences were initially aligned with Clustal X applying the default parameters (Thompson & al., 1997) and then manually adjusted for indels otherwise not properly recognized by Clustal. Careful manual adjustment was necessary, and three regions that seem to correspond to loops in secondary rRNA structure (Denduangboripant &

Cronk, 2001) involve high alignment ambiguity. These regions (103 sites in the aligned data matrix) were then excluded from subsequent phylogenetic analyses. The gaps of aligned ITS sequences were considered as missing data.

Maximum parsimony (MP) analysis. — For MP analysis, the dataset was analyzed with heuristic searches by using PAUP* v4.0b10 (Swofford, 2000). Characters were equally weighted and unordered. Branch collapse option was set to collapse if minimum length was zero. Heuristic searches were conducted in three steps. First search was made to obtain an empirical tree length. Subsequently, heuristic searches were conducted for 1000 replicates of random addition of sequences, with TBR branch swapping, ACCTRAN, MULTREES option on, STEEPEST DESCENT option off, and from each replicate a maximum of 100 trees saved. Finally, a TBR branch swapping was conducted on all the best trees found in the previous step to save a maximum of 100,000 trees. The optimal trees kept by this swapping were then filtered out and a consensus was calculated. Relative clade support was evaluated by bootstrap analyses (Felsenstein, 1985). Bootstrap values were calculated by using 500 replicates of heuristic searches, with random sequence addition, TBR branch swapping, MULTREES options on, the STEEPEST DESCENT option off, and a maximum of 1000 trees saved for each replicate.

Neighbor-joining (NJ) analysis. — As a comparison to MP analysis, a distance analysis applying NJ optimality criteria (Saitou & Nei, 1987) was also conducted on both complete and reduced data matrices by using PAUP* v4.0b10 (Swofford, 2000). The NJ analyses applied maximum likelihood distance estimates based on the model and parameters suggested by Modeltest (3.06) using an arbitrarily chosen MP tree and the Akaike information criterion (Akaike, 1974; Posada & Crandall, 1998). Bootstrap values were obtained from 1000 replicate NJ analyses.

Character-state optimization and biogeographic analyses. — Selected characters, such as presence of flower spur, shape of fruit, and the basic chromosomal numbers, have been considered as unordered binary or multi-state characters and were optimized onto the molecular phylogeny to examine their evolution by using MacClade version 3.08 (Maddison & Maddison, 1992). The geographic distribution was also examined in the same way to trace the historical biogeography of Balsaminaceae. The NJ tree based on the reduced data set was used for the tracing. Arbitrarily chosen MP trees were also compared for the tracing. Chromosomal data were based on Song & al. (2003), and for species that have no chromosome number report the basic number was considered as missing. Seven areas of endemism

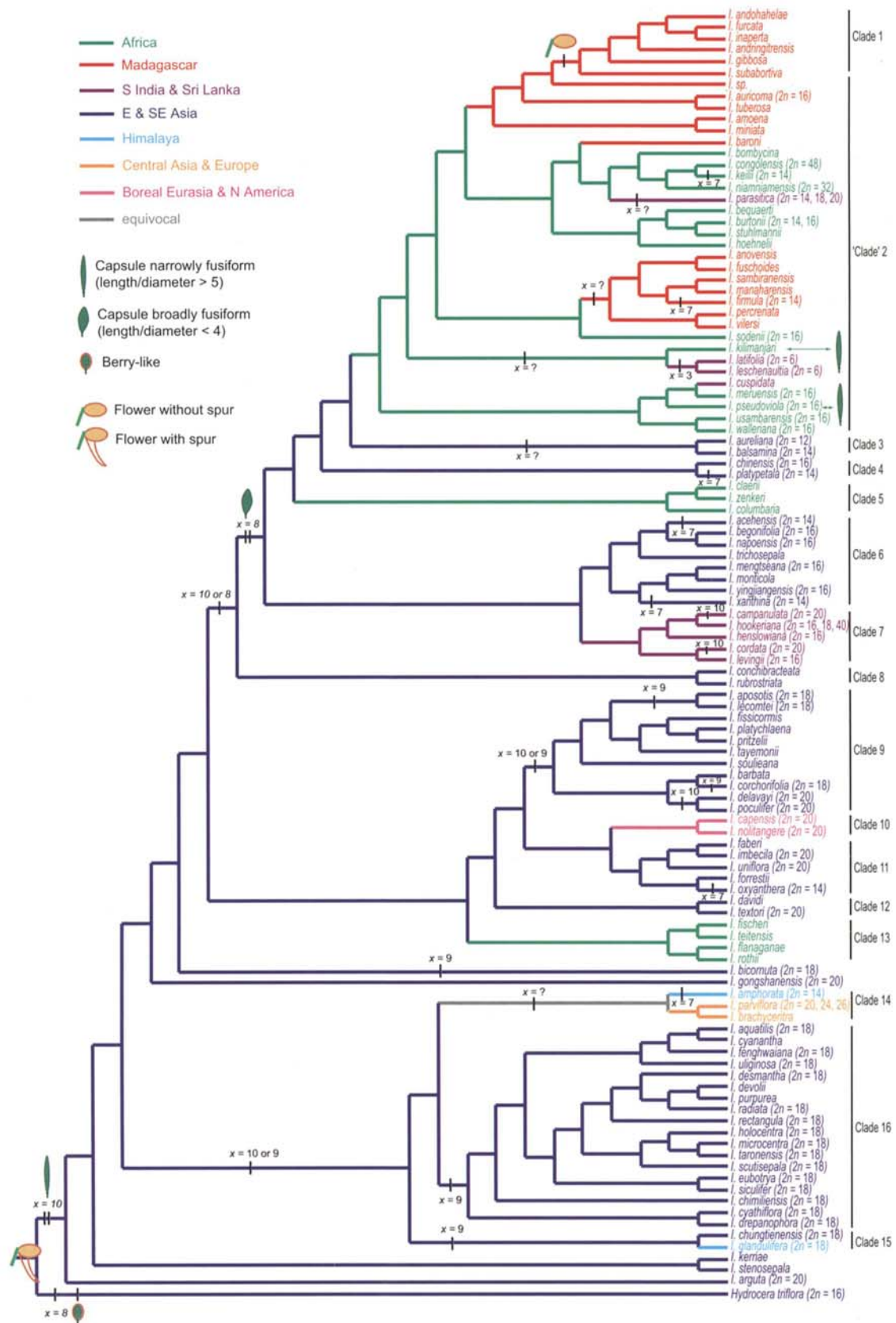


Fig. 4. Parsimonious optimization of the distribution centers of endemism, basic chromosome numbers, and selected morphological characters onto the NJ tree generated from the reduced ITS dataset. Vertical bars indicate positions of character state changes. Chromosome numbers when available are shown after the species names.

were designated on the basis of species diversity: (1) Africa, (2) Madagascar, (3) South India and Sri Lanka, (4) the Himalayas, (5) East and Southeast Asia, comprising Japan, Taiwan and the other Southeast Asian islands, the mainland Southeast Asian countries, and the area of the southern and southwestern China, (6) Central Asia and Europe, and (7) boreal Eurasia and North America. Wide distributions over more than two areas of endemism were counted as polymorphisms.

RESULTS

Sequence characteristics. — All the sequences are newly acquired and have been submitted to GenBank (Appendix). No evidence of paralogous ITS sequences was found, because all PCR products were resolved as a single band and no double peaks were encountered in sequencing. The length of the complete unaligned ITS fragments of the ingroup taxa ranged from 579 to 655 base pairs.

The full ITS dataset including ingroup and outgroup taxa consisted of 783 characters, of which 103 (13.2%) involved alignment ambiguity and were excluded from phylogenetic analyses, 180 (23.0%) are constant, 121 (15.4%) are variable but uninformative, and 379 (48.4%) are potentially informative. This dataset resulted in pairwise maximum likelihood distance estimate ranging from 0 (*I. bombycina* vs. *I. congolensis*, *I. bombycina* vs. *I. niarniamensis*, *I. congolensis* vs. *I. niarniamensis*) to 0.840 (*I. platychlaena* vs. *Pelliciera* sp.) among all taxa, and 0 to 0.542 (*I. kerriae* vs. *Hydrocera triflora*) among the ingroup taxa. The alignments were submitted to TreeBASE and are also available from the corresponding author.

ITS phylogeny. — MP analysis of the ITS dataset including all outgroup and ingroup taxa resulted in an unknown number of equally parsimonious trees (61,443 trees kept) of 2928 steps, consistency index (CI) = 0.34 including autapomorphies, CI = 0.30 excluding autapomorphies, and retention index (RI) = 0.66. The strict consensus of the optimal trees kept by heuristic searches and the bootstrap clade support (when greater than 50%) are shown in Fig. 2A. The MP strict consensus tree was moderately resolved. Despite the relative high sequence divergence, the resolution among the main lineages is poor or receives less significant support, while some lineages of closely related taxa are better resolved with strong supports. The internal branches connecting the main lineages are conspicuously shorter in comparison to the long terminal branches, suggesting a high proportion of autapomorphic variations of the sequences. The monophyly of Balsaminaceae is highly supported (bootstrap value 100%). The genus *Hydrocera* fell in a basal

polytomy with the main *Impatiens* clade and several other *Impatiens* species, viz., *I. arguta*, *I. kerriae*, and *I. stenosepala*.

We assigned numbers (clades 1 through 16) to the resolved lineages of *Impatiens* (Fig. 2A). In addition, a group of closely related non-monophyletic taxa were also referred as a clade, e.g., “clade 2”, for convenience of illustration and discussion. Noticeably, the spurless Madagascan endemic taxa were resolved as a highly supported (90%) monophyletic clade (clade 1). Together with a few species from Southeast Asia (clades 3, 4 and 6), most of the sampled African, Madagascan, and southern Indian species formed a strongly supported clade (93%), consisting of the clades 1 through 7, whereas only four African species, *I. fischeri*, *I. flanaganae*, *I. teitensis*, *I. rothii*, and the Indian *I. repens* grouped together as a weakly supported clade (52%, clade 13) that nested out of the other African species but grouped with Southeast Asian clades. The North American species, *I. capensis*, showed close affinity with the boreal temperate species, *I. nolitangere* (100%, clade 10), and both together grouped with some Southeast Asian clades (9, 11, and 12). The eastern European species, *I. parviflora*, and the central Asian species, *I. brachycentra*, grouped together with the western Himalayan species, *I. amphorata*, as a strongly supported clade (100%, clade 14), and these three species together showed affinity to some Southeast Asian and Himalayan clades (Fig. 2A).

The general-time-reversible model with invariable sites and rate heterogeneity (i.e., GTR+I+ Γ model) was suggested to best fit the ITS data based on an arbitrarily chosen MP tree (data not shown). NJ tree was reconstructed using maximum likelihood distance estimates applying the inferred model and its parameters. The topology of the neighbor-joining tree (Fig. 3A) highly resembles that of the strict consensus of the MP analyses. The monotypic genus *Hydrocera* was resolved as sister to the monophyletic *Impatiens* with moderate bootstrap support (76%). The main lineages, e.g., the clades 1 through 16 of *Impatiens*, remained almost the same as revealed by MP analyses with similar clade supports. The minor differences between NJ and MP analysis results involve different positions (Figs. 2A and 3A) of *I. bicornuta*, *I. gonshanensis*, *I. stenosepala*, clade 10 (consisting of *I. capensis* and *I. nolitangere*), and clade 15 (consisting of *I. chungtienensis* and *I. glandulifera*). However, these different resolutions did not receive significant bootstrap support in either analysis.

To determine if the relatively divergent outgroups had any potential influence on results of the phylogenetic analysis, we conducted both MP and NJ analyses on a reduced dataset that included only Balsaminaceae. The topology of the trees based on the reduced datasets strongly resembles that of trees generated from the com-

plete datasets, as shown in the simplified dendrograms (Figs. 2C, 3C). While the main topology of the trees remained the same, the positions of *I. arguta*, *I. bicornuta*, *I. gongshanensis*, *I. kerriae*, *I. stenosepala*, and clade 12 consisting of *I. davidi* and *I. textori* were slightly altered. However, none of these different resolutions receives significant support.

Character-state mapping. — Two morphological characters, occurrence of flower spur and shape of fruit, and the basic chromosome numbers were optimized onto the ITS tree generated from NJ analyses on the reduced dataset (Fig. 4). Parsimonious optimization suggests that linear-fusiform capsule, spurred flower, and basic chromosome number $x = 10$ represent the plesiomorphic states within *Impatiens*. Despite that the relationships among basal clades of the NJ tree were poorly supported, it is noteworthy that the important character-state changes suggested by the optimization correspond well to a few strongly supported nodes. For example, the broad-fusiform fruit was suggested as having evolved from linear-fusiform only once, which corresponds to the transition of basic chromosome number from $x = 10$ to $x = 8$, and the node of these transitions were highly supported by both MP and NJ analyses (bootstrap value 85–93%). The spurless flower was also suggested as a unique synapomorphic character-state for the Madagascan endemic clade 1 which received strong MP and NJ supports (90–96%). Optimization using arbitrarily chosen MP trees did not reveal any deviation of the above conclusions.

Biogeographic analysis. — Fitch parsimony optimization of centers of endemism as a multi-state character infers Southeast Asia as the ancestral area for extant Balsaminaceae (Fig. 4). *Impatiens* species in other areas are inferred to have dispersed from Southeast Asia. Two to three dispersals from Southeast Asia to Africa are suggested: one dispersal resulted in clade 13 consisting of *I. fischeri*, *I. teitensis*, etc., and another one or two dispersals involved divergence of other African species. The Madagascan *Impatiens* are inferred as having dispersed from Africa, whereas the *Impatiens* species from southern India and Sri Lanka seem to have had two origins: some were dispersed from Southeast Asia (clade 7), and others from Africa (within the assemblage “clade” 2). The central Asian and European *Impatiens* were dispersed from Southeast Asia via the Himalayas, whereas the boreal Eurasian and North American *Impatiens* are resolved as being dispersed from Southeast Asia. Optimization using the NJ tree or arbitrarily chosen MP trees revealed the same pattern.

DISCUSSION

Phylogeny of Balsaminaceae. — Recent studies, using cpDNA markers to address phylogenetic relationships at family level or above, have revealed the monophyly of Balsaminaceae and its inclusion in the order Ericales of the basal asterids (Morton & al., 1996, 1997; Soltis & al., 2000; Albach & al., 2001; Anderberg & al., 2002; Bremer & al., 2002; Geuten & al., 2004). Our present results based on nuclear ITS data are highly in agreement with these studies in concluding the monophyly of Balsaminaceae (bootstrap value 100% in both MP and NJ analyses).

At first glance, *Hydrocera triflora*, the only species of the genus, looks very much like an *Impatiens* species, in particular *I. balsamina* (Grey-Wilson, 1980d). Due to its unique floral (five free petals) and fruit (indehiscent and berry-like) morphology, *Hydrocera* is almost always considered as a genus distinct from *Impatiens* in modern botanical publications (e.g., Grey-Wilson, 1985, 1980d; Chen, 2001), despite that in some classic works such as Linnaeus' *Species Plantarum* it was considered as a species of *Impatiens*. Nevertheless, its phylogenetic relationships with the different lineages of *Impatiens* have never been explicitly studied. Since the floral morphology of *Impatiens* is highly diversified, the unique floral morphology of *H. triflora* does not necessarily ensure it as a distinct genus rather than a specialized ingroup of *Impatiens*. Our present results, however, allow the recognition of *Hydrocera* as a distinct genus sister to *Impatiens*. In the strict consensus of the MP trees, *H. triflora* falls into a basal polytomy consisting of *I. arguta*, *I. kerriae*, and *I. stenosepala*, and the clade comprising all the other *Impatiens* species (Fig. 2A), but in the NJ tree all *Impatiens* species are shown to be monophyletic with moderate support (76%) and sister to *Hydrocera* (Fig. 3A).

Implications of the molecular phylogenies on infrageneric classification of *Impatiens*. — *Impatiens* is taxonomically considered as one of the most difficult genera of angiosperms, primarily due to the delicate yet hypervariable structure and fragile nature of its flowers that are extremely difficult to examine in dried specimens if prepared conventionally (Grey-Wilson, 1980b, c). There is so far no comprehensive infrageneric classification available for the genus. The only global attempt was the early work of Warburg & Reiche (1895), in which 14 sections were recognized. However, this classification was considered neither natural nor practical (Grey-Wilson, 1980b), and was thus not followed by later authors. The only modern general treatment for the genus is the important revision of the African taxa by Grey-Wilson (1980b), which recognized six informal infrageneric groups for the African species for practical

diagnosis only. Perrier de la Bathie (1934) proposed three sections for the Madagascan *Impatiens* based on floral morphology: sect. *Preimpatiens* (= sect. *Impatiens*) (flowers with conspicuous spur), sect. *Trimorphopetalum* (flowers without spur, anthers dehiscent apically), and the monotypic sect. *Impatientella* (flowers without spur, anthers dehiscent laterally), the latter two being endemic to Madagascar. Fischer & Rahelivololona (2002) recently elevated sect. *Trimorphopetalum* to subgeneric level. The monotypic sect. *Impatientella* contains only the spurless and entirely cleistogamous species, *I. inaperta*. Our results support monophyly of the Madagascan endemic spurless species (Figs 2A and 3A: clade 1), and show that the cleistogamous *I. inaperta* is closely related to *I. furcata* of sect. *Trimorphopetalum* (or subgen. sensu Fischer & Rahelivololona, 2002). However, our molecular phylogenies show that this monophyletic clade is the most derived within *Impatiens*, and recognizing it as a distinct taxonomic identity makes the rest of the genus (taxonomically defined as section or subgenus *Impatiens*) paraphyletic. Apart from the highly supported spurless clade, molecular phylogenies also revealed a highly supported clade comprising clades 1 through 7 as shown in Figs. 2A and 3A. This clade is supported by several non-molecular characters as well, e.g., broad fusiform fruit and basic chromosome number principally $x = 8$, which have taxonomic importance.

Despite recognition of the clades mentioned above and several strongly supported small lineages (e.g., clades 5–6, 8, 10–11, 13–15), our molecular phylogenies do not resolve relationships among the lineages with confidence, and therefore offer limited taxonomic implications. Nevertheless, despite the low bootstrap support (52–64%), the African clade 13 (Figs. 2A and 3A), which nested apart from other African species, corresponds well to a species aggregate recognized by Grey-Wilson (1980b). Further phylogenetic studies, e.g., with additional molecular data from cpDNA or single-copy nuclear DNA, may improve the resolution and clade supports of the molecular phylogenies.

As pointed out by Grey-Wilson (1980b), clusters or aggregates of species are more or less clearly defined by gross morphology and geographic distribution, but the relationships among these distinguishable aggregates of species cannot be defined with confidence. Molecular phylogenies seem to show a similar pattern: small clades are recognized with strong support, but relationships among those clades are unresolved or resolved without significant support. The wide distribution of the genus, and the high morphological diversity associated with localized endemism (linked to distinguishable aggregates of species) suggest that active diversification and rapid radiation have been prevalent in the evolution of *Impatiens*. Molecular data reveal evidence of such rapid

evolution, by the relatively high sequence divergence, the noticeable difficulties in sequence alignment of certain regions, the high proportion of autapomorphic variations, and the conspicuous short internal branches connecting the main branches.

Biogeographic implications of the molecular phylogeny. — Balsaminaceae show an interesting mainly paleotropical distributional pattern, with several diversity hotspots disjunctively located in tropical Africa, Madagascar, southern India and Sri Lanka, the eastern Himalayas, and Southeast Asia. While Jones & Smith (1966) suggested that *Impatiens* originated in the Himalayan region, Grey-Wilson (1980b) formulated an hypothesis suggesting that Balsaminaceae originated in western Gondwana and spread to Southeast Asia through Madagascar and India. From Southeast Asia and the adjacent Sino-Himalayan area, *Impatiens* diversified secondarily into two lineages: one lineage radiating to the temperate Eurasian areas and North America, and the other radiating to tropical and subtropical areas of Southeast Asian islands. Meanwhile, Grey-Wilson (1980b) rejected the possibility of an overland migration between Africa and India.

Parsimonious optimization of the centers of endemism (Africa, Madagascar, southern India and Sri Lanka, the Himalayas, Central Asia and Europe, Southeast Asia, and Boreal Eurasia and North America) onto the molecular phylogenies reveal Southeast Asia as the ancestral area of extant *Impatiens* (Fig. 4). The boreal species *I. nolitangere* and North American species *I. capensis* are suggested as having dispersed from Southeast Asia. The central Asian and European species, such as *I. brachycentra* and *I. parviflora*, may have been dispersed from Southeast Asia through the Himalayas. Africa seems to have been colonized at least twice from Southeast Asia: once by a colonizer with elongated linear-fusiform fruits that led to the small clade 13, and at another time by a colonizer with broad fusiform fruits that gave rise to the others (Fig. 4). As our sampling of Indian species is still insufficient, it is not known if the colonization of Africa might have been achieved via India as stepping-stones. Nevertheless, as far as the species we sampled from India and Sri Lanka are concerned, some species (such as those of clade 7) showed Southeast Asian connection and some showed African connections (e.g., *I. cuspidata*, *I. latifolia*, *I. leschensultia*, and *I. parasitica*). Madagascan *Impatiens* shows clearly an African origin.

The historical biogeography of *Impatiens* suggested by molecular phylogenies is, in fact, in concordance with the conspicuous species similarity among Africa, Madagascar, southern India, and Southeast Asia noticed by Grey-Wilson (1980b). This similarity is shown mainly by the species with broad fusiform fruits and the basic

chromosome number of $x = 8$ (clades 1 through 7 in Fig. 4). However, our present results suggest a Southeast Asian origin of *Impatiens* and dispersals to Africa and Madagascar, whereas on the contrary, Grey-Wilson (1980b) suggested an opposite direction of dispersal from an assumed west Gondwanan (African) origin. It is worth mentioning that the present biogeographic pattern largely depends on a few nodes with strong bootstrap supports. These include the node comprising all Balsaminaceae (bootstrap 100%), the node comprising clades 1 through 8 (bootstrap 82–86%), and the node comprising clades 1 through 7 (bootstrap 85–97%) (Figs. 2A and 3A).

Morphological and chromosomal evolution.

— The fascinating floral diversity of *Impatiens* makes it an ideal subject for studying floral evolution. The greatest variation of floral morphology is seen in the lower sepal, its spur, and the lateral united petals, which seem related to different pollinators and breeding systems. Grey-Wilson (1980b) has proposed hypothetical schemes for floral evolution, e.g., loss of the upper pair of lateral sepals resulted in the evolution of a flower type with three sepals from flower types with five sepals, elaboration of the lower sepal into a nectary-tipped spur of various shapes and colors, fusion of the two lateral petals into a united pair, with reduction in size of the upper petal, and elaboration of the lower petal. Our results do not yet offer comprehensive inferences on floral evolution, due to poor resolution and weak support for many internal branches. Our ongoing studies seek more insights on these questions by acquiring more molecular data from different loci. Nevertheless, our phylogenies do bring insight on evolution of some floral traits. For example, the loss of a floral spur seen in some Madagascan endemics appears to have occurred only once (Fig. 4). Floral characters related to cleistogamy shown by *I. inaperta* have been used to diagnose section *Impatientella* (Perrier de la Bathie, 1934). Recent studies show that facultative cleistogamy is also regularly observed in the North American *I. capensis* (Antlfinger, 1986; Paoletti & Holsinger, 1999; Lu, 2002). In our molecular trees, these two species are far isolated from each other, which suggests that the same cleistogamous breeding system originated independently.

The characters of fruits have been considered of phylogenetic importance in Balsaminaceae, particularly the shape of capsules in *Impatiens* (Grey-Wilson, 1980b). When traced with our molecular phylogenies, the elongated linear or narrowly fusiform fruit was shown to be plesiomorphic, and the global or broadly fusiform fruit was shown to be derived from the linear type, while a few reversals also occurred (Fig. 4). The elongated linear type of fruit occurs mostly in Southeast Asian, Himalayan, and North-temperate species, whereas the

broadly fusiform type of fruit is more characteristic of African, Madagascan and southern Indian species.

Impatiens shows a wide range of chromosome number variation, but the most frequent numbers are $2n = 14$, 16, 18, and 20 (Song & al., 2003). Thus the major part of species show the basic chromosome numbers $x = 7$, 8, 9, and 10. This series of basic chromosome numbers suggests a typical dysploid evolutionary relationship among them. Different hypotheses have been proposed to interpret the evolution of the basic numbers. Jones & Smith (1966) and Akiyama & al. (1992b) suggested $x = 7$ to be the ancestral type, from which the other numbers were derived by ascending dysploidy, whereas Rao & al. (1986) suggested evolution of $x = 7$, 9 and 10 from the basic number $x = 8$ through both descending and ascending dysploidy. Interestingly, preliminary optimization of the incomplete chromosomal data (data are missing for most Madagascan species) onto the molecular phylogenies reveals $x = 10$ to be the most likely ancestral base number in *Impatiens*, with other base numbers being derived mainly through descending dysploidy (Fig. 4). Consistent with the transition of elongated linear to broadly fusiform fruits, the optimization suggested a transition of base chromosome number from $x = 10$ to $x = 8$ for the same clade. Considering geographic distribution, there was probably a predominantly descending dysploidy following dispersal of the clade with broadly fusiform fruits to India, Africa, and Madagascar. As an extreme case, the lowest basic chromosome number, $x = 3$, was found in two Indian species, *I. latifolia* and *I. leschenaultii*. Interestingly, these two species are confirmed to have African affinity.

Our present paper has presented the first detailed phylogenetic study on Balsaminaceae with a worldwide sampling, and it has given insight on the evolution and biogeography of the family. Despite poor resolution or weak support among main lineages of *Impatiens*, we have established evolutionary patterns of some important morphological traits, basic chromosome numbers, and geographical origins. The evolutionary history of *Impatiens* apparently involved rapid diversification and extensive range expansion particularly in the Old World tropics and subtropics.

ACKNOWLEDGEMENTS

We are very much indebted to the following persons or institutes for supplying us plant material or help in plant collections: Dr. Martin Callmänder, Dr. Philippe Chassot, Andrien Favre, Dr. Xue-Jun Ge, Dr. Jason Grant, Dr. Gang Hao, Dr. Eisuke Hayasaka, Dr. Zhen-Yu Jiang, Ray Morgan, Odile Phaehler, Michael Ronikier, Magali Schnell, Dr. Lena Struwe, Mali Wang, the National Botanic Garden of Belgium, Nationaal Herbarium

Nederland, Leiden, Berlin-Dahlem Botanic Garden, Royal Botanical Garden, Edinburgh, and Royal Botanic Gardens, Kew, and the staff in Neuchâtel Botanical Garden. We thank also Dr. Dan Potter for sharing an unpublished sequence. The study was financially supported by the Swiss National Science Foundation (FN 3100AO-102165), and the Hundreds Talents Project of Chinese Academy of Sciences granted to Y.-M. Yuan, and by research grants of the K. U. Leuven OT/01/25 and the Fund for Scientific Research - Flanders (Belgium; G.0104.01; 1.5.069.02; 1.5.061.03) granted to E. Smets. Thanks to Anja Vandepierre (K. U. Leuven) for technical assistance, Dr. Jason Grant for critical reading of the manuscript, and two anonymous reviewers for critical comments.

LITERATURE CITED

- Akaike, H.** 1974. A new look at the statistical model identification. *IEEE Trans Autom. Contr.* 19: 716–723.
- Akiyama, S. & Ohba, H.** 2000. Inflorescences of the Himalayan species of *Impatiens* (Balsaminaceae). *J. Jap. Bot.* 75: 226–240.
- Akiyama, S., Ohba, H., Sugawara, T., Yang, Y.-P. & Murata, J.** 1995. Notes on *Impatiens* (Balsaminaceae) from southwestern Yunnan, China. *J. Jap. Bot.* 70: 95–106.
- Akiyama, S., Ohba, H. & Suzuki, M.** 1992a. Notes on east Himalayan species of *Impatiens*, I. *J. Jap. Bot.* 67: 187–193.
- Akiyama, S., Ohba, H. & Wu, S.-K.** 1996. Further notes of *Impatiens* (Balsaminaceae) from Yunnan, China. *Bull. Nation. Sci. Mus., B (Tokyo)* 22: 135–144.
- Akiyama, S., Wakabayashi, M. & Ohba, H.** 1992b. Chromosome evolution in Himalayan *Impatiens* (Balsaminaceae). *Bot. J. Linn. Soc.* 109: 247–257.
- Albach, D. C., Soltis, P. S., Soltis, D. E. & Olmstead, R. G.** 2001. Phylogenetic analysis of asterids based on sequences of four genes. *Ann. Missouri Bot. Gard.* 88: 163–212.
- Anderberg, A. A., Rydin, C. & Källersjö, M.** 2002. Phylogenetic relationships in the order Ericales s. l.: analyses of molecular data from five genes from the plastid and mitochondrial genomes. *Amer. J. Bot.* 89: 677–687.
- Antlfinger, A. E.** 1986. Field germination and seedling growth of chasmogamous and cleistogamous progeny of *Impatiens capensis* (Balsaminaceae). *Amer. J. Bot.* 73: 1267–1273.
- APG.** 2003. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. *Bot. J. Linn. Soc.* 141: 399–436.
- Bremer, B., Bremer, K., Heidari, N., Erixon, P., Olmstead, R. G., Anderberg, A. A., Källersjö, M. & Barkhordarian, E.** 2002. Phylogenetics of asterids based on 3 coding and 3 non-coding chloroplast DNA markers and the utility of non-coding DNA at higher taxonomic levels. *Mol. Phylog. Evol.* 24: 274–301.
- Chen, Y.-L.** 1978. Notulae de genere *Impatiens* L. florum sinicae. *Acta Phytotaxon. Sin.* 16: 36–55.
- Chen, Y.-L.** 2000. Three new species of *Impatiens* L. from China. *Acta Phytotaxon. Sin.* 38: 557–562.
- Chen, Y.-L.** 2001. Balsaminaceae, *Flora Reipublicae Popularis Sinicae*, Tomus 47 (2). Science Press, Beijing.
- Cronquist, A.** 1981. *An Integrated System of Classification of Flowering Plants*. Columbia Univ. Press, New York.
- Denduangboripant, J. & Cronk, Q. C. B.** 2001. Evolution and alignment of the hypervariable *arm1* of *Aeschynanthus* (Gesneriaceae) ITS2 nuclear ribosomal DNA. *Mol. Phylog. Evol.* 20: 163–172.
- Doyle, J. & Doyle, J. L.** 1987. A rapid DNA isolation method for small quantities of fresh tissues. *Phytochem. Bull.* 19: 11–15.
- Felsenstein, J.** 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39: 783–791.
- Fischer, E. & Rahelivololona, M. E.** 2002. New taxa of *Impatiens* (Balsaminaceae) from Madagascar. I. *Adansonia* 24: 271–294.
- Fischer, E., Wohlhauser, S. & Rahelivololona, M. E.** 2003. New taxa of *Impatiens* (Balsaminaceae) from Madagascar. II. A collection from Masoala Peninsula. *Adansonia* 25: 17–31.
- Fujihashi, H., Akiyama, S. & Ohba, H.** 2002. Origin and relationships of the Sino-Himalayan *Impatiens* (Balsaminaceae) based on molecular phylogenetic analysis, chromosome numbers and gross morphology. *J. Jap. Bot.* 77: 284–295.
- Geuten, K., Smets, E., Schols, P., Yuan, Y.-M., Janssens, S., Küpfer, P. & Pyck, N.** 2004. Conflicting phylogenies of balsaminoid families and the polytomy in Ericales: combining data in a Bayesian framework. *Molec. Phylogen. Evol.* 31: 711–729.
- Grey-Wilson, C.** 1980a. *Impatiens* in Papuasia, Studies in Balsaminaceae: I. *Kew Bull.* 34: 661–688.
- Grey-Wilson, C.** 1980b. *Impatiens of Africa*. A. A. Balkema, Rotterdam.
- Grey-Wilson, C.** 1980c. *Hydrocera triflora*, its floral morphology and relationships with *Impatiens*. Studies in Balsaminaceae: V. *Kew Bull.* 35: 213–219.
- Grey-Wilson, C.** 1980d. Notes on collecting *Impatiens* (Balsaminaceae). *Fl. Males. Bull.* 33: 3435–3436.
- Grey-Wilson, C.** 1985. Balsaminaceae. Pp. 76–120 in: Dassanayake, M. D. & Fosberg, F. R. (eds.), *A Revised Handbook to the Flora of Ceylon*. Amerind Publ. Co., Pvt. Ltd., New Delhi.
- Grey-Wilson, C.** 1989a. A revision of Sumatran *Impatiens*. Studies in Balsaminaceae: VIII. *Kew Bull.* 44: 67–105.
- Grey-Wilson, C.** 1989b. *Semeiocardium* Zoll.; is it a good genus? Studies in Balsaminaceae: IX. *Kew Bull.* 44: 107–113.
- Hooker, J. D. & Thompson, T.** 1859. Praecursores ad floram Indicum. Balsaminaceae. *Bot. J. Linn. Soc.* 4: 106–157.
- Huang, S.-H., Shui, Y.-M. & Chen, W.-H.** 2003. New taxa of *Impatiens* from Yunnan. *Acta Bot. Yunnan.* 25: 261–280.
- Humbert, H.** 1956. Contributions a l'étude de la flore de Madagascar et des Comores (fascicule 5). *Notul. Syst. (Paris)* 15: 113–134.
- Jones, K. & Smith, J. B.** 1966. The cytogeography of *Impatiens* L. (Balsaminaceae). *Kew Bull.* 20: 63–72.
- Lu, Y.-Q.** 2002. Why is cleistogamy a selected reproductive strategy in *Impatiens capensis* (Balsaminaceae)? *Biol. J. Linn. Soc.* 75: 543–553.
- Maddison, W. P. & Maddison, D. R.** 1992. *MacClade*:

- Analysis of Phylogeny and Character Evolution, version 3.0.* Sinauer, Sunderland, Massachusetts.
- Morton, C. M., Chase, M. W., Kron, K. A. & Swensen, S. M.** 1996. A molecular evaluation of the monophyly of the order Ebenales based upon *rbcL* sequence data. *Syst. Bot.* 21: 567–586.
- Morton, C. M., Mori, S. A., Prance, G. T., Karol, K. G. & Chase, M. W.** 1997. Phylogenetic relationships of Lecythidaceae: a cladistic analysis using *rbcL* sequence and morphological data. *Amer. J. Bot.* 84: 530–540.
- Paoletti, C. & Holsinger, K. E.** 1999. Spatial patterns of polygenic variation in *Impatiens capensis*, a species with an environmentally controlled mixed mating system. *J. Evol. Biol.* 12: 689–696.
- Perrier de la Bathie, H.** 1934. Les *Impatiens* de Madagascar. *Arch. Bot. Mém.* 7: 1–124.
- Perrier de la Bathie, H.** 1948. Révision des *Impatiens* de Madagascar et des Comores. *Mem. Acad. Sci.* 67: 1–16.
- Posada, D. & Crandall, K. A.** 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* 14, 817–818.
- Rao, R. V. S., Ayyangar, K. R. & Sampathkumar, R.** 1986. On the karyological characteristics of some members of Balsaminaceae. *Cytologia* 51: 251–260.
- Saitou, N. & Nei, M.** 1987. The neighbor-joining methods: a new method for reconstructing phylogenetic trees. *Molec. Biol. Evol.* 4: 406–425.
- Shimizu, T.** 2000. New species of the Thai *Impatiens* (Balsaminaceae): 2. *Bull. Nation. Sci. Mus., B (Tokyo)* 26: 35–42.
- Soltis, D. E., Soltis, P. S., Chase, M. W., Mort, M. E., Albach, D. C., Zanis, M., Savolainen, V., Hahn, W. H., Hoot, S. B., Fay, M. F., Axtell, M., Swensen, S. M., Prince, L. M., Kress, W. J., Nixon, K. C. & Farris, J. S.** 2000. Angiosperm phylogeny inferred from 18S rDNA, *rbcL*, and *atpB* sequences. *Bot. J. Linn. Soc.* 133: 381–461.
- Song, Y., Yuan, Y.-M. & Küpfer, P.** 2003. Chromosomal evolution in Balsaminaceae, with cytological observations on 45 species from Southeast Asia. *Caryologia* 56: 463–481.
- Swofford, D. L.** 2000. *PAUP* Phylogenetic Analysis Using Parsimony (*and other methods)*, version 4.0b10. Sinauer Associates, Sunderland, Massachusetts.
- Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G.** 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 24: 4876–4882.
- Warburg, O. & Reiche, K.** 1895. Balsaminaceae. Pp. 383–392 in: Engler, H. G. A. & Prantl, K. A. E. (eds.), *Die Natürlichen Pflanzenfamilien*, Teil 3, Abteil. 5. Wilhelm Engelmann, Leipzig.
- Yuan, Y.-M., Wohlhauser, S., Möller, M., Chassot, P., Mansion, G., Grant, J., Küpfer, P. & Klackenberg, J.** 2003. Monophyly and relationships of the tribe Exaceae (Gentianaceae) inferred from nuclear ribosomal and chloroplast DNA sequences. *Molec. Phylog. Evol.* 28: 500–517.