A three-genome phylogeny of *Momordica* (Cucurbitaceae) suggests seven returns from dioecy to monoecy and recent long-distance dispersal to Asia

Hanno Schaefer *, Susanne S. Renner

**A R T I C L E  I N F O**

**Article history:**
Received 17 May 2009
Revised 9 August 2009
Accepted 10 August 2009
Available online 15 August 2009

**Keywords:**
Ancestral state reconstruction
Dioecy
Leafy
Molecular clock
Monoecy
Sexual system transitions

**A B S T R A C T**

The bitter gourd genus *Momordica* comprises 47 species in Africa and 12 in Asia and Australia. All have unisexual flowers, and of the African species, 24 are dioecious, 23 monoecious, while all Asian species are dioecious. Maximum likelihood analyses of 6257 aligned nucleotides of plastid, mitochondrial and nuclear DNA obtained for 122 accessions of *Momordica* and seven outgroups show that *Momordica* is monophyletic and consists of 11 well-supported clades. Monoecy evolved from dioecy seven times independently, always in Africa and mostly in savanna species with low population densities. Leaky dioecy, with occasional fruit-producing males, occurs in two African species and might be the first step in an evolutionary transition towards monoecy. Dated biogeographic analyses suggest that *Momordica* originated in tropical Africa and that the Asian species are the result of one long-distance dispersal event about 19 million years ago. The pantropical vegetable *Momordica charantia* is of African, not Asian origin as had previously been suggested.

© 2009 Elsevier Inc. All rights reserved.

1. Introduction

In the gourd family (Cucurbitaceae), flowers are mostly unisexual, and about 50% of the 900–1000 species are monoecious, the remainder dioecious (Schaefer and Renner, unpublished). Details of sexual system evolution in the family, however, are poorly known. Zhang et al. (2006) reconstructed bisexual flowers as ancestral in the order Cucurbitales (seven families) and dioecy as having evolved in the common ancestor of the four core cucurbitaceous families Begoniaceae, Cucurbitaceae, Datiscaceae, and Tetramelaceae. Dioecy was inferred as ancestral for the Cucurbitaceae, and monoecy in the family would then be the result of repeated sexual system shifts. The only evolutionary analysis to date focused on the small cucurbit genus *Bryonia* with about ten species, where either two shifts from dioecy to monoecy or four shifts from monoecy to dioecy occurred, depending on the assumed ancestral state (Volz and Renner, 2008).

*Momordica* comprises 59 species growing in rainforest, deciduous forest, bushland, savannas, or grassland. Most species are perennial climbers, two species are small shrubs and two species are annuals. *Momordica* flowers measure 1–10 cm in diameter and are short-lived. About a third of the species are monoecious, the remaining dioecious. Three of the monoecious species have been shown to be self-compatible (Palada and Chang, 2003; Loke-sha and Vasudeva, 2001; H.S., pers. obs.), and self-incompatibility is unknown in Cucurbitaceae. Details of flowering biology are known only for the bitter gourd, *M. charantia*, which is an important crop species in parts of Asia and Africa. In this species, the first flowers appear 45–55 days after sowing, and under long-day conditions male flowers bloom 2 weeks before the first female flowers; the female to male flower ratio is ca. 1 in 25, and flowering lasts ca. 6 months (Palada and Chang, 2003). Sex expression and inheritance have been studied in *M. dioica*, which is used as a vegetable in India (Hossain et al., 1996). Crossing experiments that used artificially induced monoecious plants indicated that sex in *M. dioica* is controlled by a single factor, with heterozygous males and homozygous recessive females.

In this study, we use field and herbarium data, molecular markers from three genomes, ancestral state reconstruction, and molecular dating to infer how often, when, and in which direction transitions between monoecy and dioecy have occurred in *Momordica* and if they are related to habitat, life form, or pollinators.

2. Materials and methods

2.1. Taxon sampling

Online Supplement Table 1 provides a complete list of the 58 species of *Momordica* included in this study (out of 59 accepted species in the genus), with authors, geographic origin, sexual system, habitat, life form, pollinator, fruit type, chromosome number, and GenBank Accession Numbers for all sequences. We included...
six as yet undescribed species and one recently described African species (Schaefer, 2009), a full taxonomic revision of the genus will be published elsewhere. To detect intra-specific variation, we sequenced 2–6 accessions for 80% of the species; for 11 rare species only one accession was available. Our analysis includes 605 sequences, of which 593 were generated for this study. In addition, the first author studied several thousand herbarium specimens in 21 European, African, American and Chinese institutions (A, B, BM, BR, CAS, DSm, E, EA, FI, FT, GH, K, KUN, L, M, MO, P, UPS, W, WAG and WU, herbarium acronyms follow Index Herbariorum at http://sciweb.nybg.org/science2/IndexHerbariorum.asp), and 19 of the 59 species in their natural habitat in Australia, Tanzania, Nigeria, Sierra Leone, and China. Ten species were cultivated at the botanical garden Munich. Two species could not be included: M. macrosperma (Cogn.) Chiov., which is known only from the leafless type at FT and was not available for DNA extraction, and M. sahyadrica Kattuk. & V. T. Antony, a recently described entity from India, which based on morphological characters seems to be a hybrid (H.S., unpublished). Following the Cucurbitaceae phylogeny of Schaefer et al. (2009), we included two distinct outgroups, Bolbostemma paniculatum and Gerrandanthus grandiflorus, and the close relatives Baijiaania bornensis, B. yunnanensis, Thladiantha cordifolia, T. nudiflora, and T. pustulata.

2.2. DNA isolation, amplification, sequencing, and alignment

Total genomic DNA was isolated from silica-dried leaves and herbarium specimens using the NucleoSpin plant kit (Macherey-Nagel, Düren, Germany). DNA amplification by polymerase chain reaction (PCR), fragment purification and cycle sequencing followed standard protocols. We amplified six markers from the three genomes (i) one coding and three non-coding plastid regions: the plastid maturation K (matK) gene, the trnL intron and adjacent trnL-F spacer, the rpl20-rps12 intergenic spacer between the ribosomal protein genes S12 and L20, and the trnH-psbA spacer. Primers used to amplify the first three regions are given in Schaefer et al. (2009). The trnH-psbA spacer was amplified with primers trnH and psbA given in Volz and Renner (2009). (ii) From the nucleus, we amplified the second intron of the Leafy gene (referred to herein simply as LFY) using the primers LFY CucF 5-TCT TCC ACC TST ATG ARC AGT GTC GTG AAT-3 and LFY CucR 5-CGA AAT CAC AAA AGY TAT TGS GYA KTY CA-3 (developed by HS). (iii) The mitochondrial region used in this study is the nad1 intron (NADH dehydrogenase subunit 1, exon B-C), which was amplified using the nad1/B-C universal primers from Demesure et al. (1995).

Five randomly chosen samples of the nuclear and mitochondrial products were cloned as described in Volz and Renner (2008) to assess within-plant sequence divergence. For all markers, sequencing primers were the same as for amplification. Both strands were sequenced and used to generate consensus sequences in Sequencher (ver. 4.1.2, GeneCodes Corp.). Sequence alignment was done manually in MacClade ver. 4.0.6 (Maddison and Maddison, 2003). The alignments have been deposited in TreeBASE (www.treebase.org).

2.3. Phylogenetic analyses

Maximum likelihood (ML) tree searches and ML bootstrapping were performed using RAxML ver. 7.0.4 (Stamatakis et al., 2008; available at http://phylobench.vital-it.ch/raxml-bb/). RAxML searches relied on the general time-reversible (GTR) model with among-site rate heterogeneity modelled by a gamma distribution with 25 rate categories (all model parameters were estimated over the duration of specified runs). Although RAxML can incorporate a further parameter for the proportion of invariant sites, we chose not to use it because the 25 gamma rate categories sufficiently account for rate heterogeneity (Stamatakis et al., 2008). Analyses of the separate plastid, nuclear, and mitochondrial data partitions produced congruent phylogenetic estimates (not shown), with all areas of discordance being restricted to branches with low support (i.e., bootstrap proportions <60%). Given this overall congruence, the three data partitions were concatenated, with specific model parameters estimated for each partition. To detect if analyses had become trapped in local optima, we conducted ten separate ML searches. The tree with the best likelihood score was then selected as the optimal tree. Non-parametric bootstrapping as implemented in the “fast bootstrap” algorithm of RAxML used 500 replicates.

Bayesian inference used the GTR + Γ model (with the default four rate categories) plus a proportion of invariant sites and relied on MrBayes (Huelsenbeck and Ronquist, 2001). We again analyzed the combined dataset with three partitions (plastid, nuclear, mitochondrial), allowing partition models to vary by unlinking gamma shapes, transition matrices, and proportions of invariant sites. Markov chain Monte Carlo (MCMC) runs started from independent random trees, were repeated twice, and extended for 1 million generations, with trees sampled every 100th generation. We used the default priors in MrBayes, namely a flat Dirichlet prior for the relative nucleotide frequencies and rate parameters, a discrete uniform prior for topologies, and an exponential distribution (mean 1.0) for the gamma-shape parameter and branch lengths. Convergence was assessed by checking that the standard deviations of split frequencies were <0.01; that the log probabilities of the data given the parameter values fluctuated within narrow limits; that the convergence diagnostic (the potential scale reduction factor given by MrBayes) approached one; and by examining the plot provided by MrBayes of the generation number versus the log probability of the data. Trees saved prior to convergence were discarded as burn-in (1000 trees) and a consensus tree was constructed from the remaining trees.

2.4. Analysis of sexual system evolution

To analyze sexual system evolution, we used a reduced dataset of 65 species with one accession per species (58 Momordica and seven outgroup species). We mapped the evolution of sexual systems onto the phylogeny by defining three unordered states, (0) unknown, (1) monoecious, and (2) dioecious. Information on sexual systems came from personal observation in the natural habitat and greenhouse, from the study of herbarium specimens, floras, and monographs (online Supplement Table 1). To infer ancestral states, we used parsimony and maximum likelihood as implemented in Mesquite ver. 2.6 (Maddison and Maddison, 2009; http://mesquiteproject.org/mesquite/mesquite.html). All analyses were carried out on the preferred highest likelihood tree, with branch lengths set to the maximum likelihood values obtained under the GTR + Γ model (above). Likelihood analyses in Mesquite used the Markov k-state one-parameter model, which is a generalization of the Jukes–Cantor model (Lewis, 2001) and assumes a single rate for all transitions between character states. We let Mesquite estimate the transition parameters of the model, based on the tip trait states in the 65-taxon tree and its branch lengths.

2.5. Molecular clock analyses

To transform the likelihood tree inferred from the combined dataset into a time-calibrated phylogeny, we used Bayesian clock estimations with an uncorrelated rates model (Drummond et al., 2006). The 65–species alignment was first imported in BEAUti ver. 1.4.8 (part of the BEAST package; Drummond and Rambaut, 2007) to generate BEAST input files. We used the GTR + Γ model with six rate categories. Based on the results of a family-wide analysis with a four-plastid-marker matrix and four fossil-calibration points
Schaefer et al. (2009), we used the following five constraints (all with a normal prior distribution): *Momordica* stem node 46 ± 3 Ma, *Momordica* crown node 31 ± 5 Ma, Thladianthinae (*Thladiantha* and *Baijiania*) crown node 45 ± 4 Ma, *Baijiania* crown node 32 ± 6 Ma, and *Thladiantha* crown node 14 ± 5 Ma. The performance of the BEAST runs was checked using Tracer ver. 1.4 (Rambaut and Drummond, 2007). Following convergence, the resulting trees were combined using TreeAnnotator ver. 1.4.8 (part of the BEAST package), with a burn-in of 2000 trees. Final trees were checked and edited in FigTree ver. 1.2.2 (Rambaut, 2006-2009).

2.6. Biogeographical analyses

To estimate the biogeographical history of *Momordica*, we used ancestral area reconstruction (Ronquist, 1994). We estimated geographic regions as an unordered multi-state character using the following six regions as character states (i) East and Southeast Asia including New Guinea and tropical Australia, (ii) West Asia, (iii) East and Northeast Africa, (iv) Central Africa, (v) West Africa, and (vi) Southern Africa. For the widespread species *M. balsamina*, *M. charantia*, and *M. foetida*, we used an additional state “all Africa.” Recent human-mediated range expansions of *M. balsamina* and *M. charantia* into the New World tropics, Asia, and the Pacific islands were ignored. *Momordica cymbalaria* (syn. *M. tuberosa*) is known from North and East Africa to India and Pakistan but likely was introduced to Asia as a vegetable and medicinal plant (Lokesha and Vasudeva, 2001); it was therefore coded as East and Northeast Africa only. We then used maximum likelihood as implemented in Mesquite to reconstruct shifts between biogeographic regions. This method incorporates branch-length information, which is why we...
employed it as an alternative to dispersal-vicariance analysis (DIVA; Ronquist, 1997).

3. Results

3.1. Morphological analyses and field observations

Of the 59 described and new Momordica species (58 of them included in Fig. 1), 34 are dioecious and 22 monoecious. For three of the new species it remains unclear whether they are monoecious or dioecious because the available collections are too incomplete. Based on field observations and herbarium specimens, all species of the closely related genera Indofevillea, Thladiantha, Baijiania, Telfairia, and Cogniauxia are dioecious. In monoecious Momordica, the number of female flowers is much lower than that of male flowers, ranging from about 1:10 to 1:25. In dioecious species studied in the field, female plants are much less common than male plants (ca. 1:50). In the two predominantly dioecious species M. friesiorum and M. pterocarpa, three herbarium specimens of 150 and 4 of ca. 500 were monoecious (i.e., they were fruit-producing males).

3.2. Phylogenetic analyses

The final plastid sequence matrix consisted of 1217 aligned nucleotides of the matK gene, 1177 nucleotides of the trnL and trnL-F region, 869 nucleotides of the rpl20-rps12 intron, and 620 nucleotides of the trnH-psbA spacer. Plastid sequences were obtained for all included species and accessions. The nuclear dataset consisted of 976 aligned nucleotides of LFY, which were obtained for 49 accessions representing 36 species. The mitochondrial matrix consisted of 1659 aligned nucleotides of nad1, which were obtained for 31 samples belonging to 31 species.

Fig. 2. The evolution of monoecy and dioecy in Momordica as inferred on the preferred maximum likelihood tree (character states: dioecious – white, monoecious – black, unclear – grey); numbers on the right indicate sexual system shifts.
Maximum likelihood analyses of the individual datasets revealed no topological contradictions with significant bootstrap support (Section 2), and the four partitions were therefore combined into a matrix with 6257 aligned nucleotides. The topology of the resulting maximum likelihood tree (Fig. 1) is identical to the Bayesian consensus tree (not shown) in all statistically supported nodes and shows that *Momordica* is monophyletic (with 100% bootstrap support as well as Bayesian posterior probability) and that the genus can be divided into 11 clades that mostly correspond to the morphological clades proposed by Jeffrey and de Wilde (2006). Multiple accessions of the same morphological species clustered together, but in *M. cissoides*, *M. charantia*, and *M. enneaphylla*, large genetic differences among accessions (long branches) suggest additional cryptic species.

3.3. Sexual system evolution

Parsimony and maximum likelihood produced identical inferences about changes between monoecy and dioecy in *Momordica*. We therefore show only the reconstruction obtained under likelihood (Fig. 2). The ancestral state in *Momordica* most likely is dioecy. Monoecy then evolved independently in seven species or groups of species (1) *M. calantha*, a perennial liana of secondary rainforests in Northern Tanzania and Kenya; (2) the *M. charantia* clade, three savanna species of southern and southwestern Africa and two weed species common along forest margins and on disturbed ground throughout tropical Africa; (3) the sister species *M. corymbifera* and *M. repens*, tuberous perennials from southern African savannas; (4) the Raphanocarpus group of six perennials.
with dry fruits, mostly from savannas or desert habitats of Eastern or sub-Saharan Africa; (5) the *M. multiflora* clade, six species from tropical forest in Central Africa; for two of them the sexual system is unknown; (6) *M. cardiospermoides*, a perennial savanna species of East and Southern Africa; (7) *M. jeffreyana*, a perennial liana of the West African rainforests.

### 3.4. Biogeography

Reconstructions of the ancestral area (Fig. 3) and habitat (not shown) yielded humid forests in East and Southeast Asia as the most likely states for the *Momordica* stem node. The area occupied by the crown group node is reconstructed as most likely East and Northeast Africa. The genus subsequently diversified in “East and Northeast Africa” and “West Africa.” Of the 11 monophyletic groups within *Momordica*, four are reconstructed as East and Northeast African with >60% likelihood, two with >50%. The vegetative species *M. charantia* and *M. balsamina* are both reconstructed as originating in Africa.

### 3.5. Molecular clock analyses

Divergence time estimates for *Momordica* are shown in Fig. 4. The oldest of the 11 clades within the genus (labelled in Fig. 1) is the *Dimorphochlamys* group (29 (32–14) Ma) with four tall perennial liana species in the East and West African Tropics and Mozambique. The Asian *M. cochinchinensis* group is estimated to be ca. 19 (24–14) My old. Age estimates for the nodes inferred to have shifted

---

**Fig. 4.** Chronogram for 58 species of *Momordica* and outgroups obtained under a model of uncorrelated rate change using five calibration points (see text for details). Node heights are the median ages, 95% highest posterior density intervals are indicated by grey bars at the nodes.
from dioecy to monoecy range from 34 to 2 Ma. The South East Asian and Australian *M. sphaeroidea* is about 3.9 (4.9–1.5) My old.

4. Discussion

The inferred seven transitions from dioecy to monoecy in *Momordica* illustrate the flexibility of Cucurbitaceae sexual systems. The genus also provides a so far unique example of a consistent trend away from dioecy and towards monoecy. In animals, with their more determined body plans, this reversal may be impossible (Bull and Charnov, 1985), but in plants such reversals have now been reported from several other genera, such as *Fuchsia* (Onagraceae; Sytsma et al., 1991), *Sagittaria* (Alismataceae; Dorken et al., 2002), *Mercurialis* (Euphorbiaceae; Obbard et al., 2006), *Bryo-

nia* (Volz and Renner, 2008), and *Iphigenia/Wurmbea* (Colchicaceae; Case et al., 2008). In *Iphigenia* (nested in the genus *Wurmbea*) and *Fuchsia*, monoecy appears to be associated with long-distance dispersal, in agreement with Baker’s law (1967) that a single plant can only start a new population if it is self-fertile and produces hermaphroditic flowers or unisexual flowers of both sexes. In *Momor-

dica*, reversals to monoecy are not associated with long-distance dispersal events, and the Asian lineage, which is the result of the longest dispersal event in the genus, is entirely dioecious. It seems more likely that the transitions to monoecy in *Momordica* are related to plant–pollinator relationships and how these relationships are moulded by population density, which in turn relates to habi-

tat. The pollinators of all but three species of *Momordica* are special-

isled oil bees of the genus *Ctenoplectra* (Schafer and Renner, 2008; online Supplement Table 1). Male flowers in *Momordica* (ex-

cept those of *M. balsamina, M. charantia,* and *M. dioica*) produce nectar, pollen and fatty oil, while female flowers produce fatty oils only (Vogel, 1990; H.S., unpubl.). Male flowers are borne singly or in few-flowered umbels in leaf axils throughout the flowering sea-

son. Female flowers are produced towards the end of the season and always singly. Moreover, the amount of oil in female flowers tends to be lower than that offered in male flowers (Vogel, 1990, p. 75), and in some species, female flowers are entirely rewardless. This is the case in the dioecious *M. cabrae* (Keraudren-Aymonin, 1975, p. 36). *Ctenoplectra* bees need a mixture of pollen and oil (and perhaps small amounts of nectar) as larval food. Adults feed mainly on nectar. Male flowers are therefore more rewarding for the bees than are female flowers. In populations of dioecious species, female plants are much less frequent than male plants (Section 3.1) and visited mainly accidentally: isolated female plants are rarely if ever visited by oil bees (H.S., pers. obs.). In *Momordica* species with low population density, dioecy may therefore corre-

late with low pollination success and thus low fruit set, favoring the establishment of monoecious mutants, which on average should have higher reproductive success. Population densities in *Momordica* tend to be higher among forest species than among sa-

vanna species (H.S., pers. obs.), and low-density populations may not be able to persist in savannas as long as they remain strictly dioecious. These correlations fit with thirteen monoecious species occurring in savannas, 11 in forests, while of the dioecious species, seven occur in savannas and bushland, 24 in forest.

In the dioecious species *M. f r e i s i o r i u m* and *M. pterocarpa*, some herbarium specimens have a single female flower or fruit on branches with male flowers. Occasional flowers of the other sex in normally unisexual individuals, termed leaky dioecy (Baker and Cox, 1984; Humeau et al., 1999), are not uncommon in dioecious Cucurbitaceae (H.S., unpubl. field and herbarium observations) and angiosperms in general (Darwin, 1877; Lewis, 1941; Ross, 1982; Ehlers and Bataillon, 2007). However, in spite of the phenom-

enon being well known and widespread, we little know about the ecological and evolutionary consequences of rare fruit-bearing males. These consequences must depend on the quantity and qual-

ity of the seeds produced by “leaky” males relative to co-occurring females. Without data on these aspects, one cannot predict whether a particular leaky-dioecious species is on its way back to monoecy or not. Nevertheless it is clear that the spread of bisexual mutants must be at the start of any transition from dioecy to monoecy.

The biogeographic history of *Momordica* reconstructed in this study fits well into the scenario developed for the history of the cucurbits family in Schaefer et al. (2009) if a tropical Asian origin of the *Momordica* stem lineage is assumed (so far no fossil evidence available). Having spread into Northern and Eastern Africa, inde-

pendent *Momordica* lineages expanded their range into Central and West Africa, and Southern Africa. About 19 Ma, one lineage then dispersed back to Asia with one species, *M. sphaeroidea*, finally re-

aching Australia less than 4 Ma. The important vegetable species *M. charantia* and the less commonly grown vegetable *M. balsamina* both belong to a South African clade and are therefore most likely of African origin and not Asian as suggested by Marr et al. (2004).

Acknowledgments

The authors thank B. Duyfjes, C. Jeffrey, M. Thulin, W. de Wilde, and J. Wieringa for material and helpful comments, and the curators of A, B, BM, BR, CAS, DSM, E, EA, FL, FT, GH, K, KUN, L, M, MO, P, UPS, W, WAG and WU for hospitality and/or loan of mate-

rial. Financial support for the project came from the German Sci-

cence Foundation (DFG) Grant No. RE603/3-1.

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version, at doi:10.1016/j.ympev.2009.08.006.

References


d. 71, 244–253.


Keraudren-Aymonin, M., 1975. Cucurbitaceae. In: Flore d’Afrique Centrale (Zaire-


Rambaut, A., 2006-2009. Fig.Tree. Tree Figure Drawing Tool, Version 1.2.2. Available from: <http://tree.bio.ed.ac.uk/software/figtree/>.