Nuclear DNA content in *Sinningia* (Gesneriaceae); intraspecific genome size variation and genome characterization in *S. speciosa*

David Zaitlin and Andrew J. Pierce

Abstract: The Gesneriaceae (Lamiales) is a family of flowering plants comprising >3000 species of mainly tropical origin, the most familiar of which is the cultivated African violet (*Saintpaulia* spp.). Species of Gesneriaceae are poorly represented in the lists of taxa sampled for genome size estimation; measurements are available for three species of *Ramonda* and one each of *Haberlea*, *Saintpaulia*, and *Streptocarpus*, all species of Old World origin. We report here nuclear genome size estimates for 10 species of *Sinningia*, a neotropical genus largely restricted to Brazil. Flow cytometry of leaf cell nuclei showed that holoploid genome size in *Sinningia* is very small (approximately two times the size of the *Arabidopsis* genome), and is small compared to the other six species of Gesneriaceae with genome size estimates. We also documented intraspecific genome size variation of 21%–26% within a group of wild *Sinningia speciosa* (Lodd.) Hiern collections. In addition, we analyzed 1210 genome survey sequences from *S. speciosa* to characterize basic features of the nuclear genome such as guanine–cytosine content, types of repetitive elements, numbers of protein-coding sequences, and sequences unique to *S. speciosa*. We included several other angiosperm species as genome size standards, one of which was the snapdragon (*Antirrhinum majus* L.; Veronicaceae, Lamiales). Multiple measurements on three accessions indicated that the genome size of *A. majus* is $\sim 633 \times 10^6$ base pairs, which is approximately 40% of the previously published estimate.

Key words: Gesneriaceae, Sinningieae, *Sinningia*, genome size, intraspecific variation, *Antirrhinum*.

Résumé : Les Gesnériacées (Lamiales) sont une famille de plantes à fleurs qui compte >3000 espèces, principalement d'origine tropicale, dont la plus connue est la violette africaine (*Saintpaulia* spp.). Les espèces appartenant aux Gesnériacées sont mal représentées parmi les listes de taxons dont la taille des génomes a été estimée ; des mesures sont disponibles pour seulement trois espèces de *Ramonda* et une espèce chacune pour les genres *Haberlea*, *Saintpaulia* et *Streptocarpus*, toutes des espèces provenant de l'Ancien Monde. Les auteurs rapportent des estimés de la taille du génome nucléaire pour 10 espèces du genre *Sinningia*, un genre néotropical largement limité au Brésil. L'analyse par cytométrie en flux de noyaux de cellules foliaires a montré que la taille du génome holoploïde chez les *Sinningia* est très petite ($\sim$2X *Arabidopsis*) et qu'elle est petite par rapport aux tailles estimées pour six autres espèces de Gesnériacées. Les auteurs ont également documenté de la variation intraspécifique quant à la taille du génome de l'ordre de 21 à 26 % au sein d'un groupe de *Sinningia speciosa* (Lodd.) Hiern sauvages. De plus, les auteurs ont analysé 1210 séquences génomiques aléatoires chez le *S. speciosa* afin de déterminer certaines caractéristiques de base du génome nucléaire comme le contenu en G+C, les types d'éléments répétés, le nombre de séquences codant pour des protéines et les séquences uniques au *S. speciosa*. Plusieurs autres angiospermes ont été employées comme standard, incluant le mufflier, *Antirrhinum majus* L. (Véronicacées ; Lamiales). De multiples mesures réalisées sur trois accessions indiquent que la taille du génome de l'*A. majus* est de $\sim$633 Mb, soit environ 40 % de la valeur rapportée antérieurement.

Mots-clés : Gesnériacées, Sinningieae, *Sinningia*, taille du génome, variation intraspécifique, *Antirrhinum*.

[Traduit par la Rédaction]

Introduction

Determination of genome size is an active area of investigation in plant research, with estimates currently available for >5100 species of land plants (Bennett and Leitch 2005). Genome sizes vary enormously in angiosperms, from <0.1 pg to >125 pg ($\sim$1900-fold; Bennett et al. 2008), although the distribution is strongly skewed towards the "very small" (≤1.4 pg) and "small" (1.4–3.5 pg) size classes, with a modal value of 0.6 pg (Soltis et al. 2003; Greilhuber et al.

Received 3 June 2010. Accepted 4 August 2010. Published on the NRC Research Press Web site at genome.nrc.ca on 27 November 2010.

Corresponding Editor: P. Gustafson.

D. Zaitlin.[1] Kentucky Tobacco Research and Development Center, 1401 University Drive, University of Kentucky, Lexington, KY 40546, USA.
A.J. Pierce. Department of Microbiology, Immunology and Molecular Genetics, Markey Cancer Center, University of Kentucky, Lexington, KY 40536, USA.

[1]Corresponding author (e-mail: dzait2@uky.edu).

2006). Phylogenetic reconstructions suggest that the common ancestor of flowering plants had a small genome and that a small genome is the ancestral state for most of the major clades within angiosperms (Leitch et al. 2005). Indeed, the extant basal angiosperms that have been sampled all have small genomes (Soltis et al. 2003). Genome size can be given either as a mass in picograms or in million base pairs (Mbp), where 1 pg = 978 Mbp (Doležel et al. 2003), and is expressed in terms of "C value," where 1C is the amount of DNA in a reduced nonreplicated nucleus (Greilhuber et al. 2005). All species examined in *Sinningia* have the same diploid chromosome number ($2n = 26$; Möller and Kiehn 2003), and polyploidy is unknown in the natural species (Perret et al. 2003; Weber 2004). Therefore, we use the term "genome size" to indicate the mass of DNA in the unreplicated holoploid genome (equivalent to 1C) and "nuclear DNA content" to refer to the amount of DNA present in the nonreduced (diplophasic) nucleus (2C), after Greilhuber et al. (2005).

Nuclear DNA content is considered to be a "key diversity character" in plants (Bennett et al. 2008) and has been correlated with the relative sizes of the cell and nucleus, cellular processes such as rate of DNA synthesis and timing of the mitotic cell cycle, growth and other agronomic traits in crops (collectively referred to as "nucleotypic effects"; Francis et al. 2008; Korban et al. 2009), as well as a range of ecological characters (Reeves et al. 1998; Weiss-Schneeweiss et al. 2006). From a purely practical standpoint, genome size data can have taxonomic utility and has been used to distinguish species groups in *Petunia* (Mishiba et al. 2000), to delineate taxa of hybrid origin in the European hawkweeds (*Hieracium*; Suda et al. 2007), and as a systematic marker to clarify evolutionary relationships among species and subgenera of *Artemisia* (Garcia et al. 2006). Estimates of genome size can also be useful to the plant breeder; for example, in Kentucky bluegrass (*Poa pratensis* L.), knowledge of 2C values is crucial when selecting parental germplasm and identifying hybrids because chromosome numbers can vary between 28 and 140 as a result of polyploidy (Eaton et al. 2004). The impact of flow cytometry and genome size on plant breeding has recently been reviewed by Ochatt (2008).

Sinningia (Gesneriaceae) is an ecologically diverse genus of ~70 species of tropical herbs or shrubs, most of which produce annual shoots from a perennial tuber. The majority of species are native to southern Brazil, although the genus ranges from Central America to Argentina (Weber 2004). *Sinningia incarnata* (Aubl.) Denham has the broadest contiguous range, extending into southern Mexico, where occasional use of the tuber as a treatment for dysentery has been documented by the indigenous Mixe people of Oaxaca (Heinrich et al. 1992). Species vary considerably in vegetative habit, floral morphology, and mature plant size. *Sinningia incarnata* and *Sinningia sceptrum* (Mart.) Wiehler, for example, can be well over 1 m in height, whereas the miniature species *Sinningia pusilla* (Mart.) Baill., *Sinningia concinna* (Hook. f.) Nichols. and *Sinningia muscicola* Chautems, Lopes & Peixoto are smaller than the model dicot *Arabidopsis thaliana* (L.) Heynh. and are among the smallest of terrestrial angiosperms. Ecologically, most *Sinningia* species are saxicolous (living on or among rocks) and commonly occur on granite-gneiss inselbergs ("island mountains") and other rocky habitats in the Atlantic coastal mountains in Brazil (Meirelles et al. 1999), although some are terrestrial or epiphytic. Pollen transfer can be mediated by birds, bees, bats, or moths, and variable traits associated with these pollination syndromes include flower color (white, lavender, purple, yellow, green, orange, and red), flower shape, floral scent, and nectar sugar composition. Nearly one-half of all species in *Sinningia* possess tubular flowers that are red or red-orange in color, traits that are commonly associated with hummingbird pollination (Perret et al. 2003). In one of these species, *Sinningia cardinalis* (Lehm.) H.E. Moore, corolla color was shown to be due to the presence of 3-deoxyanthocyanins, a class of variant flavonoid pigments that are rare in angiosperms (Winefield et al. 2005).

Two recent taxonomic treatments of New World Gesneriaceae have placed *Sinningia* in subfamily Gesnerioideae, tribe Gloxinieae in the company of *Gloxinia*, *Smithiantha*, *Eucodonia*, *Kohleria*, and others (Wiehler 1983; Burtt and Wiehler 1995). However, the presence of a tuber in most species, rather than a subterranean rhizome, separates *Sinningia* from the rest of the tribe. The cladistic analysis of Boggan (1991) based on 46 morphological characters strongly suggested that species of *Paliavana* and *Vanhouttea* are nested within the broadly defined *Sinningia*. More recent studies using nuclear and plastid DNA sequences supported this arrangement and showed that *Sinningia*, including *Paliavana* and *Vanhouttea*, comprises a separate evolutionary lineage within Gesnerioideae. Thus, it has been proposed that tribe Sinningieae of Fritsch (1893–1894) (cited in Weber 2004) be resurrected and redefined to include these three genera (Smith et al. 1997; Zimmer et al. 2002). Phylogenetic analysis of combined plastid and nuclear DNA sequence data confirmed the monophyly of Sinningieae and showed that *Paliavana* and *Vanhouttea* are embedded in *Sinningia*, which sets the stage for a future recircumscription of *Sinningia* (Perret et al. 2003).

Most species in *Sinningia* have conspicuous, colorful flowers and (or) attractive foliage, making them desirable as ornamentals. A few species are sold commercially, but only the cultivated forms of *Sinningia speciosa* (Lodd.) Hiern, commonly known as the florist's gloxinia, are widely available in the horticultural trade. *Sinningia speciosa* was introduced into cultivation in Great Britain in 1815 (Citerne and Cronk 1999). Its unfortunate inclusion in the rhizomatous genus *Gloxinia* (as *Gloxinia speciosa* Lodd.; Loddiges 1817) has had a long-lasting effect on horticulture, and the common name persists to this day. *Sinningia speciosa* is native to the Atlantic coastal forest of southeastern Brazil, where it occurs over a range of $\sim 4 \times 10^5$ km^2. Populations are known from the states of Minas Gerais, Rio de Janeiro, São Paulo, and Espirito Santo, with a possible extension into Santa Catarina (Skog & Boggan 2007; A. Chautems, Conservatoire et Jardin botaniques, Geneva, Switzerland, personal communication, 2010). The species is morphologically diverse in the wild, although all natural forms of *S. speciosa* have nodding, bilaterally symmetrical (zygomorphic) flowers with corollas that are lavender or purple, rarely white or pink (Fig. 1). In contrast, commercial gloxinias have larger, radially symmetrical (actinomorphic)

Fig. 1. Morphological diversity in *Sinningia speciosa*. Wild collections shown are 'Espirito Santo' (upper left), 'Carangola' (upper right), 'Cardoso Moreira-pink' (lower left), and 'Cardoso Moreira' (lower right). 'Cardoso Moreira' and 'Cardoso Moreira-pink' are from the Brazilian state of Rio de Janeiro, 'Carangola' is from Minas Gerais, and 'Espirito Santo' was collected in the southern part of Espirito Santo state.

flowers in shades of purple, red, or white that are held fully erect. Wild-type flowers have five unequal petal lobes, five calyx lobes, and four fertile stamens, whereas the erect-flowered cultivars have five to nine identical petal lobes, a similar number of calyx lobes, and five or more fertile stamens. In fact, some of the cultivated forms are so different in appearance from their wild progenitors that it can be difficult to identify them as *S. speciosa*.

Very little is known about genome size and its possible evolutionary significance in the Gesneriaceae. At the present time, estimates are available for only six taxa; four European alpine resurrection plants (*Ramonda serbica* Panc., *Ramonda myconi* Pau, *Ramonda nathaliae* Panc. & Petrovic, and *Haberlea rhodopensis* Friv.) and two species from southern Africa (*Streptocarpus cyaneus* S. Moore and *Saintpaulia ionantha* Wendl.) (Müller et al. 1997; Bennett and Leitch 2005; Zonneveld et al. 2005; Loureiro et al. 2007; Siljak-Yakovlev et al. 2008). Here, we report on nuclear DNA content for 10 species of *Sinningia*, with a focus on wild forms of *S. speciosa*. This work is novel because these are the first estimates of genome size for any of the ~1500 species of New World Gesneriaceae. We show that all species of *Sinningia* examined have small genomes and that reproducible intraspecific genome size variation is present in *S. speciosa*. We also report a corrected 1C estimate for the snapdragon (*Antirrhinum majus* L.).

Materials and methods

Plant material

All plants in the genus *Sinningia* used in this study were grown from seed, with the exception of *Sinningia tubiflora*

(Hook.) Fritsch in Engl. & Prantl, which was obtained as stem cuttings from C. Babcock of the Desert Botanical Garden (Tempe, Ariz.) in 1995. Seeds of *Sinningia* species were provided by Mr. M. Peixoto (Mogi das Cruzes, Brazil) and are available through his web site (www.brazilplants.com) or from the first author. Species and sources are summarized in Table 1. Fresh leaflet tissue of *Medicago truncatula* Gaertn. F83005 was provided by H. Zhu. Leaves of *Arabidopsis thaliana* ecotype Col-0 were provided by J. Smalle. Seed of tomato (*Solanum lycopersicum* L. 'Micro Tom') was obtained from Totally Tomatoes (Randolph, Wis.). Maize (*Zea mays* L.) 2C nuclei were used as an internal reference standard for flow cytometry — seeds of an unidentified yellow dent field *Z. mays* hybrid were from Hartz Mountain Industries (Secaucus, N.J.). Wild accessions of *A. majus* were obtained from the US Department of Agriculture, Agricultural Research Service, Ornamental Plant Germplasm Center (Columbus, Ohio). *Sinningia* seeds were germinated in 10 cm pots containing Pro-Mix BX (Premier Horticulture Inc., Quakertown, Pa.), and the seedlings were grown under fluorescent lighting (80 $\mu mol \cdot m^{-2} \cdot s^{-1}$) at 25 °C until they were large enough to sample for flow cytometry.

Isolation of nuclei and flow cytometry

Intact nuclei were isolated from young expanded leaves using reagents from the Partec CyStain PI Absolute P kit (Partec GmbH, Münster, Germany). Small samples of freshly harvested leaf tissue (10–30 mg) were minced for 30–60 s with a razor blade (Personna Double Edge Platinum Chrome, American Safety Razor Co., Verona, Va.) in 0.3 mL of cold (4 °C) nuclei extraction buffer (NEB) in a plastic Petri dish essentially as described by Galbraith et al. (1983). The fine slurry was poured through a 50 μm filter (Partec Celltrix) into a 1.5 mL microcentrifuge tube. The dish was rinsed with an additional 0.3 mL of NEB, which was poured through the same filter, and the combined filtrates were centrifuged (500 × *g* for 10 min) to pellet the nuclei. The cleared supernatant was removed carefully by aspiration, and the pellet was gently resuspended in 0.25 mL NEB. Crude suspensions of nuclei were transferred to polycarbonate tubes (10 mm × 55 mm) and stained by the addition of four volumes (1 mL) of staining solution containing propidium iodide (PI) and ribonuclease prepared as directed by the manufacturer. Stained samples were placed in the dark for 1–2 h before being analyzed on a Partec PAS flow cytometer equipped with a 20 mW, 488 nm argon laser. Cytometric data acquisition (10 000 events/sample) was triggered by red fluorescence using a long-pass 630 nm filter and acquired on a three-decade logarithmic scale, which allowed the full range of peaks from *Arabidopsis* 2C to *Z. mays* 4C to be displayed together. Noise from debris was further reduced by deriving one-dimensional fluorescent DNA-content histograms from two-dimensional plots of fluorescence versus either forward or side scatter (Fig. 2A). The plant accessions assayed varied for each run based on the availability of living material at the time. At least three individuals per species or accession were used for the genome size calculations, although limited plant material was available for *S. eumorpha* H.E. Moore ($n = 1$), *S. speciosa* 'Antonio Dias' ($n = 2$), and *S. speciosa* 'Chiltern Seeds' ($n = 2$). Five plants each of *S. pusilla*, *S. muscicola* (Chautems et al. 2010), and *S. speciosa* 'Espirito Santo' were used in the analyses.

Data analysis

Flow cytometric data for the *Sinningia* species and *A. majus* were collected in 13 independent runs performed on 13 different days over the course of 18 months. Replicated genome size standards (see below) were included in every run. Data files from the Partec PAS flow cytometer were analyzed using FlowJo version 7.5 (Tree Star, Inc., Ashland, Ore.). High-quality histograms were obtained by manually defining a gated area that included the fluorescence signals from PI-stained G_0/G_1 nuclei and that excluded interference due to cellular debris (Arumuganathan and Earle 1991). *Zea mays* nuclei, which were prepared separately, were added to all samples as an internal reference standard (2C nuclear DNA content = 5.73 pg; Johnston et al. 1999). Peak fluorescence values (geometric mean) were then divided by the geometric mean of the *Z. mays* 2C peak to give a normalized ratio.

Nuclear DNA contents for the 10 species of *Sinningia* and *A. majus* were calculated against genomes of known size: *Arabidopsis thaliana* Col-0 2C, 4C, and 8C peaks (higher ploidy nuclei are due to endopolyploidy in mature leaves; Galbraith et al. 1991; Lee et al. 2009), *M. truncatula* 2C, and *Solanum lycopersicum* 2C peaks were appropriate size standards for the species examined here. Numerical data was analyzed using Prism 5 software (GraphPad Software, Inc. La Jolla, Calif.). For each run, the fluorescence ratios (see above) derived from the standard species were plotted against their known genome sizes. Genome sizes for the unknowns were interpolated from the resulting linear regressions, and a mean genome size was then calculated for each species or accession. Statistical significance of the differences in mean genome sizes in *S. speciosa* was determined in pairwise comparisons using the Student's *t* test (QuickCalcs; GraphPad Software Inc., La Jolla, Calif.).

Genomic library construction

Total DNA (20 μg) isolated from young leaves of *S. speciosa* AC1503 was hydrodynamically sheared in a volume of 0.2 mL using a HydroShear instrument (Genomic Solutions Inc., Ann Arbor, Mich.). The DNA-containing solution was passed 10 times through a standard shearing block on a speed setting of 6, which produced fragments with a median size of ~2500 bp. The fragmented DNA was purified and concentrated on an affinity column (QIAquick PCR purification kit; Qiagen Sciences, Valencia, CA.), and a 2.5 μg sample was subjected to end repair using the DNA Terminator kit (Lucigen, Middleton, Wis.) in a volume of 25 μL. A single dA residue was added to the 3′ ends of the DNA fragments with 2.5 U of *Taq* DNA polymerase (New England Biolabs) in 1× *Taq* buffer (Lucigen) containing 2 mmol/L dATP in a volume of 50 μL at 70 °C for 60 min. End-repaired and modified DNA fragments between 1.5 and 4 kb were then purified from a 1% agarose gel in 0.5× TBE using the QIAquick gel extraction kit (Qiagen). The recovered genomic DNA fragments were ligated into the pGEM-T vector (Promega) in a 1:1 molar ratio in 10 μL at 15 °C for 16 h. A 1 μL sample of this ligation reaction was used to transform chemically competent *E. coli* strain NEB 10β

Table 1. Species and accessions of *Sinningia* and *Antirrhinum* used in this study.

Species	Accession No. or cultivar name	Collection location or origin	Source
Sinningia pusilla (Mart.) Baill.	'Itaoca'	Brazil, Rio de Janeiro	M. Peixoto
S. aggregata (Ker Gawl.) Wiehler	AC1461[a] and unnamed	Southeastern Brazil	M. Peixoto
S. guttata Lindl.	MP0977[b]	Brazil, Rio de Janeiro	M. Peixoto
S. macrophylla (Nees & Mart.) Benth. & Hook.f. ex Fritsch in Engl. & Prantl	MP1003	Brazil, Bahia	M. Peixoto
S. tubiflora (Hook.) Fritsch in Engl. & Prantl		Bolivia	C. Babcock
S. eumorpha H.E. Moore	'Saltão'	Brazil, São Paulo	M. Peixoto
S. muscicola Chautems, Lopes, & Peixoto	MP1094	Brazil, Rio de Janeiro	M. Peixoto
S. richii Clayberg	'Robson Lopes'	Brazil, Espirito Santo	M. Peixoto
S. harleyi Wiehler & Chautems	MP0483	Brazil, Bahia	M. Peixoto
S. speciosa (Lodd.) Hiern	AC1503	Brazil, Rio de Janeiro	GS Seed Fund[c]
S. speciosa	'Antonio Dias'	Brazil, Minas Gerais	M. Peixoto
S. speciosa	'Buzios'; MP0728	Brazil, Rio de Janeiro	M. Peixoto
S. speciosa	'Carangola'MP0634	Brazil, Minas Gerais	M. Peixoto
S. speciosa	'Cardoso Moreira'	Brazil, Rio de Janeiro	M. Peixoto
S. speciosa	'Cardoso Moreira-pink'	Brazil, Rio de Janeiro	M. Peixoto
S. speciosa	'Espirito Santo'	Brazil, Espirito Santo	M. Peixoto
S. speciosa	'Jurapê'	Brazil, Santa Catarina (unconfirmed)	M. Peixoto
S. speciosa	'Serra da Vista'	Brazil, Rio de Janeiro	M. Peixoto
S. speciosa	'Chiltern Seeds'	Brazil, unknown	GS Seed Fund
S. speciosa	'Dona Lourdes'	Horticulture	M. Peixoto
S. speciosa	'Guatapara'	Horticulture	M. Peixoto
S. speciosa	'White Slipper'	Horticulture	M. Peixoto
Antirrhinum majus L.	PI#420370	Spain	OPGC/USDA-ARS[d]
A. majus	PI#542417	Germany	OPGC/USDA-ARS
A. majus	PI#603105	Portugal	OPGC/USDA-ARS

[a]AC, Alain Chautems.
[b]MP, Mauro Peixoto.
[c]GS, The Gesneriad Society (www.gesneriadsociety.org).
[d]Ornamental Plant Germplasm Center, Agricultural Research Service, US Department of Agriculture.

(New England Biolabs) as directed by the manufacturer. Under these conditions, we estimated that >230 000 plasmids could be recovered from the 10 μL ligation reaction, of which ~90% were recombinant molecules.

Nucleotide sequencing of 1316 clones from the small-insert sheared genomic DNA library was performed at the University of Kentucky Advanced Genetics Technology Center on an Applied Biosystems 3730xl 96 capillary DNA sequencing instrument. All reads were trimmed to remove plasmid vector sequences and low-quality base calls, which resulted in high-quality reads averaging 920 bases in length. The resulting 1225 high-quality sequences were queried in batches against the European Bioinformatics Institute non-redundant protein database (uniprot) with the program BlastStation (TM Software, Inc., Arcadia, Calif.) using the WU-blast option and the BLOSUM62 protein substitution matrix. The output of the BLASTX searches (Altschul et al. 1990) was then manually categorized by best hit (lowest *E* value) based on homology to known or predicted protein sequences.

Results

Genome size estimation

We used flow cytometry to measure the fluorescence of isolated PI-stained *Sinningia* leaf cell nuclei, and calculated the nuclear DNA content against plant species with established genome sizes. A two-dimensional plot of forward scatter versus red fluorescence for leaf cell nuclei populations prepared from *A. thaliana* and *Z. mays* is shown in Fig. 2A. Plotting the peak fluorescence ratios (unknown peak geometric mean / *Z. mays* 2C peak geometric mean) against genome size for the 2C, 4C, and 8C peaks of *Arabidopsis thaliana* Col-0 (1C = 0.16 pg = 157 Mbp; Bennett et al. 2003) and the 2C peaks from *M. truncatula* (500 Mbp; Young et al. 2005) and *Solanum lycopersicum* (950 Mbp; Mueller et al. 2009) gave a straight line (Fig. 2B). In three independent runs, the coefficients of determination (r^2) were 0.9965, 0.9975, and 0.9984 (mean 0.9975), establishing the linearity of this sizing procedure. For another 10 runs where only *Arabidopsis* was used as a standard, mean r^2 was 0.9993 (range 0.9973–0.9999). Our use of known size standards is similar to the method of Hijri and Sanders (2004), who estimated genome size in a mycorrhizal fungus against haploid, diploid, and tetraploid nuclei from yeast.

Flow cytometry of PI-stained leaf cell nuclei showed that all species in *Sinningia* included in this study have small genomes. Excluding *S. speciosa,* genome sizes ranged from 251 ± 23 Mbp (mean ± SD) in *Sinningia harleyi* Wiehler & Chautems to 375 ± 6 Mbp in *S. macrophylla* (Table 2). The

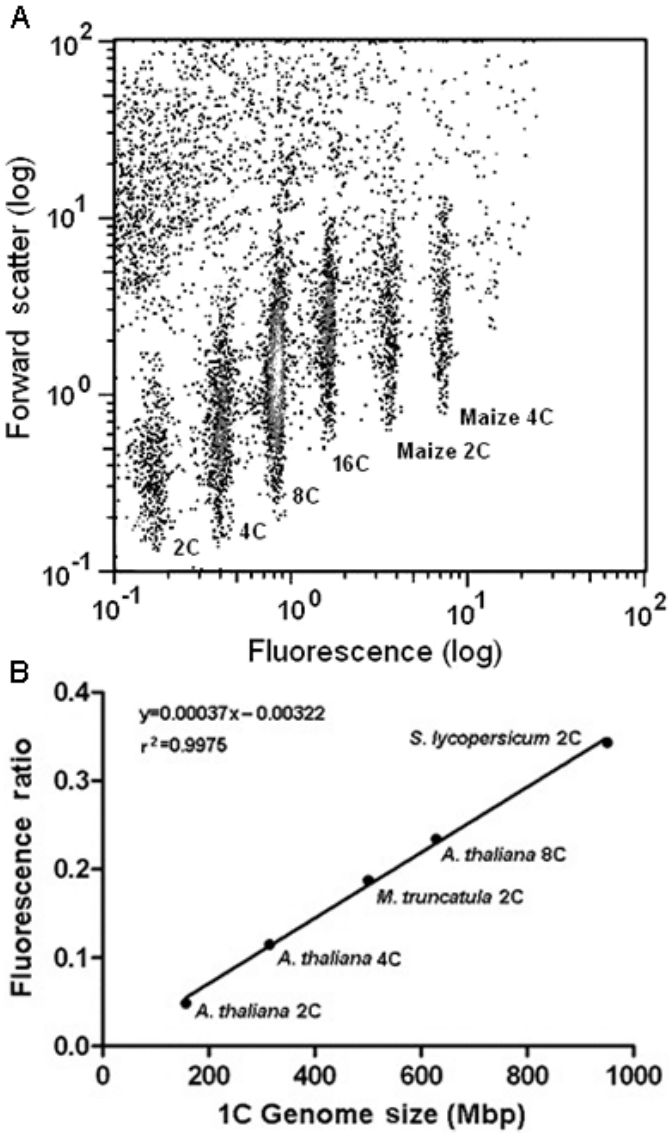


Fig. 2. (A) Flow cytometry of *Arabidopsis thaliana* Col-0 and maize (*Zea mays*) leaf cell nuclei, showing forward scatter plotted against propidium iodide fluorescence. Nuclei populations correspond exactly to the six peaks in the histogram shown in Fig. 4A. (B) Linear relationship between plant genomes of known size for *A. thaliana*, barrel medic (*Medicago truncatula*) and tomato (*Solanum lycopersicum*). Peak fluorescence ratios were determined from logarithmic data using *Z. mays* 2C nuclei as internal standard. The 4C and 8C peaks from *A. thaliana* are due to endopolyploidy (Lee et al. 2009).

overall mean for this group of nine species was 314 ± 24 Mbp. In the group of seven documented *S. speciosa* collections for which we had multiple measurements, genome sizes varied from 243 ± 37 Mbp in 'Cardoso Moreira-pink' to 309 ± 26 Mbp in 'Espirito Santo'. We also obtained estimates of 174 ± 8 Mbp for 'Antonio Dias' (based on three measurements), and 337 ± 14 Mbp for the commercial wild-type accession 'Chiltern Seeds'. Excluding 'Antonio Dias' but including 'Chiltern Seeds', the mean genome size for this group was 282 ± 29 Mbp. The two older cultivars 'Dona Lourdes' and 'Guatapara' gave genome size estimates of 573 ± 49 and 565 ± 39 Mbp, respectively, which is approximately twice the mean for the seven wild collections of *S. speciosa*. For reference, we also examined three wild accessions of *A. majus*. The mean genome size for this species based on 17 measurements for the three accessions was 633 ± 33 Mbp.

DNA sequencing of *S. speciosa* genomic clones

A small-insert clone library was prepared using randomly sheared, size-selected genomic DNA fragments from *S. speciosa* AC1503. AC1503 is a second wild collection that is morphologically identical to the 'Búzios' accession used here (A. Chautems, personal communication 2010); both were collected from the Cabo Frio area of eastern Rio de Janeiro state. A total of 1316 genomic clones were subjected to nucleotide sequencing, which resulted in 1210 unique genome survey sequences (GSS). Clones originating from organelle genomes identified by BLASTN and BLASTX searches were removed from the data set. The remaining 1 362 070 bases of nuclear DNA had a mean G + C content of 37.11%.

The output files from the BLASTX searches were examined, and each database "hit" was assigned to one of four general categories; predicted proteins, retrotransposon (RT) elements, DNA transposons, or organellar sequences. These categories were further subdivided by species, type of RT, and either chloroplast or mitochondrion (Table 3). Of 1210 GSS, 610 (50.4%) showed homology to predicted protein sequences. Of these, 500 had E values of $< 1 \times 10^{-15}$, whereas another 110 had E values between 1×10^{-5} and 1×10^{-15}. One hundred six clones (8.8%) originated from either the chloroplastic (91) or mitochondrial (15) genomes, and 97 (8%) were homologous to Ty3/*gypsy*-like (40) or Ty1/*copia*-like (57) LTR-type RTs. Another 61 clones had significant matches to RT polyproteins giving a total of 158 (13%) GSS with protein homology to known plant retroelements. The BLASTX searches also identified clones with homology to other plant repetitive elements such as a single LINE from sweet potato (*Ipomoea batatas* (L.) Lam.), and 13 class II (DNA) transposons from the hAT, Mutator, and CACTA superfamiles (Wicker et al. 2007). Most importantly, we found that 327 (27%) of the GSS had homology to known or predicted plant proteins of nuclear origin. For 210 (64.2%) of these, the highest identity matches were to sequences from one of three angiosperm species: black cottonwood (*Populus trichocarpa* Torr. & Gray), castor bean (*Ricinus communis* L.), or grape (*Vitis vinifera* L.). One hundred thirty-nine (42.5%) of the matches, most of which were from *Ricinus*, had an associated protein identity or functional annotation. Only 6 (0.98%) of the 610 clones had homology matches to predicted protein sequences from species other than plants.

We also searched the 1210 genomic clones for simple sequence repeats (SSRs) using a web-based tool (http://www.genome.clemson.edu/resources/online_tools/ssr). Searches were limited to a minimum length of 16 bases for 2–6 base motifs and 20 bases for mononucleotide repeats. We identified 174 SSRs in 141 clones, 24 of which (17%) contained between 2 and 4 SSRs. Sixty-nine percent (120) of the SSRs that met the above search criteria were comprised solely of dA and dT; there were 14 A/T mononucleotide repeats, 73

Table 2. Calculated genome size estimates for *Sinningia* species and *Antirrhinum majus*.

Species	Accession No. or cultivar name	No. of measurements	2C Nuclear DNA content (pg)	Calculated 1C genome size (Mbp)*
Sinningia pusilla	'Itaoca'	10	0.732±0.051	358±25
S. aggregata	AC1461 and unnamed	7	0.577±0.087	282±43
S. guttata	MP0977	12	0.685±0.043	335±21
S. macrophylla	MP1003	6	0.767±0.012	375±6
S. tubiflora		9	0.761±0.102	372±50
S. eumorpha	'Saltão'	8	0.521±0.020	255±10
S. muscicola	MP1094	18	0.577±0.045	282±22
S. richii	'Robson Lopes'	11	0.644±0.033	315±16
S. harleyi	MP0483	11	0.513±0.047	251±23
S. speciosa	'Antônio Dias'	3	0.356±0.016	174±8
S. speciosa	'Búzios'	11	0.628±0.072	307±35
S. speciosa	'Carangola'	8	0.632±0.106	309±52
S. speciosa	'Cardoso Moreira'	7	0.517±0.049	253±24
S. speciosa	'Cardoso Moreira-pink'	9	0.497±0.076	243±37
S. speciosa	'Espirito Santo'	22	0.632±0.053	309±26
S. speciosa	'Jurapê'	12	0.507±0.033	248±16
S. speciosa	'Serra da Vista'	16	0.515±0.055	252±27
S. speciosa	'Chiltern Seeds'	9	0.689±0.029	337±14
S. speciosa	'Dona Lourdes'	9	1.172±0.100	573±49
S. speciosa	'Guatapara'	6	1.155±0.080	565±39
S. speciosa	'White Slipper'	6	0.536±0.106	262±52
Antirrhinum majus	PI#420370	6	1.237±0.055	605±27
A. majus	PI#542417	6	1.342±0.067	656±33
A. majus	PI#603105	5	1.276±0.045	624±22
A. majus combined		17	1.294±0.067	633±33

*Values are means ± SDs.

Table 3. Classification of sequenced genomic clones from *S. speciosa* AC1503 based on manual assignment of blastx search results.

	E-value categories		
BLASTX predictions	$E < 1 \times 10^{-15}$	$1 \times 10^{-15} < E < 1 \times 10^{-5}$	Total no. (percent)
Dicots			
Arabidopsis, *Brassica*	13	10	23 (3.8)
Legumes (*Glycine*, *Medicago*, *Vicia*)	10	3	13 (2.1)
Populus	28	16	44 (7.2)
Vitis	61	17	78 (12.8)
Ricinus	65	23	88 (14.40)
Capsicum, *Nicotiana*, *Petunia*, *Solanum*	20	7	27 (4.4)
Monocots (*Oryza*, *Sorghum*, *Musa*, *Zea*, *Zingiber*, orchid)	23	11	34 (5.7)
Other (plant)	14	6	20 (3.3)
Other (nonplant)	4	2	6 (0.98)
Retrotransposons			
gypsy-like	37	3	40 (6.6)
copia-like	54	3	57 (9.3)
Provirus	1	0	1 (0.16)
RT proteins (*gag-pol*, integrase)	55	4	59 (9.7)
LINES, SINES	1	0	1 (0.16)
DNA transposons	10	3	13 (2.1)
Chloroplast genome	89	2	91 (14.9)
Mitochondrial genome	15	0	15 (2.5)
Total	500	110	610

AT/TA dinucleotide repeats, 20 trinucleotide repeats (AAT/ATT, TAA/TTA, and TAT/ATA), six tetranucleotide repeats (AAAT/TTTA, ATAA/TTAT, ATTA/TAAT, TAAA/TTTA, TATT/AATA and TTAA/TTAA), four pentanucleotide repeats (TATTT/AAATA, AATTA/TAATT and TTATT/AATAA), and three hexanucleotide repeats (AAAATT/AATTTT, ATTTTT/AAAAAT and TTTTTA/TAAAAA). Of the other 28 unique classes of SSRs, the two dinucleotide repeats AC/GT and AG/CT were the most frequent, accounting for 23 (42.6%) of the remaining 54 SSRs (not shown). Fifty-two of the SSRs identified in this search fell within predicted open reading frames (ORFs); of these, only 17 were AT/TA repeats. We also found 14 C/G mononucleotide repeats between 10 and 19 bases in length, but our search did not identify any SSR motifs of 2–5 bases that consisted of only dC and dG.

Discussion

The major finding we report here is the very small genome size of *Sinningia* s.l. Given the other published genome sizes in Gesneriaceae (662–1372 Mbp; see below), this was unexpected; however, in hindsight, the low yields of DNA isolated from leaf tissue provided anecdotal evidence for a small genome. We found two major genome size clusters in *S. speciosa* of ~250 Mbp and ~300 Mbp, estimates that were calculated against plant species with established nuclear genome sizes. We used 157 Mbp for the 1C genome size of *Arabidopsis thaliana* (Bennett et al. 2003) because it was determined by flow cytometry against the sequenced genomes of a fruit fly (*Drosophila melanogaster* Meigen) and a nematode worm (*Caenorhabditis elegans* Maupas). However, based on a comparison of the assembled genome with a subset of randomly generated shotgun sequence data, Liu and Bennetzen (2008) make a compelling argument that the *Arabidopsis* genome might be no larger than ~134 Mbp. If this estimate proves to be correct, then the two genome size clusters in wild *S. speciosa* would be reduced by 15% to approximately 213 and 256 Mbp, respectively.

Flow cytometry with *Z. mays* 2C nuclei as an internal standard proved to be a reliable technique for estimating genome size in *Sinningia*. Fluorescence ratios derived from logarithmic data gave a linear relationship for the five standard peaks (Fig. 2B). We found that the presence of plant cell nuclei from both the standards and unknowns had only a small effect on the PI fluorescence of *Z. mays* nuclei; in two runs, the *Z. mays* fluorescence readings varied between –5.8% and 3.9% when mixed with other plant cell nuclei as compared with *Z. mays* nuclei alone (data not shown). Such minimal variation is likely due to the fact that we analyzed isolated nuclei in the presence of saturating concentrations of PI. Genome size coefficients of variation (CVs) were 6.3% for *Sinningia guttata* Lindl. and 7.8% for *S. muscicola*, two species for which we had 12 and 18 measurements, respectively. The CV was also low for *S. eumorpha* (4%), which might be expected because only one individual was available for our experiments, although it was sampled three times over 12 months. The CVs were higher for both *S. pusilla* (13.6%) and *Sinningia aggregata* (Ker Gawl.) Wiehler (15%); in *S. pusilla*, a single low estimate of 257.75 Mbp from one run (2 July 2008) accounted for the high SD, and removing this value changed the overall genome size estimate from 338 ± 46 Mbp to 358 ± 25 Mbp and reduced the CV to 7%. Although there is no way to account for the one aberrant value, inadvertent sampling of an older leaf could possibly result in lower PI fluorescence because of physiological changes (such as chromatin condensation and DNA degradation) that occur during senescence (Simeonova et al. 2000; Lim et al. 2007). Because leaves of *S. pusilla* are very small, it can be difficult to determine an individual leaf's position on the rosette, and thus, an older leaf might inadvertently have been used. For *S. aggregata* (CV = 15%), we sampled three plants grown from two different seed collections (7 measurements), so it is possible that some level of actual genome size polymorphism may exist within this wide-ranging, variable species.

Intraspecific genome size variation in *S. speciosa*

Flow cytometry of leaf cell nuclei from nine wild *S. speciosa* collections indicated that considerable variation in genome size could exist within this species. Pairwise analyses (unpaired, two-tailed Student's *t* test) were performed on the 1C estimates to determine whether the observed differences in the mean values were statistically significant. The results (Table 4) confirm the presence of three genome size classes within this group of nine accessions. The genome size estimate for 'Antonio Dias' (~174 Mbp) is significantly different from all of the others (*P* values ranged from <0.0001 to 0.0264) and is discussed in detail below. 'Espirito Santo', 'Búzios', and 'Carangola' gave estimates of around 300 Mbp, whereas 'Cardoso Moreira', 'Cardoso Moreira-pink', and 'Serra da Vista' were closer to 250 Mbp. The estimate of 337 ± 14 Mbp for 'Chiltern Seeds' was the highest that we obtained for any of the wild-type accessions of *S. speciosa*, although the *t* test indicated that it is not significantly different from those in the 300 Mbp group. The accession called 'Jurapê', which is believed to have been collected in the Brazilian state of Santa Catarina, had a calculated genome size of 235 ± 34 Mbp. However, examination of the data revealed a single, anomalous estimate of 158.72 Mbp from the experimental run of 29 August 2008. Like the situation posited for *S. pusilla* above, this low value could have been due to an older leaf that was beginning to senesce. Given that the other six calculated 1C genome size estimates for 'Jurapê' were well over 200 Mbp, the larger estimates are probably closer to the true genome size for this accession. Removing the low value from the data set increased the estimate to 248 ± 16 Mbp, reduced the CV from 14.5% to 6.5%, and placed it firmly in the 250 Mbp size class. Genome sizes for both 'Carangola' and the cultivar 'White Slipper' are uncertain because of their high CVs (16.8% and 19.8%, respectively), making further investigation of these two accessions necessary. *Sinningia macrophylla*, a species closely related to *S. speciosa* from southern Bahia state (Perret et al. 2003; Skog and Boggan 2007), had a considerably larger 1C genome size of 375 ± 6 Mbp.

Sinningia speciosa 'Antonio Dias' appears to have the smallest genome size of all eight wild collections included in this study. With limited plant material ($n = 2$), we were only able to obtain three independent flow cytometric meas-

Table 4. Unpaired, two-tailed Student's *t* test comparisons of mean genome sizes in wild collections of *Sinningia speciosa*.

Comparison	*t*	df	*P*
'Buzios' vs. Cardoso Moreira'	2.6770	9	0.0253
'Buzios' vs. 'Cardoso Moreira-pink'	2.8109	10	0.0184
'Buzios' vs. 'Jurapê'	3.5104	10	0.0056
'Buzios' vs. 'Serra da Vista'	3.2272	13	0.0066
'Carangola' vs.' Jurapê'	2.4851	9	0.0347
'Carangola' vs. 'Serra da Vista'	2.4978	12	0.0280
'Cardoso Moreira' vs. 'Chiltern Seeds'	6.0458	8	0.0003
'Cardoso Moreira' vs. 'Espirito Santo'	3.9970	15	0.0012
'Chiltern Seeds' vs. 'Cardoso Moreira-pink'	4.7901	9	0.0010
'Chiltern Seeds' vs. 'Jurapê'	8.9922	9	<0.0001
'Chiltern Seeds' vs. 'Serra da Vista'	6.0634	12	<0.0001
'CM-pink' vs. 'Espirito Santo'	4.2275	16	0.0006
'Jurapê' vs. 'Espirito Santo'	5.1070	16	0.0001
'Espirito Santo' vs. 'Serra da Vista'	4.8295	19	0.0001

Note: Only comparisons where $P < 0.05$ are shown. All pairwise comparisons with 'Dona Lourdes' and 'Antonio Dias' were highly significant ($P < 0.0001$) (see text).

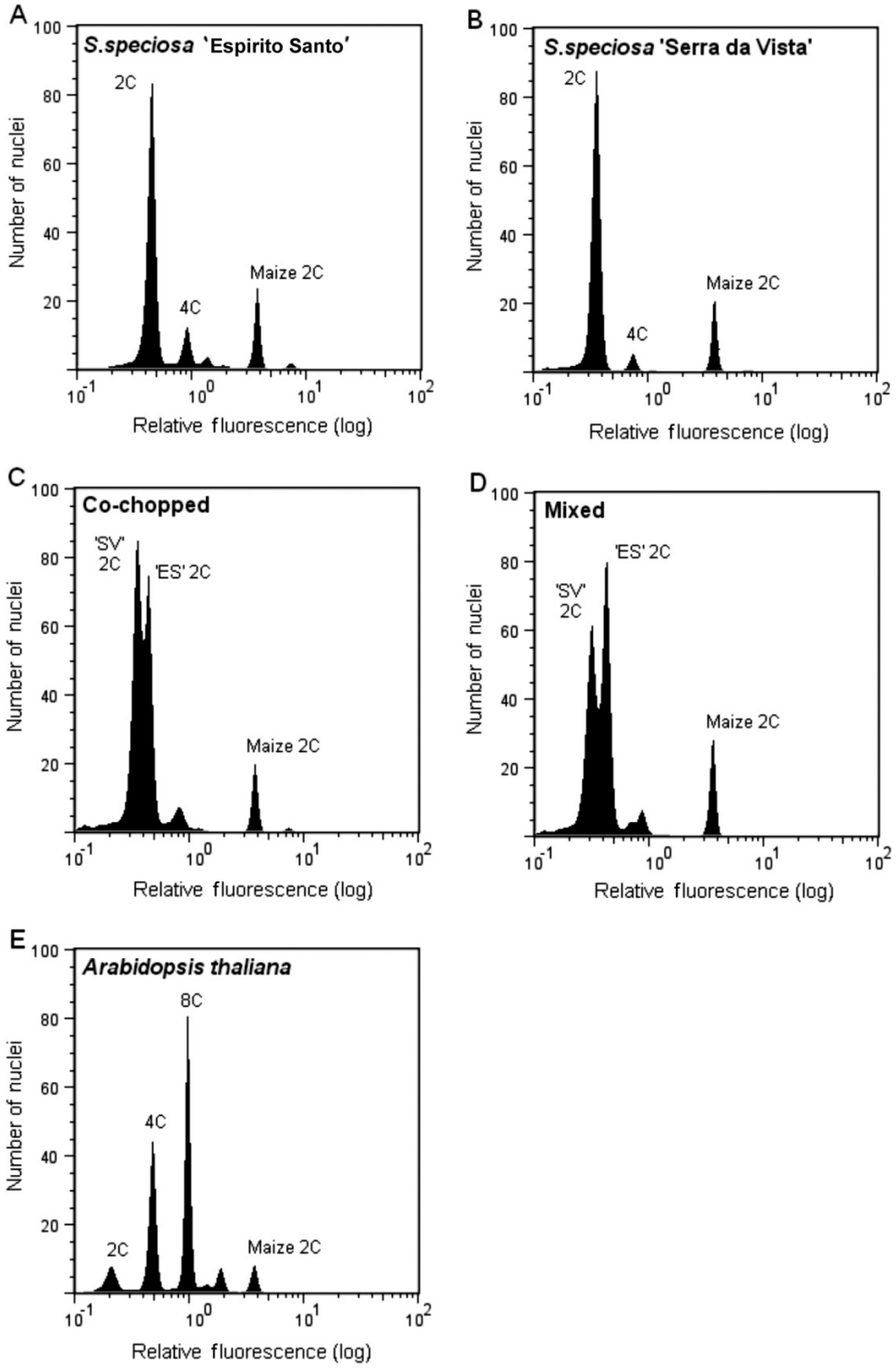
urements over the course of three runs. The resulting 1C estimate was 174 ± 8 Mbp, which is ~30% smaller than the accessions in the 250 Mbp class (Table 2), possibly making 'Antonio Dias' the sole representative of a third genome size class for the species. Therefore, it is important to repeat these measurements for this accession using more individuals. 'Antonio Dias' is a morphologically distinct collection of *S. speciosa* from Minas Gerais state, where the species is rare. This particular population is found further inland than any other documented occurrence of *S. speciosa* and is isolated from other populations (such as 'Carangola') by ~150 km (A. Chautems, personal communication, 2010). It is presently unclear whether 'Antonio Dias' is a relict population of *S. speciosa* that survived extinction in the region or if it represents a long-range dispersal to a suitable habitat. If so, 'Antonio Dias' could represent a distinct lineage within *S. speciosa* that has experienced ecogeographic isolation, a process that can lead to reproductive isolation and, in turn, speciation (Rieseberg and Willis 2007). Reproductive isolation could also account for the apparent difference in genome size (Greilhuber 1998).

The largest genomes presently known in *Sinningia* belong to the two *S. speciosa* cultivars 'Guatapara' and 'Dona Lourdes'. The origins of these cultivars are unknown, although they may well have originated in the 19th century. Both have 1C genomes of nearly 600 Mbp, which is twice the mean size determined for the wild collections (Table 2). Without cytogenetic confirmation, we cannot state with absolute certainty that these two cultivars are tetraploids, although it is strongly indicated by the genome size data. Were this finding to be independently confirmed, it would be the first reported instance of polyploidy in *S. speciosa*.

Interspecific genome size variation is common in angiosperms and has been well documented in many genera and species complexes such as *Cistus* (Ellul et al. 2002), *Hordeum* (Jakob et al. 2004), *Orobanche* (Weiss-Schneeweiss et al. 2006), *Artemisia arborescens* L. (Garcia et al. 2006), *Hieracium* (Chrtek et al. 2009), and *Tulipa* (Zonneveld 2009). However, the existence and magnitude of intraspecific genome size variation in plants is a topic of some controversy. Nuclear DNA content is generally considered to be fairly constant within diploid species (Bennett and Smith 1976; Greilhuber 1998), although there were several reports showing that genome sizes in crop species such as soybean (*Gkycine max* (L.) Merr.; Rayburn et al. 1997), peanut (*Arachis hypogaea* L.; Singh et al. 1996), *Z. mays* (Laurie and Bennett 1985), and sunflower (*Helianthus annuus* L.; Johnston et al. 1996) varied from ~11 to >20%. Upon re-examination, many of these earlier claims of substantial intraspecific genome size variation were subsequently rejected as being due to experimental error (Price et al. 2000; Greilhuber 1998). However, this is not to say that statistically significant intraspecific variation in nuclear DNA content does not exist in angiosperms. For example, more recent studies have documented variation of up to 10% in Eurasian collections of *Arabidopsis thaliana* (Schmuths et al. 2004), ~4% in cultivated accessions of pumpkin (*Cucurbita pepo* L. subsp. *pepo*; Rayburn et al. 2008), up to 13% in the bottle gourd (*Lagenaria siceraria* (Molina) Standl.; Achigan-Dako et al. 2008), ~4% in a diverse group of *Glycine max* landraces and cultivars (Rayburn et al. 2004), and ≤18.8% between European populations of the perennial grass *Festuca pallens* Host (Šmarda et al. 2008).

Certain plant secondary metabolites are known to interfere with genome size measurements obtained via flow cytometry. One of the first recognized examples of this phenomenon comes from Price et al. (2000), who showed that uncharacterized inhibitors present in *Helianthus annuus* leaves reduced the fluorescence of PI-stained *Helianthus annuus* and pea (*Pisum satvum* L.) nuclei when the tissues were processed together. Subsequent investigations implicated several classes of phenylpropanoid compounds in this effect; flavonoids, fumarocoumarins, and chlorogenic acid cause artefactual underestimates of genome size in flow cytometry through their interactions with DNA and (or) PI (Noirot et al. 2003; Walker et al. 2006; Bennett et al. 2008). The experiments of Bennett et al. (2008) and Noirot et al. (2003) are pertinent to our work with *Sinningia*. Leaf color in *S. speciosa* is quite variable (Fig. 1), and there is good reason to suspect that the red or purple pigmentation observed in collections, such as 'Serra da Vista' and 'Carangola', is due to the presence of anthocyanins. Processing pigmented leaf tissue for flow cytometry gave solutions that were pale pink or lavender in color (not shown). Bennett et al. (2008) showed that addition of the natural anthocyanin cyanidin-3-rutinoside to *Pisum staivum* nuclei prepared in Galbraith buffer (Galbraith et al. 1983) reduced PI fluorescence up to ~9%, but the effect was only significant at concentrations >50 μmol/L (Table 4 and Fig. 3 of Bennett et al. 2008). These authors also stressed the importance of using an internal standard in such studies. In our experiments, we took every precaution to avoid potential interfering effects of cytosolic compounds on genome size estimation in *Sinningia*; leaf tissue was chopped in cold buffer, filtrates were kept on ice, and nuclei were recovered by centrifugation to remove potential inhibitors present in the supernatant. *Zea mays* nuclei were added to all samples as an internal control, and peak fluorescence ratios were used to calculate genome size. When two samples were coprocessed, alternating pieces of tissue were stacked and chopped together so that all nuclei were exposed to the

Fig. 3. Flow cytometric histograms of PI-stained leaf cell nuclei from *S. speciosa*. Figures 3C and 3D clearly show that the nuclear genome of 'Espirito Santo' ('ES') is larger than that of 'Serra da Vista' ('SV'). *Arabidopsis thaliana*, which was used as a size standard in this study, is shown in Fig. 3E. The terms 2C, 4C and 8C are defined in the text. Nuclei isolated from maize (*Zea mays*) seedlings were included in all samples.



same liquid environment simultaneously. We analyzed nuclei from 'Espirito Santo' and 'Serra da Vista' prepared in this fashion, as well as a mixture of these two accessions prepared separately and combined after PI staining. Flow cytometry gave a clearly resolved, bifurcated 2C peak for both methods of preparation (Fig. 3), demonstrating unequivocally that nuclear DNA content differs between these two collections of *S. speciosa*. The use of nonsequence-specific intercalating fluorescent dyes, such as PI, has become the "gold standard" for measuring nuclear DNA content in plants by flow cytometry. We would extend this to include coprocessing of plant tissue samples from both ends of the

distribution in cases where intraspecific genome size variation is suspected. It is clear that flow cytometry can resolve small differences in genome size when samples are coprocessed. Examples are 4.5%–5% in *Festuca pallens* (2C = 5 pg; Šmarda et al 2007), 4% in tetraploid *Hieracium brachiatum* Bertol. ex DC. (2C = 7.5 pg; Suda et al. 2007), and ~12% in *Bituminaria bituminosa* (L.) C.H. Stirton (2C = 2 pg; Walker et al. 2006). We were able to detect intraspecific nuclear DNA differences in *S. speciosa* of ~20% (Fig. 3), which is approximately 0.11 pg in diploid nuclei. In *Arabidopsis thaliana*, Schmuths et al. (2004) did not analyze coprocessed samples, possibly because the reported 1.1-fold difference (0.04 pg) between the largest and smallest genomes was too small to be detected by flow cytometry. The situation was similar for *Curcurbita pepo* var. *pepo*, where 2C = 1 pg, and the ~4% reported variation would also be 0.04 pg (Rayburn et al. 2008). Approximately 4% intraspecific variation was also observed in *Glycine max* (Rayburn et al. 2004), which has a genome size of 2C = 2.5 pg. Four percent variation would amount to ~0.1 pg DNA, a difference that probably could have been resolved by flow cytometry.

Genome size in Sinningieae

The molecular phylogeny of Perret et al. (2003) divides the tribe Sinningieae into five well-supported evolutionary lineages. The 10 species included in the present study fall into three major monophyletic clades designated A (Dircaea; *S. eumorpha*), B (Corytholoma; *S. aggregata*, *S. tubiflora*, *S. pusilla*, *Sinningia richii* Clayberg, *S. harleyi*, and *S. muscicola*), and C (Sinningia; *S. speciosa*, *S. guttata*, and *S. macrophylla*), which together account for all but two or three species in *Sinningia*. Because of the limited number of species examined for genome size, no clear pattern emerges when considering this character against known species relationships. Deeper sampling of the Dircaea and Sinningia clades will be necessary to determine whether genome size has any phylogenetic relevance within Sinningieae. Of particular interest are species classified in the genera *Vanhouttea* and *Paliavana*, which are distributed, respectively, among two (C and D) and three (C, D, and E) clades each (Perret et al. 2003).

With so few genome size estimates available for species of Gesneriaceae, we can only speculate as to the range of nuclear DNA content within the family. In plants, changes in genome size result from a dynamic, bidirectional process in which some diploid lineages expand and others contract in DNA content over evolutionary time. Statistical analysis of genome size data in a phylogenetic context clearly shows this bidirectionality for diploid species of Brassicaceae, a family in which small genomes predominate and where larger genomes (four to five times the inferred ancestral genome size) occur in only a few lineages (Lysak et al. 2009). A similar interpretation was reached in an earlier study of the cotton tribe (Gossypieae), where 3 of 10 lineages showed significant increases and 4 showed deceases in genome size over time (Wendel et al. 2002). Our results suggest that a very small nuclear genome (~300 Mbp) is the common state in *Sinningia* (Table 2). This is in contrast to the reported genome sizes for the other six species of Gesneriaceae examined: 662 Mbp in *Streptocarpus cyaneus* (Bennett and Leitch 2005), 732 Mbp in *Saintpaulia ionantha* (Loureiro et al. 2007), 1372 Mbp in *Haberlea rhodopensis* (Zonneveld et al. 2005), and 1125 and 1278 Mbp in diploid species of *R. nathaliae* and *R. myconi*, respectively (Siljak-Yakovlev et al. 2008). Although the trend appears to be toward larger nuclear genomes for Old World versus New World species, more genome size estimates in the Gesneriaceae are required before valid hypotheses regarding evolution and family origins can be tested. Species in the New World subfamily Gesnerioideae (tribes Episcieae, Gesnerieae, and Gloxineae), the Coronantheroid genera (e.g., *Coronanthera*, *Mitraria*, and *Fieldia*), and diploid representatives of Didymocarpoid genera from Africa, Asia, and the paleotropics (e.g., *Streptocarpus*, *Saintpaulia*, *Petrocosmea*, *Chirita*, and *Cytrandra*) remain to be sampled (for taxonomy see Weber 2004 and Möller et al. 2009).

Characterizing the *S. speciosa* genome

We obtained single-pass DNA sequence data from 1210 random DNA fragments cloned from total *S. speciosa* DNA. The average G + C content was 37.11%, which is within the range reported for sequenced angiosperm genomes (Ming et al. 2008). All sequences were queried against the European Bioinformatics Institute protein sequence database, which resulted in 610 BLASTX matches with *E* values between 1×10^{-15} and 1×10^{-5}. Approximately 27% of the GSS showed homology to predicted plant proteins, and 72% of these had *E* values less than 1×10^{-15}. Removing the 106 organellar clones (8.8% of the total) left 600 of 1104 GSS (54%) that either had no match in the database or had protein homologies with *E* values $> 1 \times 10^{-5}$.

The 327 GSS with predicted homologies to putative plant proteins were categorized by plant family or genus that gave the most significant match (lowest *E* value). As expected, the majority of database hits were to dicotyledonous species, with 83.5% coming from six families (Brassicaceae, Fabaceae, Euphorbiaceae, Solanaceae, Vitaceae, and Salicaceae) (Table 3). The concentration of hits from the Salicaceae, Vitaceae, and Euphorbiaceae undoubtedly reflects the extent of genomic resources available for *Populus trichocarpa*, *Vitis vinifera*), and *Ricinus communis*, respectively (Tuskan et al. 2006; Jaillon et al. 2007; Foster et al. 2010). There were also 20 clones with protein homologies to species in other dicot genera such as *Antirrhinum*, *Citrus*, *Catharanthus*, *Ipomoea*, *Malus*, *Daucus*, *Sesamum*, *Gossypium*, and *Vaccinium* (not shown). Only one clone was identified that had homology to a Gesneriaceae sequence; *bdn1* is a dehydrin-like gene from the Chinese species *Boea crassifolia* Hemsley in F.B. Forbes & Hemsley (GenBank accession AF190474). Another 34 clones had predicted protein homologies to sequences from the monocotyledonous genera *Oryza*, *Sorghum*, *Zea*, and *Musa*, although the actual number is much higher because many of the predicted RT sequence matches were to retroelements from the rice (*Oryza sativa* L.) genome. One clone classified with the RT-like sequences had significant homology (both nucleotide and protein) to an endogenous copy of *petunia vein clearing virus* (PVCV; *Caulimoviridae*) isolated from *Petunia ×hybrida* Vilm. The PVCV provirus exists in multiple, dispersed, repeated copies (most of which appear to be methylated) in the *P. ×hybrida* genome and was shown to be inducible and infectious (Ri-

chert-Pöggeler et al. 2003). The presence of PVCV-like sequences in the genome of *S. speciosa* could indicate the existence of a pararetrovirus with a host range that includes *Sinningia* or, possibly, an ancient infection and integration via horizontal transfer from another plant host.

Further support for a small genome in *S. speciosa* comes from comparing our GSS results with those for *Arabidopsis*. Rabinowicz et al. (2005) sequenced small insert DNA libraries from 16 plant species of varying ploidy level and genome size (including *Arabidopsis*), and showed that hypermethylation of repeated sequences in the genomes of higher plants can be exploited in gene enrichment strategies (''methyl filtration''). For *Arabidopsis thaliana*, BLASTX searches of 1089 GSS (mean length of 520 bases) gave 22% matches to known plant proteins. Our finding of 27% predicted protein matches in 1210 *S. speciosa* GSS is slightly higher than the results for *Arabidopsis* and could be due to the increase in the number and type of plant genomic sequences that have become available since 2005. Also, we did not correct for potential pseudogenes and truncated genes nor did we account for edge effects that would result from the relatively short length of the cloned DNA inserts. Rabinowicz et al. (2005) also found that ~23% of their clones contained sequences matching known plant repeats (BLASTN *E* value cutoff = 1×10^{-10}). For the *S. speciosa* GSS, 14% (171 of 1210) of the translated sequences had matches to RT proteins present in the uniprot database, an estimate that doesn't include any unknown RT families unique to *Sinningia*. For comparison, we also generated 203 methyl-filtered GSS from the original ligation reaction by using an *Escherichia coli* (Migula) Castellani & Chalmers strain (NEB 5-alpha; New England Biolabs) that degrades DNA containing 5-methylcytosine (*McrBC*$^+$) (not shown). BLASTX searches identified 72 clones (35.5%) that gave matches at $E < 1\times 10^{-5}$ to putative plant proteins, which represents a 31% enrichment for genic sequences compared with the unfiltered library. Only four sequences (2%) were found that had homology to known plant retroelements. This is an 86% reduction in the number of RT-containing GSS and could indicate that the majority of such sequences in the genome of *S. speciosa* are probably methylated.

Genome size in *Antirrhinum majus*

In planning the genome size experiments, we had intended to include a species with a genome larger than that of *Solanum lycopersicum* to extend the range of our size standards. The reported genome size of *A. majus* is 1C = 1.6 pg = 1580 Mbp (Bennett and Smith 1976; Bennett and Leitch 2005), making this species a good choice. A preliminary investigation showed that nuclei prepared from a cultivated *A. majus*, 'Burpee's Tall Mix' (W. Atlee Burpee & Co., Warminster, Pa.), gave a prominent 2C peak that was nearly coincident with the 8C peak from *Arabidopsis thaliana* (1256 Mbp) (Fig. 4). To further investigate this observation, we grew plants from seed of three wild accessions of *A. majus* (Table 1). Flow cytometry of leaf cell nuclei from PI Nos. 420370, 542417, and 603105 revealed that the 2C peaks were also very similar in size to the *Arabidopsis thaliana* 8C peak, indicating that the genome of *A. majus* is much smaller than the published value. All *A. majus* preparations also gave a peak of twice the 2C fluorescence intensity that was not due to aggregated nuclei. These peaks were of much lower amplitude than the 2C peaks, accounting for <10% the number of events recorded by the flow cytometer. We used leaves from near the shoot apex, so the observed 4C peaks could represent nuclei in the G_2 phase of the cell cycle, or they could result from endopolyploidy (Lee et al. 2009). We never observed peaks >4C in nuclei preparations from *A. majus*. Based on 17 measurements from the three wild accessions, we conclude that the 1C genome size of diploid *A. majus* is 633 ± 33 Mbp, which is 40% of the original estimate. Our calculated genome size for PI No. 542417 was higher than for either of the other two accessions (~8% greater than PI No. 420370), and it is possible that intraspecific genome size variation exists within this species. Because *A. majus* was not the focus of the present study, this was not pursued further.

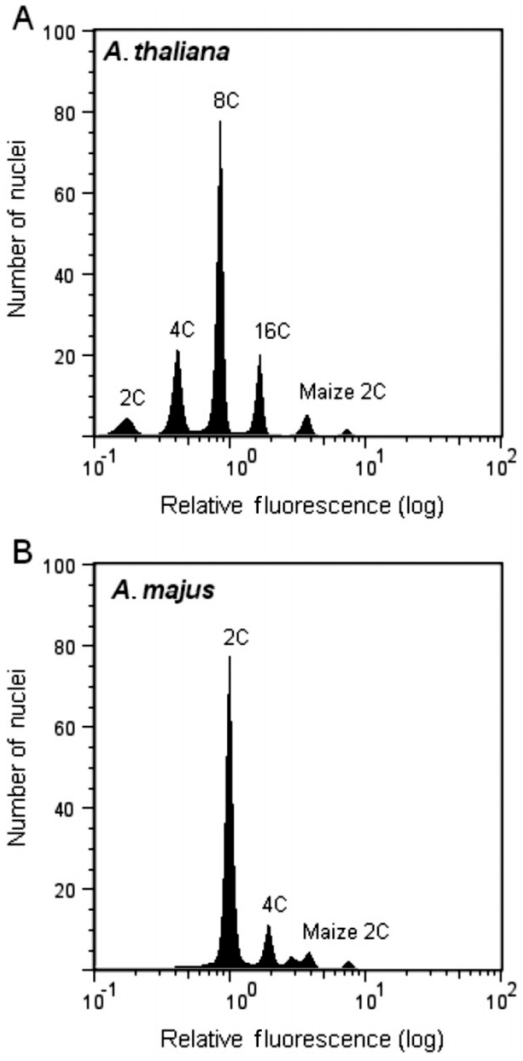


Fig. 4. Histograms of PI-stained leaf cell nuclei from *Arabidopsis thaliana* (A) and cultivated *Antirrhinum majus* 'Burpee's Tall Mix' (B). Comparison of the two panels clearly shows that the 2C peak of *Antirrhinum majus* has slightly higher fluorescence than the 8C peak of *Arabidopsis thaliana* (~1256 Mbp), indicating that the 1C genome of *Antirrhinum majus* is much smaller than 1580 Mbp as reported by Bennett and Smith (1976). Note that the unlabeled peak in Fig. 4A corresponds to *Zea mays* 4C nuclei.

We can only speculate as to why our genome size results for *A. majus* do not agree with those of Bennett and Smith (1976). The two methods used to estimate genome size (laser-based flow cytometry versus Feulgen microdensitometry) are very different, and the studies were conducted more than 30 years apart by different people in widely separated venues. We favor our present result because there are a number of well-known sources of potential error inherent to the Feulgen method, some of which could be species specific (Bennett and Smith 1976). With the exception of certain plant secondary metabolites that tend to reduce PI fluorescence (discussed above), laser-based flow cytometry is methodologically straightforward and has generally proven to be a reliable method for plant genome size estimation. The angiosperm DNA C-values database (Bennett and Leitch 2005) gives a single citation for genome size in *A. majus*, although a manual search of the literature revealed that Galbraith et al. (1983) published an estimate based on flow cytometry. Even though Galbraith and coworkers used the antibiotic fluorochrome mithramycin, which is known to bind preferentially to GC-rich regions of DNA (Doležel and Bartoš 2005), rather than the nonbase-specific PI, their genome size of 1.03 pg/nucleus (1C = 504 Mbp) is much closer to our estimate than is the original of Bennett and Smith 1976.

How do we account for intraspecific genome size variation in *S. speciosa*? Within monophyletic groups of higher plants, it is well established that genome size variation (in diploids) is mainly due to the net gain or loss of repeated sequences, mainly class I long terminal repeat (LTR) RTs, which can comprise well over 50% of genomic DNA (SanMiguel et al. 1996; Bennetzen et al. 2005; Hawkins et al. 2006). The LTR RTs are widespread mobile elements that are classified based on their terminal repeat sequences and the order of gene products encoded in the internal regions, and individual species can harbor many diverse families (Vitte and Bennetzen 2006; Wicker et al. 2007). The LTR-type RTs play a major role in genome expansion, although there are also opposing mechanisms in some species to eliminate repetitive DNA and counteract "genomic obesity" (Bennetzen et al. 2005). The emerging picture of RT proliferation in plant genomes is that it often occurs episodically, in "bursts" where some families of RT elements can be preferentially amplified over others. Retrotransposon amplification can also be specific to certain evolutionary lineages, as has been shown in *Gossypium*, *Oryza australiensis* Domin, and three hybrid *Helianthus* species (Hawkins et al. 2006; Piegu et al. 2006; Ungerer et al. 2009). In our survey of the *S. speciosa* genome, we identified 158 random DNA clones that contained sequences with predicted protein homology to plant LTR RT elements. Using the parameters given in Hawkins et al. (2006) for *Gossypium* and a genome size of 285 Mbp, we calculated an abundance of ~1645 *gypsy*-like and ~3400 *copia*-like elements per haploid genome based on the 158 RT-like GSS. These two elements account for ~34 Mbp (~12%) of genomic DNA in *S. speciosa*, which is similar to the LTR RT fraction of *Arabidopsis* (~18%) and ~16% predicted for the 200 Mbp genome of *Fragaria vesca* (Liu and Bennetzen 2008; Pontaroli et al. 2009). Our calculation of ~12% for *S. speciosa* is certainly a minimal estimate because of the relatively short sequence reads, the size of the data set, and the fact that we could not include any other classes of repeats or RTs unique to *Sinningia*. Thus, we can hypothesize that differences in RT content could account for the intraspecific genome size variation found between different populations of *S. speciosa*. It would be interesting to compare the types and relative frequencies of LTR RTs that are present in the major *S. speciosa* lineages. Genome sequencing or resequencing in this species could provide an excellent opportunity to define the nature and disposition of ~50 Mbp of genomic DNA, the difference in genome size between *S. speciosa* accessions such as Jurapê and Búzios. Also of interest is whether the genome is becoming larger through accumulation of additional RTs or whether it is decreasing in size because of specific removal of these and (or) other repeated sequences.

The small genome of *S. speciosa* is but one of several factors that make it an excellent candidate for genomic research. The species' basic biology is quite different from that of the *Arabidopsis thaliana*; *S. speciosa* is a long-lived perennial (not an ephemeral annual) with large flowers that accumulate anthocyanins, and most importantly, it produces a prominent, long-lived tuber. Tuber induction and initiation have been studied extensively in the potato (*Solanum tuberosum* L.), and our understanding of the molecular signaling events leading to tuberization has expanded greatly in the past decade (Hannapel 2007). Diploid potato species are highly heterozygous, self-incompatible outbreeders, and all cultivated varieties are polyploid (Bradshaw et al. 2006). Therefore, *Solanum tuberosum* is not readily amenable to genetic investigations because of its complex patterns of inheritance (Fernie and Willmitzer 2001; Prat 2010). Tuberization in *Solanum tuberosum* is a complex process involving multiple levels of regulation in the perception of environmental signals that redirect the developmental fate of meristematic cells in the stolon tip (Prat 2010). The "exact mechanism" of tuber induction is presently unknown (Hannapel 2007), and it will be difficult to identify the genes involved using forward genetic methods. Unlike *Solanum tuberosum*, *S. speciosa* is completely self-compatible and does not suffer from inbreeding depression in the development of homozygous lines. *Sinningia speciosa* is also easily grown under artificial lighting, has a relatively short generation time (4–6 months), produces thousands of small seeds that do not require stratification, and can be readily transformed by *Agrobacterium tumefaciens* Smith & Townsend (Zhang et al. 2008). Although tuberization in *Solanum tuberosum* requires daylengths shorter than a critical photoperiod with the stolon tips in the dark (Ewing and Wareing 1978), no such photomorphogenic requirements are known for tuber induction in *Sinningia*.

Recent technological advances in DNA sequencing have drastically reduced the cost, level of effort, and the time required to generate previously unimaginable amounts of sequence data at single base-pair resolution. The several competing massively parallel, high-throughput platforms have made whole genome sequencing available to a wider community of scientists and should open the door to genomic studies in many nonmodel organisms (Lister et al. 2009). For higher plants, such next-generation sequencing can be expected to have a major impact in the areas of phylogenetics and comparative genomics. A recent discussion

on prioritizing species for whole genome sequencing advocated a systematic approach to the selection process, one based on a "robust phylogeny" (Jackson et al. 2006). A genome sequence for *S. speciosa* would provide a reference genome for the Gesneriaceae and would enable evolutionary comparisons to *Solanum lycopersicum*, *Solanum tuberosum*, and monkeyflower (*Mimulus*), the sequenced genomes in the Euasterid I clade (lamiids) of angiosperms (Mueller et al. 2009; Joint Genome Institute 2010; Potato Genome Sequencing Consortium 2010). Sequencing *S. speciosa* could also potentially enable the discovery of novel plant-specific genes, such as those involved in tuberization. Tubers are fundamental storage and survival organs that have evolved independently in many angiosperm families, and their importance to the human diet cannot be overstated. To illustrate this, Mansfeld's World Database of Agricultural and Horticultural Crops (Liebniz Institute of Plant Genetics and Crop Plant Research 1998–2010) lists 191 tuberous angiosperm taxa in 77 genera (14 monocot families and 20 dicot families) grown for food or medicinal value, and this list does not include noncrop or ornamental species. Thus, to even consider engineering nontuberous crops (i.e., *Solanum lycopersicum*) to store carbon for novel applications such as biofuels, genes involved in tuber induction and development will need to be identified and the pathways characterized. Such knowledge will also make it possible to determine whether the genetics of tuberization differs significantly between *Sinningia* and important crop species such as *Solanum tuberosum*.

Acknowledgements

The authors thank Mauro Peixoto for generously providing seeds of *Sinningia* ssp. and for his boundless enthusiasm for the plants of his native Brazil; John Boggan, Alain Chautems, and John R. Clark for botanical information and for critically reading the manuscript; and Jennifer Webb and Kelley Davis for excellent assistance with DNA sequencing. We are also indebted to Professor C.L. Schardl for partial support of this research under US Department of Agriculture CSREES special grant 2009-34457-20125.

References

Achigan-Dako, E.G., Fuchs, J., Ahanchede, A., and Blattner, F.R. 2008. Flow cytometric analysis in *Lagenaria siceraria* (Cucurbitaceae) indicates correlation of genome size with usage types and growing elevation. Plant Syst. Evol. **276**(1–2): 9–19. doi:10.1007/s00606-008-0075-2.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic local alignment search tool. J. Mol. Biol. **215**(3): 403–410. PMID:2231712.

Arumuganathan, K., and Earle, E.D. 1991. Nuclear DNA content of some important plant species. Plant Mol. Biol. Rep. **9**(3): 208–218. doi:10.1007/BF02672069.

Bennett, M.D., and Leitch, I.J. 2005. Angiosperm DNA C-values database, release 6.0 edition, Oct. 2005 [online]. Board of Trustees of the Royal Botanic Gardens, Kew. Available from http://data.kew.org/cvalues/ [accessed 28 April 2010].

Bennett, M.D., and Smith, J.B. 1976. Nuclear DNA amounts in angiosperms. Philos. Trans. R. Soc. Lond. B. Biol. Sci. **274**(933): 227–274. doi:10.1098/rstb.1976.0044.

Bennett, M.D., Leitch, I.J., Price, H.J., and Johnston, J.S. 2003. Comparisons with *Caenorhabditis* (~100 Mb) and Drosophila (~175 Mb) using flow cytometry show genome size in *Arabidopsis* to be ~157 Mb and thus ~25% larger than the Arabidopsis Genome Initiative estimate of ~125 Mb. Ann. Bot. (Lond.), **91**(5): 547–557. doi:10.1093/aob/mcg057.

Bennett, M.D., Price, H.J., and Johnston, J.S. 2008. Anthocyanin inhibits propidium iodide DNA fluorescence in *Euphorbia pulcherrima*: implications for genome size variation and flow cytometry. Ann. Bot. (Lond.), **101**(6): 777–790. doi:10.1093/aob/mcm303.

Bennetzen, J.L., Ma, J., and Devos, K.M. 2005. Mechanisms of recent genome size variation in flowering plants. Ann. Bot. (Lond.), **95**(1): 127–132. doi:10.1093/aob/mci008.

Boggan, J.K. 1991. A morphological study and cladistic analysis of *Sinningia* and associated genera with particular reference to *Lembocarpus*, *Lietzia*, *Paliavana*, and *Vanhouttea* (Gesneriaceae: Gloxinieae). M.S. thesis, Cornell University, Ithaca, NY.

Bradshaw, J.E., Bryan, G.J., and Ramsay, G. 2006. Genetic resources (including wild and cultivated *Solanum* species) and progress in their utilization in potato breeding. Potato Res. **49**(1): 49–65. doi:10.1007/s11540-006-9002-5.

Burtt, B.L., and Wiehler, H. 1995. Classification of the family Gesneriaceae. Gesneriana, **1**: 1–4.

Chautems, A., Costa Lopes, T.C., Peixoto, M., and Rossini, J. 2010. Taxonomic revision of *Sinningia* Nees (Gesneriaceae) IV: six new species from Brazil and a long overlooked taxon. Candollea, **65** (2). In press.

Chrtek, J., Zahradnícek, J., Krak, K., and Fehrer, J. 2009. Genome size in *Hieracium* subgenus *Hieracium* (Asteraceae) is strongly correlated with major phylogenetic groups. Ann. Bot. (Lond.), **104**(1): 161–178. doi:10.1093/aob/mcp107.

Citerne, H., and Cronk, Q.C.B. 1999. The origin of the peloric *Sinningia*. New Plantsman, **6**: 219–222.

Doležel, J., and Bartoš, J. 2005. Plant DNA flow cytometry and estimation of nuclear genome size. Ann. Bot. (Lond.), **95**(1): 99–110. doi:10.1093/aob/mci005.

Doležel, J., Bartoš, J., Voglmayr, H., and Greilhuber, J. 2003. Nuclear DNA content and genome size of trout and human. Cytometry, **51A**: 127–128.

Eaton, T.D., Curley, J., Williamson, R.C., and Jung, G. 2004. Determination of the level of variation in polyploidy among Kentucky bluegrass cultivars by means of flow cytometry. Crop Sci. **44**(6): 2168–2174. doi:10.2135/cropsci2004.2168.

Ellul, P., Boscaiu, M., Vicente, O., Moreno, V., and Rosselló, J.A. 2002. Intra- and interspecific variation in DNA content in *Cistus* (Cistaceae). Ann. Bot. (Lond.), **90**(3): 345–351. doi:10.1093/aob/mcf194.

Ewing, E.E., and Wareing, P.F. 1978. Shoot, stolon, and tuber formation on potato (*Solanum tuberosum* L.) cuttings in response to photoperiod. Plant Physiol. **61**(3): 348–353. doi:10.1104/pp.61.3.348. PMID:16660290.

Fernie, A.R., and Willmitzer, L. 2001. Molecular and biochemical triggers of potato tuber development. Plant Physiol. **127**(4): 1459–1465. doi:10.1104/pp.010764. PMID:11743089.

Foster, J.T., Allan, G.J., Chan, A.P., Rabinowicz, P.D., Ravel, J., Jackson, P.J., et al. 2010. Single nucleotide polymorphisms for assessing genetic diversity in castor bean (*Ricinus communis*). BMC Plant Biol. **10**(1): 13. doi:10.1186/1471-2229-10-13. PMID:20082707.

Francis, D., Davies, M.S., and Barlow, P.W. 2008. A strong nucleotypic effect on the cell cycle regardless of ploidy level. Ann. Bot. (Lond.), **101**(6): 747–757. doi:10.1093/aob/mcn038.

Galbraith, D.W., Harkins, K.R., Maddox, J.M., Ayres, N.M., Sharma, D.P., and Firoozabady, E. 1983. Rapid flow cytometric

analysis of the cell cycle in intact plant tissues. Science (Washington, D.C.), **220**(4601): 1049–1051. doi:10.1126/science.220.4601.1049. PMID:17754551.

Galbraith, D.W., Harkins, K.R., and Knapp, S. 1991. Systemic endopolyploidy in *Arabidopsis thaliana*. Plant Physiol. **96**(3): 985–989. doi:10.1104/pp.96.3.985. PMID:16668285.

Garcia, S., Garnatje, T., Twibell, J.D., and Vallès, J. 2006. Genome size variation in the *Artemisia arborescens* complex (Asteraceae, Anthemideae) and its cultivars. Genome, **49**(3): 244–253. doi:10.1139/G05-105. PMID:16604107.

Greilhuber, J. 1998. Intraspecific variation in genome size: a critical reassessment. Ann. Bot. (Lond.), **82**: 27–35. doi:10.1006/anbo.1998.0725.

Greilhuber, J., Doležel, J., Lysák, M.A., and Bennett, M.D. 2005. The origin, evolution and proposed stabilization of the terms 'genome size' and 'C-value' to describe nuclear DNA contents. Ann. Bot. (Lond.), **95**(1): 255–260. doi:10.1093/aob/mci019.

Greilhuber, J., Borsch, T., Müller, K., Worberg, A., Porembski, S., and Barthlott, W. 2006. Smallest angiosperm genomes found in Lentibulariaceae, with chromosomes of bacterial size. Plant Biol (Stuttg.), **8**(6): 770–777. doi:10.1055/s-2006-924101. PMID:17203433.

Hannapel, D.J. 2007. Signalling the induction of tuber formation. *In* Potato biology and biotechnology: advances and perspectives. *Edited by* D. Vreugdenhil, J. Bradshaw, C. Gebhardt, F. Govers, M.A. Taylor, D.K.L. MacKerron, and H.A. Ross. Elsevier, Oxford, UK. pp. 237–256.

Hawkins, J.S., Kim, H., Nason, J.D., Wing, R.A., and Wendel, J.F. 2006. Differential lineage-specific amplification of transposable elements is responsible for genome size variation in *Gossypium*. Genome Res. **16**(10): 1252–1261. doi:10.1101/gr.5282906. PMID:16954538.

Heinrich, M., Rimpler, H., and Barrera, N. 1992. Indigenous phytotherapy of gastrointestinal disorders in a lowland Mixe community (Oaxaca, Mexico): ethnopharmacologic evaluation. J. Ethnopharmacol. **36**(1): 63–80. doi:10.1016/0378-8741(92)90062-V.

Hijri, M., and Sanders, I.R. 2004. The arbuscular mycorrhizal fungus *Glomus intraradices* is haploid and has a small genome size in the lower limit of eukaryotes. Fungal Genet. Biol. **41**(2): 253–261. doi:10.1016/j.fgb.2003.10.011. PMID:14732270.

Jackson, S., Rounsley, S., and Purugganan, M. 2006. Comparative sequencing of plant genomes: choices to make. Plant Cell, **18**(5): 1100–1104. doi:10.1105/tpc.106.042192. PMID:16670439.

Jaillon, O., Aury, J.M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., et al.; French–Italian Public Consortium for Grapevine Genome Characterization. 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature (Lond.), **449**(7161): 463–467. doi:10.1038/nature06148. PMID:17721507.

Jakob, S.S., Meister, A., and Blattner, F.R. 2004. The considerable genome size variation of *Hordeum* species (Poaceae) is linked to phylogeny, life form, ecology, and speciation rates. Mol. Biol. Evol. **21**(5): 860–869. doi:10.1093/molbev/msh092. PMID:15014163.

Johnston, J.S., Jensen, A., Czeschin, D.G., Jr., and Price, H.J. 1996. Environmentally induced nuclear 2C DNA content variation in *Helianthus annuus* (Asteraceae). Am. J. Bot. **83**: 1113–1120. doi:10.2307/2446194.

Johnston, J.S., Bennett, M.D., Rayburn, A.L., Galbraith, D.W., and Price, H.J. 1999. Reference standards for determination of DNA content of plant nuclei. Am. J. Bot. **86**(5): 609–613. doi:10.2307/2656569. PMID:10330063.

Joint Genome Institute. 2010. Mimulus genome project. JGI release version 1.0 edition [online]. US Department of Energy, Joint Genome Institute, Walnut Creek, Calif. Available from http://www.phytozome.org/mimulus.php [accessed 3 June 2010].

Korban, S.S., Wannarat, W., Rayburn, C.M., Tatum, T.C., and Rayburn, A.L. 2009. Genome size and nucleotypic variation in *Malus* germplasm. Genome, **52**(2): 148–155. doi:10.1139/G08-109. PMID:19234563.

Laurie, D.A., and Bennett, M.D. 1985. Nuclear DNA content in the genera *Zea* and *Sorghum*. Intergeneric, interspecific and intraspecific variation. Heredity, **55**(3): 307–313. doi:10.1038/hdy.1985.112.

Lee, H.O., Davidson, J.M., and Duronio, R.J. 2009. Endoreplication: polyploidy with purpose. Genes Dev. **23**(21): 2461–2477. doi:10.1101/gad.1829209. PMID:19884253.

Leitch, I.J., Soltis, D.E., Soltis, P.S., and Bennett, M.D. 2005. Evolution of DNA amounts across land plants (Embryophyta). Ann. Bot. (Lond.), **95**(1): 207–217. doi:10.1093/aob/mci014.

Liebniz Institute of Plant Genetics and Crop Plant Research. 1998–2010. Mansfeld's World Database of Agricultural and Horticultural Crops [online]. Liebniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany. Available from http://mansfeld.ipk-gatersleben.de [accessed 19 July 2010].

Lim, P.O., Kim, H.J., and Nam, H.G. 2007. Leaf senescence. Annu. Rev. Plant Biol. **58**(1): 115–136. doi:10.1146/annurev.arplant.57.032905.105316. PMID:17177638.

Lister, R., Gregory, B.D., and Ecker, J.R. 2009. Next is now: new technologies for sequencing of genomes, transcriptomes, and beyond. Curr. Opin. Plant Biol. **12**(2): 107–118. doi:10.1016/j.pbi.2008.11.004. PMID:19157957.

Liu, R., and Bennetzen, J.L. 2008. Enchilada redux: how complete is your genome sequence? New Phytol. **179**(2): 249–250. doi:10.1111/j.1469-8137.2008.02527.x. PMID:19086172.

Loddiges, C. 1817. *Gloxinia speciosa*. Bot. Cab. **1**: t. 28.

Loureiro, J., Rodriguez, E., Doležel, J., and Santos, C. 2007. Two new nuclear isolation buffers for plant DNA flow cytometry: a test with 37 species. Ann. Bot. (Lond.), **100**(4): 875–888. doi:10.1093/aob/mcm152.

Lysak, M.A., Koch, M.A., Beaulieu, J.M., Meister, A., and Leitch, I.J. 2009. The dynamic ups and downs of genome size evolution in *Brassicaceae*. Mol. Biol. Evol. **26**(1): 85–98. doi:10.1093/molbev/msn223. PMID:18842687.

Meirelles, S.T., Pivello, V.R., and Joly, C.A. 1999. The vegetation of granite rock outcrops in Rio de Janeiro, Brazil, and the need for its protection. Environ. Conserv. **26**(1): 10–20. doi:10.1017/S0376892999000041.

Ming, R., Hou, S., Feng, Y., Yu, Q., Dionne-Laporte, A., Saw, J.H., et al. 2008. The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). Nature (Lond.), **452**(7190): 991–996. doi:10.1038/nature06856. PMID:18432245.

Mishiba, K.-I., Ando, T., Mii, M., Watanabe, H., Kokubun, H., Hashimoto, G., et al. 2000. Nuclear DNA content as an index character discriminating taxa in the genus *Petunia sensu* Jussieu (Solanaceae). Ann. Bot. (Lond.), **85**(5): 665–673. doi:10.1006/anbo.2000.1122.

Möller, M., and Kiehn, M. 2003. A synopsis of cytological studies in Gesneriaceae. Edinb. J. Bot. **60**(03): 425–447. doi:10.1017/S0960428603000337.

Möller, M., Pfosser, M., Jang, C.-G., Mayer, V., Clark, A., Hollingsworth, M.L., et al. 2009. A preliminary phylogeny of the 'didymocarpoid Gesnericaeae' based on three molecular data sets: incongruence with available tribal classifications. Am. J. Bot. **96**(5): 989–1010. doi:10.3732/ajb.0800291.

Mueller, L.A., Lankhorst, R.K., Tanksley, S.D., Giovannoni, J.J.,

White, R., Vrebalov, J., et al. 2009. A snapshot of the emerging tomato genome sequence. Plant Genome, **2**(1): 78–92. doi:10.3835/plantgenome2008.08.0005.

Müller, J., Sprenger, N., Bortlik, K., Boller, T., and Wiemken, A. 1997. Dessication increases sucrose levels in *Ramonda* and *Haberlea*, two genera of resurrection plants in the Gesneriaceae. Physiol. Plant. **100**(1): 153–158. doi:10.1111/j.1399-3054.1997.tb03466.x.

Noirot, M., Barre, P., Duperray, C., Louarn, J., and Hamon, S. 2003. Effects of caffeine and chlorogenic acid on propidium iodide accessibility to DNA: consequences on genome size evaluation in coffee tree. Ann. Bot. (Lond.), **92**(2): 259–264. doi:10.1093/aob/mcg139.

Ochatt, S.J. 2008. Flow cytometry in plant breeding. Cytometry A, **73**(7): 581–598. PMID:18431774.

Perret, M., Chautems, A., Spichiger, R., Kite, G., and Savolainen, V. 2003. Systematics and evolution of tribe Sinningieae (Gesneriaceae): evidence from phylogenetic analyses of six plastid DNA regions and nuclear *ncpGS*. Am. J. Bot. **90**(3): 445–460. doi:10.3732/ajb.90.3.445.

Piegu, B., Guyot, R., Picault, N., Roulin, A., Saniyal, A., Kim, H., et al. 2006. Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. Genome Res. **16**(10): 1262–1269. doi:10.1101/gr.5290206. PMID:16963705.

Pontaroli, A.C., Rogers, R.L., Zhang, Q., Shields, M., Davis, T.M., Folta, K.M., et al. 2009. Gene content and distribution in the nuclear genome of *Fragaria vesca*. Plant Genome, **2**(1): 93–101. doi:10.3835/plantgenome2008.09.0007.

Potato Genome Sequencing Consortium. 2010. PGSC data release, 2010 [online]. Available from http://potatogenomics.plantbiology.msu.edu/ [accessed 3 June 2010].

Prat, S. 2010. Hormonal and daylength control of potato tuberization. *In* Plant hormones: biosynthesis, signal transduction, action! Revised 3rd ed. *Edited by* P.J. Davies. Springer, Dortrecht, Heidelberg, London, New York. pp. 574–596.

Price, H.J., Hodnett, G., and Johnston, J.S. 2000. Sunflower (*Helianthus annuus*) leaves contain compounds that reduce nuclear propidium iodide fluorescence. Ann. Bot. (Lond.), **86**(5): 929–934. doi:10.1006/anbo.2000.1255.

Rabinowicz, P.D., Citek, R., Budiman, M.A., Nunberg, A., Bedell, J.A., Lakey, N., et al. 2005. Differential methylation of genes and repeats in land plants. Genome Res. **15**(10): 1431–1440. doi:10.1101/gr.4100405. PMID:16204196.

Rayburn, A.L., Biradar, D.P., Bullock, D.G., Nelson, R.L., Gourmet, C., and Wetzel, J.B. 1997. Nuclear DNA content diversity in Chinese soybean introductions. Ann. Bot. (Lond.), **80**(3): 321–325. doi:10.1006/anbo.1997.0445.

Rayburn, A.L., Biradar, D.P., Nelson, R.L., McCloskey, R., and Yeater, K.M. 2004. Documenting intraspecific genome size variation in soybean. Crop Sci. **44**(1): 261–264. doi:10.2135/cropsci2004.0261.

Rayburn, A.L., Kushad, M.M., and Wannarat, W. 2008. Intraspecific genome size variation in pumpkin (*Cucurbita pepo* subsp. *pepo*). Hortscience, **43**(3: 949–951.

Reeves, G., Francis, D., Davies, M.S., Rogers, H.J., and Hodkinson, T.R. 1998. Genome size is negatively correlated with altitude in natural populations of *Dactylis glomerata*. Ann. Bot. (Lond.), **82**(Suppl A): 99–105. doi:10.1006/anbo.1998.0751.

Richert-Pöggeler, K.R., Noreen, F., Schwarzacher, T., Harper, G., and Hohn, T. 2003. Induction of infectious petunia vein clearing (pararetro) virus from endogenous provirus in petunia. EMBO J. **22**(18): 4836–4845. doi:10.1093/emboj/cdg443. PMID:12970195.

Rieseberg, L.H., and Willis, J.H. 2007. Plant speciation. Science (Washington, D.C.), **317**(5840): 910–914. doi:10.1126/science.1137729. PMID:17702935.

SanMiguel, P., Tikhonov, A., Jin, Y.-K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., et al. 1996. Nested retrotransposons in the intergenic regions of the maize genome. Science (Washington, D.C.), **274**(5288): 765–768. doi:10.1126/science.274.5288.765. PMID:8864112.

Schmuths, H., Meister, A., Horres, R., and Bachmann, K. 2004. Genome size variation among accessions of *Arabidopsis thaliana*. Ann. Bot. (Lond.), **93**(3): 317–321. doi:10.1093/aob/mch037.

Siljak-Yakovlev, S., Stevanovic, V., Tomasevic, M., Brown, S.C., and Stevanovic, B. 2008. Genome size variation and polyploidy in the resurrection plant genus *Ramonda*: cytogeography of living fossils. Environ. Exp. Bot. **62**(2): 101–112. doi:10.1016/j.envexpbot.2007.07.017.

Simeonova, E., Sikora, A., Charzyńska, M., and Mostowska, A. 2000. Aspects of programmed cell death during leaf senescence of mono- and dicotyledonous plants. Protoplasma, **214**(1–2): 93–101. doi:10.1007/BF02524266.

Singh, K.P., Raina, S.N., and Singh, A.K. 1996. Variation in chromosomal DNA associated with the evolution of *Arachis* species. Genome, **39**(5): 890–897. doi:10.1139/g96-112. PMID:18469942.

Skog, L.E., and Boggan, J.K. 2007. World checklist of Gesneriaceae [online]. Department of Botany, Smithsonian Institution, Washington, D.C. Available from http://botany.si.edu/Gesneriaceae/Checklist [accessed 28 April 2010].

Šmarda, P., Bureš, P., Horová, L., Foggi, B., and Rossi, G. 2007. Genome size and GC content evolution of *Festuca*: ancestral expansion and subsequent reduction. Ann. Bot. (Lond.), **101**(3): 421–433. doi:10.1093/aob/mcm307.

Šmarda, P., Bureš, P., Horová, L., and Rotreklová, O. 2008. Intrapopulation genome size dynamics in *Festuca pallens*. Ann. Bot. (Lond.), **102**(4): 599–607. doi:10.1093/aob/mcn133.

Smith, J.F., Wolfram, J.C., Brown, K.D., Carroll, C.L., and Denton, D.S. 1997. Tribal relationships in the Gesneriaceae: evidence from DNA sequences of the chloroplast gene *ndhF*. Ann. Mo. Bot. Gard. **84**(1): 50–66. doi:10.2307/2399953.

Soltis, D.E., Soltis, P.S., Bennett, M.D., and Leitch, I.J. 2003. Evolution of genome size in the angiosperms. Am. J. Bot. **90**(11): 1596–1603. doi:10.3732/ajb.90.11.1596.

Suda, J., Krahulcová, A., Trávnícek, P., Rosenbaumová, R., Peckert, T., and Krahulec, R. 2007. Genome size variation and species relationships in *Hieracium* sub-genus *Pilosella* (Asteraceae) as inferred by flow cytometry. Ann. Bot. (Lond.), **100**(6): 1323–1335. doi:10.1093/aob/mcm218.

Tuskan, G.A., Difazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., et al. 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science (Washington, D.C.), **313**(5793): 1596–1604. doi:10.1126/science.1128691. PMID:16973872.

Ungerer, M.C., Strakosh, S.C., and Stimpson, K.M. 2009. Proliferation of *Ty3/gypsy*-like retrotransposons in hybrid sunflower taxa inferred from phylogenetic data. BMC Biol. **7**(1): 40. doi:10.1186/1741-7007-7-40. PMID:19594956.

Vitte, C., and Bennetzen, J.L. 2006. Analysis of retrotransposon structural diversity uncovers properties and propensities in angiosperm genome evolution. Proc. Natl. Acad. Sci. USA, **103**(47): 17 638 – 17 643. doi:10.1073/pnas.0605618103. PMID:17101966.

Walker, D.J., Monino, I., and Correal, E. 2006. Genome size in *Bituminaria bituminosa* (L.) C.H. Stirton (Fabaceae) populations:

separation of "true" difference from environmental effects on DNA determination. Environ. Exp. Bot. **55**(3): 258–265. doi:10.1016/j.envexpbot.2004.11.005.

Weber, A. 2004. Gesneriaceae. *In* The families and genera of vascular plants. *Edited by* K. Kubitzki. Vol. VII Flowering plants — dicotyledons. Lamiales (except Acanthaceae incl. Avicenniaceae). *Volume editor*: J.W. Kadereit. Springer-Verlag, Berlin. pp. 63–158.

Weiss-Schneeweiss, H., Greilhuber, J., and Schneeweiss, G.M. 2006. Genome size evolution in holoparasitic *Orobanche* (Orobanchaceae) and related genera. Am. J. Bot. **93**(1): 148–156. doi:10.3732/ajb.93.1.148.

Wendel, J.F., Cronn, R.C., Johnston, J.S., and Price, H.J. 2002. Feast and famine in plant genomes. Genetica, **115**(1): 37–47. doi:10.1023/A:1016020030189. PMID:12188047.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., et al. 2007. A unified classification system for eukaryotic transposable elements. Nat. Rev. Genet. **8**(12): 973–982. doi:10.1038/nrg2165. PMID:17984973.

Wiehler, H. 1983. A synopsis of the neotropical Gesneriaceae. Selbyana, **6**: 1–219.

Winefield, C.S., Lewis, D.H., Swinny, E.E., Zhang, H., Arathoon, H.S., Fischer, T.C., et al. 2005. Investigation of the biosynthesis of 3-deoxyantocyanins in *Sinningia cardinalis*. Physiol. Plant. **124**(4): 419–430. doi:10.1111/j.1399-3054.2005.00531.x.

Young, N.D., Cannon, S.B., Sato, S., Kim, D., Cook, D.R., Town, C.D., et al. 2005. Sequencing the genespaces of *Medicago truncatula* and *Lotus japonicus*. Plant Physiol. **137**(4): 1174–1181. doi:10.1104/pp.104.057034. PMID:15824279.

Zhang, M.-Z., Ye, D., Wang, L.-L., Pang, J.-L., Zhang, Y.-H., Zheng, K., et al. 2008. Overexpression of the cucumber *LEAFY* homolog *CFL* and hormone treatments alter flower development in gloxinia (*Sinningia speciosa*). Plant Mol. Biol. **67**(4): 419–427. doi:10.1007/s11103-008-9330-8. PMID:18392697.

Zimmer, E.A., Roalson, E.H., Skog, L.E., Boggan, J.K., and Idnurm, A. 2002. Phylogenetic relationships in the Gesnerioideae (Gesneriaceae) based on nrDNA ITS and cpDNA *trnL-F* and *trnE-T* spacer region sequences. Am. J. Bot. **89**(2): 296–311. doi:10.3732/ajb.89.2.296.

Zonneveld, B.J.M. 2009. The systematic value of nuclear genome size for "all" species of *Tulipa* L. (*Liliaceae*). Plant Syst. Evol. **281**(1-4): 217–245. doi:10.1007/s00606-009-0203-7.

Zonneveld, B.J.M., Leitch, I.J., and Bennett, M.D. 2005. First nuclear DNA amounts in more than 300 angiosperms. Ann. Bot. (Lond.), **96**(2): 229–244. doi:10.1093/aob/mci170.