Genetica **115:** 37–47, 2002.

Feast and famine in plant genomes

Jonathan F. Wendel[1], Richard C. Cronn[1], J. Spencer Johnston[2] & H. James Price[3]
[1]*Department of Botany, Iowa State University, Ames, IA, 50011, USA (Phone: +1-515-294-7172; Fax: +515-294-1337; E-mail: jfw@iastate.edu);* [2]*Department of Entomology,* [3]*Department of Soil and Crop Sciences, Texas A & M University, College Station, TX 77843, USA*

Key words: cotton, DNA content, genome evolution, genome size, *Gossypium*, Malvaceae, molecular evolution, 2C values

Abstract

Plant genomes vary over several orders of magnitude in size, even among closely related species, yet the origin, genesis and significance of this variation are not clear. Because DNA content varies over a sevenfold range among diploid species in the cotton genus (*Gossypium*) and its allies, this group offers opportunities for exploring patterns and mechanisms of genome size evolution. For example, the question has been raised whether plant genomes have a 'one-way ticket to genomic obesity', as a consequence of retroelement accumulation. Few empirical studies directly address this possibility, although it is consistent with recent insights gleaned from evolutionary genomic investigations. We used a phylogenetic approach to evaluate the directionality of genome size evolution among *Gossypium* species and their relatives in the cotton tribe (Gossypieae, Malvaceae). Our results suggest that both DNA content increase and decrease have occurred repeatedly during evolution. In contrast to a model of unidirectional genome size change, the frequency of inferred genome size contraction exceeded that of expansion. In conjunction with other evidence, this finding highlights the dynamic nature of plant genome size evolution, and suggests that poorly understood genomic contraction mechanisms operate on a more extensive scale that previously recognized. Moreover, the research sets the stage for fine-scale analysis of the evolutionary dynamics and directionality of change for the full spectrum of genomic constituents.

Introduction

Plant genomes exhibit extraordinary variation in size, ranging from approximately 125 Mbp in *Arabidopsis thaliana*, known with exquisite precision now that the genome has been nearly completely sequenced (The *Arabidopsis* Genome Initiative, 2000), to over 120,000 Mbp in some relatives of lilies (Bennett & Leitch, 1995; Bennett et al., 1997; Leitch et al., 1998). Because this remarkable range in DNA content is not associated with variation in the basic complement of genes required for growth and development, it has become renowned as the 'C-value paradox' (Thomas, 1971). Mechanistic explanations for the observed size variation include repeated cycles of polyploidy over evolutionary time (Leitch & Bennett, 1997; Soltis & Soltis, 1999; Otto & Whitton, 2000; Wendel, 2000), leading to progressively larger genomes, as well as massive accumulation of retroelements, as elegantly shown for some species in the grass family (Bennetzen, 1996; SanMiguel et al., 1996; Bennetzen, 2000). The latter observations led naturally to a proposal that genome size evolution in plants may largely be unidirectional, from small to large (Bennetzen & Kellogg, 1997a), with an overall pattern of ratcheting upward due to the combined effects of polyploidization and retroelement accumulation.

The suggestion that plants may have a 'one-way ticket to genomic obesity' (Bennetzen & Kellogg, 1997a) provided an appealing scenario that appeared to account, at least in part, for the C-value paradox. Plants with small genomes are suggested to have been relatively effective in suppressing retrotransposon activity; whereas those with large genomes achieved this condition by release from suppression (or new invasion) and massive retroelement amplification. Be-

cause the generality of this proposal is unknown and because alternative explanations exist, the subject has attracted considerable attention (Bennetzen & Kellogg, 1997b; Petrov, 1997; Leitch et al., 1998; Vicient et al., 1999; Federoff, 2000; Rabinowicz, 2000; Shirasu et al., 2000; Petrov, 2001). The hypothesis of unidirectional genome size change has been rigorously tested in only a few cases (Bennetzen & Kellogg, 1997a; Cox et al., 1998; Leitch et al., 1998). In addition, accumulating evidence suggests that genome size contraction may also be a common evolutionary occurrence. Supporting this suggestion is recent evidence that underscores the diversity and scope of genomic deletional mechanisms (Petrov et al., 1996; Petrov, 1997; Vicient et al., 1999; Kirik et al., 2000; Petrov et al., 2000; Shirasu et al., 2000; Petrov, 2001), the phylogenetically widespread distribution of plants with small genomes (Bennett & Leitch, 1995; Bennett et al., 1997; Leitch et al., 1998), and the prevalence of high levels of infraspecific DNA content variation (Price, 1988; Kalendar et al., 2000) that are not easily reconciled with an assumption of unidirectional evolutionary change.

A prerequisite for any discussion of the directionality of genome size change is that the phylogenetic relationships of the organisms under study are understood. Phylogenetic analysis provides an essential framework for inferences of ancestral genome sizes (Bennetzen & Kellogg, 1997a) as well as for polarizing genomic changes for specific homologous genomic regions. Hence, robust phylogenies are essential for both global (genome size) and local (fine-scale) inferences of past genomic evolutionary events. A particularly useful model in this respect is the group of plants comprising the cotton genus (*Gossypium* L.) and its relatives in the small Malvaceous tribe Gossypieae Alefeld. This monophyletic (LaDuke & Doebley, 1995) tribe includes eight genera (Fryxell, 1979), the largest being *Gossypium* with approximately 50 species (Fryxell, 1992). Remarkable genome size variation exists among the 45 diploid species of *Gossypium*, with 2C DNA contents ranging from approximately 2.0 to 7.0 pg (Wendel et al., 1999). This information motivated an exploration of genome sizes among its relatives, as well as phylogenetic relationships among the genera in the tribe. Our intention was to generate the information required to understand the genome size variation that exists among diploid species within this single tribe, and thereby facilitate inferences of the directionality of genome size change. To accomplish this we estimated the phylogeny of the tribe using exemplar sampling, filled in gaps in our knowledge of chromosome numbers to ensure that only diploids were included in the study, and employed analytical methods for inferring ancestral conditions for quantitative varying characters (Maddison, 1991; Martins & Hansen, 1997; Cunningham et al., 1998; Martins, 1999; Oakley & Cunningham, 2000).

Materials and methods

Organismal context and taxon sampling

Included in the eight genera of the Gossypieae are four small genera with restricted geographic distributions (Fryxell, 1979). *Lebronnecia* Fosberg, from the Marquesas Islands, is a monotypic genus, as is *Cephalohibiscus* Ulbrich, from New Guinea and the Solomon Islands. *Gossypioides* Skovsted ex J.B. Hutchinson contains two species from East Africa and Madagascar. *Kokia* Lewton, endemic to the Hawaiian Islands, includes three extant and one extinct species. In addition to these four small genera, the tribe includes four moderately sized genera with broader geographic ranges: *Hampea* Schlechtendal comprises 21 neotropical species; *Cienfuegosia* Cavanilles includes 25 species with an aggregate range that includes the neotropics and parts of Africa; 17 species are recognized in the pantropical genus *Thespesia* Solander; and *Gossypium*, as noted above, is the largest genus in the tribe with about 50 species.

An initial phylogenetic scaffolding for the tribe was erected several years ago based on sequences from the chloroplast gene *ndhF* and from nuclear ribosomal ITS sequences (Seelanan et al., 1997). These analyses provided unequivocal support for several relationships among genera but left other aspects of their evolutionary history unclear. Particularly strongly supported was the close relationship between *Gossypioides* and *Kokia* and their collective sister-taxon relationship to *Gossypium*. *Thespesia* was implicated to be at least biphyletic and perhaps polyphyletic, with a portion being sister to *Lebronnecia* and perhaps *Hampea* and the remainder being cladistically unresolved along with *Cienfuegosia* and the (*Hampea* + *Lebronnecia* + *Thespesia*- in part) clade. *Cephalohibiscus* was not included in these analyses as it was not available. This earlier work provided some insights into the evolutionary history of the group, but left unanswered several key questions of branching order, particularly as regards the earliest divergences within the tribe.

Table 1. Species in the Gossypieae used for phylogenetic analysis, chromosome counts, and genome size measurements[a]

Taxon	Chromosome no. ($2n$)	Genome size (2C)
C. hitchcockii (Ulbrich ex Kearney) Blanchard	*20*	2.3
C. tripartita H.B.K. Gürke	*20*	1.9
C. yucatanensis Millspaugh	*22*	2.0
G. kirkii (Mast.) J.B. Hutchinson	24	1.2
G. herbaceum L.	26	3.7
G. raimondii Ulbrich	26	2.0
H. appendiculata (J. Donnell-Smith) Standley	26	5.9
K. drynarioides (Seemann) Lewton	24	1.2
L. kokioides Fosberg	*26*	3.6
T. lampas (Cavanilles) Dalzell ex Dalzell and Gibson	26	3.2
T. populnea (L.) Solander ex Correa	*26*	8.2
T. thespesioides (R. Brown ex Bentham) Fryxell	*26*	3.2
M. sylvestris L.	42	3.0

[a]All species were included in the molecular phylogenetic analysis with the exception of *C. hitchcockii* and *C. yucatanensis*. Chromosome counts determined in this study are in italics; others are from the literature. Genome sizes were estimated in the present study, except for the outgroup species *M. sylvestris*, which is from Bennett et al. (2000).

Chromosome numbers have been reported for six of the eight genera (all except *Cephalohibiscus* and *Lebronnecia*). As summarized in Fryxell (1979), most species are diploid, with somatic numbers ranging from $2n = 20$ or 22 (*Cienfuegosia*) to $2n = 24$ or 26 (other genera). Polyploidy is unknown in the tribe outside of one species of *Thespesia* (*T. populneoides*) (Miege & Josserand, 1972) and five species of *Gossypium*; the latter, in fact, provide one of the classic models of allopolyploid evolution (Wendel et al., 1999; Wendel, 2000). DNA content has been determined only for members of *Gossypium*, where a rich history of genetic and cytogenetic work has stimulated a systematic survey of genome sizes (summarized in Endrizzi et al., 1985; Wendel et al., 1999). These studies demonstrate a 3.5-fold range in genome sizes among closely related diploid species in this single genus, with values ranging from a low of 2.0 pg per 2C nucleus for the 'D-genome' species from the New World to a high of approximately 7.0 pg per 2C nucleus for the 'K-genome' taxa from NW Australia. Genome sizes are relatively homogeneous *within* genome groups (summarized in Endrizzi et al., 1985; Wendel et al., 1999) and each is also monophyletic (Wendel & Albert, 1992; Seelanan et al., 1997; Wendel et al., 1999). The five allopolyploid species have genome sizes that are additive with respect to those of their diploid progenitor genomes.

To further resolve the phylogenetic history of the group and to better develop the organismal framework as a model for studies of genome size evolution, we conducted molecular phylogenetic analysis, made new chromosome counts to confirm diploidy and fill in gaps in the data, and measured genome sizes outside of the genus *Gossypium*. The species evaluated (Table 1) in the molecular phylogenetic analysis included one representative of each of the available genera in the tribe, except that two species of *Gossypium* were sampled, which vary twofold in DNA content, and three species of *Thespesia* were sampled because of their variable 2C DNA content and previous suggestions of generic paraphyly or polyphyly. To supplement the existing data base, chromosome counts were obtained for three species of *Cienfuegosia*, two species of *Thespesia*, and *L. kokioides*. Genome sizes were estimated for all of the above. For rooting phylogenetic trees, *Malva sylvestris* L. was included as an outgroup.

Phylogenetic analysis

An approximately 1 kb segment of the 5′ end of cellulose synthase A1 (*CesA1*) was amplified and sequenced as described in Cronn et al. (1999). This nuclear gene was selected because Southern hybridization surveys indicated that it was single-copy and because previous work established both experimental protocols (conserved primer regions; amplification conditions) and potential phylogenetic utility. The amplified portion spans exon 1 through exon 5 and

consists of approximately 60% exon nucleotide positions. Gaps caused few problems in aligning sequences, which were readily aligned using ClustalW (Johnson et al., 1994). The aligned matrix was subjected to both maximum likelihood and maximum parsimony analysis using PAUP* 4.0b8 (Swofford, 2001). Nucleotide alignments are available at http://www.botany.iastate.edu/~jfw/HomePage/jfwdata_sets.html. *CesA1* sequences have been deposited in GenBank under accession numbers AF139442, AF139444, AF201886, AF376040, and AF376042–AF376048. Parsimony analysis was conducted using a branch and bound search strategy with equal weighting of all character-state transformations and alignment gaps treated as missing data. Support for resolved clades was evaluated by jackknife resampling (Farris, 1996) using 1000 replications and branch and bound searches. Maximum likelihood analysis was performed using the HKY85 model of sequence evolution and empirically derived estimates of base frequencies ($A = 0.29$; $C = 0.17$; $G = 0.21$; $T = 0.33$), transition/transversion rates (1.55), and the gamma shape parameter (1.56).

Genome size estimation

Newly expanded leaves from greenhouse-grown plants were manually diced to release nuclei as described (Johnston et al., 1999). In all cases, a leaf of the internal standard *Pisum sativum* cv. Minerva Maple was included in nuclear isolation. Chopped leaves were filtered through a 53 μm nylon mesh and propidium iodide was added to a final concentration of 50 ppm. The mean fluorescence of nuclei was quantified using a Coulter Epics Elite (Coulter Electronics, Hialeah FL) flow cytometer equipped with a water-cooled laser tuned at 514 nm and 500 mW. Fluorescence at >615 nm was detected with a photomultiplier screened by a long pass filter. Mean 2C DNA content of each target species was calculated by comparing its mean fluorescence with the mean fluorescence of the standard pea nuclei, which has a 2C value of 9.56 pg DNA (Johnston et al., 1999).

Chromosome counts

Chromosome counts were obtained from enzymatically-digested root tips, pretreated in a saturated aqueous solution of alpha bromonaphthaline for 90 min, as described by (Hanson et al., 1996) and (Jewell & Islam-Faridi, 1994).

Results

Phylogenetic analysis of CesA1

Sequence data were generated for an approximately 1 kb portion of the cellulose synthase gene *CesA1*, including 615 exon nucleotides from the first five exons and between 381 and 460 intron positions from introns 1 through 4. Absolute lengths varied among taxa from 996 bp in the outgroup, *M. sylvestris*, to 1075 bp in *G. kirkii*. The aligned data matrix contained 1146 nucleotide positions, including 67 parsimony-informative characters.

Maximum parsimony analysis led to the recovery of a single tree (Figure 1) that is fully resolved and has high internal consistency with respect to the data (consistency index = 0.92; retention index = 0.81). As shown in Figure 1, most clades are relatively strongly supported, with all but one being supported by five or more character-state changes and having jackknife support greater than 75%. The topology shown extends and corroborates the initial phylogenetic scaffolding for the tribe Gossypieae based on the chloroplast gene *ndhF* and nuclear ribosomal ITS sequences (Seelanan et al., 1997). As expected from the earlier analyses, *Kokia* and *Gossypioides* are sister-genera, these two being collectively sister to *Gossypium*. This clade of three genera is resolved as sister to a *Thespesia* + *Hampea* + *Lebronnecia* clade, with the latter clade additionally supported by a synapomorphic one-codon deletion (lysine) in exon 4 (the only indel in the coding region of the *CesA1* data set). The internal structure of the taxa included in the *Thespesia* + *Hampea* + *Lebronnecia* grouping is not robustly resolved, as there is only weak support for the basal separation of *H. appendiculata* from the clade composed of *L. kokioides*, *T. lampas*, and *T. thespesioides*. Another point of agreement with the earlier analysis of Seelanan et al. (1997) is that *Thespesia* is paraphyletic or polyphyletic, suggesting once again the need for a more thorough study of this group. Finally, the *CesA1* data clearly and strongly support the basal separation of *Cienfuegosia* from the remainder of the tribe, a position not unexpected given its unusual chromosome numbers (Table 1) and divergent morphology (Fryxell, 1979).

Maximum likelihood analysis yielded the same topology as that recovered from parsimony analysis, indicating that the topology is unaffected by choice of phylogeny reconstruction method. As shown by the branch lengths in Figure 1, there was little het-

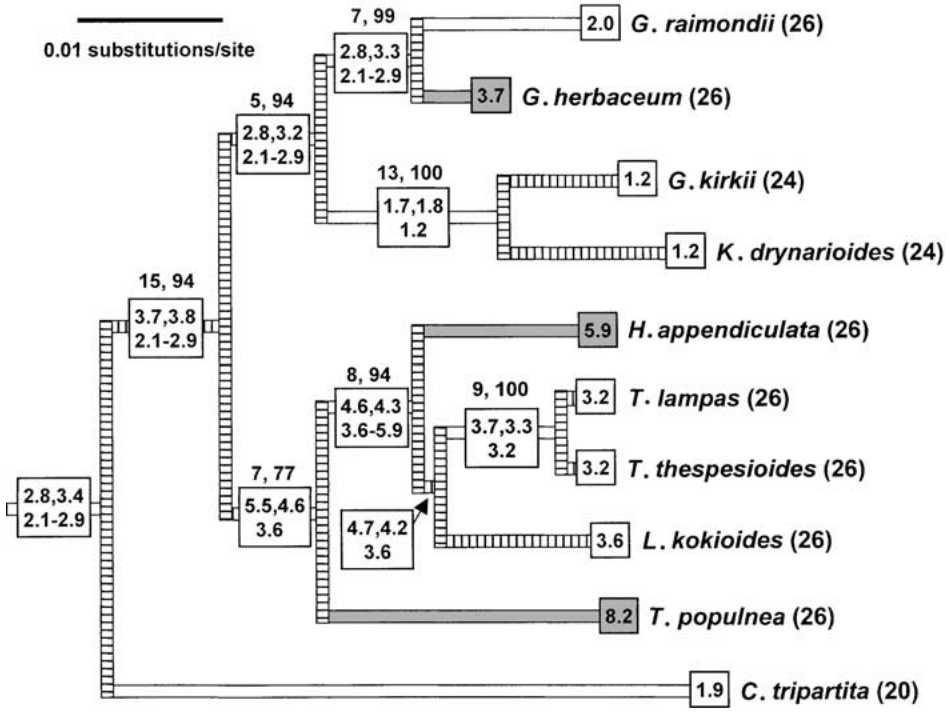

Figure 1. Genome size evolution in the cotton tribe (Gossypieae). The shortest tree recovered from parsimony analysis, inferred from *CesA1* sequence data, is identical in topology to the maximum likelihood phylogeny shown here. The number of character-state changes and jackknife support (%) from maximum parsimony analysis are shown *above* each internal branch, except for the *T. lampas* through *L. kokioides* clade, which was supported only by a single character (in addition to a one-codon deletion). The tree was rooted with the outgroup *M. sylvestris* (ingroup-outgroup branch length = 0.12). Genome sizes are shown at branch tips before species names, which are followed by somatic chromosome numbers. Ancestral genome sizes were estimated using sum-of-squared-changes parsimony analysis (Maddison, 1991), a generalized least squared method (Martins & Hansen, 1997), and linear (= Wagner) parsimony (Swofford & Maddison, 1987). These three estimates are shown in boxes on the internal branches, with the first two on the top line and the linear parsimony reconstructions on the bottom line (including the full range from minimum to maximum in cases where there was a range of equally parsimonious reconstructions). Inferred genome size *increases* are shown by shaded branches, *decreases* are indicated by un-filled branches, and ambiguities or stasis are denoted by hatched branches.

erogeneity in evolutionary rates for *CesA1* among the taxa studied. To formally evaluate this suggestion, we used the Kishino–Hasegawa likelihood ratio test to evaluate whether the *CesA1* sequences exhibited clocklike behavior; log likelihoods with and without clock enforcement were not significantly different (-ln L, no clock = 3108.03; molecular clock enforced = 3113.26; $X^2 = 10.459$; $p = 0.16$), consistent with equal evolutionary rates.

Genome size estimation

Flow cytometry was used to measure DNA content in all taxa included in the phylogenetic analysis. Replicate measurements from individual species were highly repeatable with low standard errors, and internal controls yielded the expected genome sizes (data not shown). Measured 2C values are shown in Table 1. These data demonstrate a nearly sevenfold variation in DNA content between the largest (*T. populnea*; 2C = 8.2 pg) and smallest (*G. kirkii* and *K. drynarioides*, each with 2C = 1.2 pg) genomes measured. Whereas relatively little variation was observed among the three *Cienfuegosia* species measured, the two *Gossypium* species vary nearly twofold in DNA content and the three species of *Thespesia* exhibit greater than a 2.5-fold difference in genome sizes.

Chromosome counts

To verify that the observed genome size variation reflects changes at the same ploidy level, that is, that they do not result from polyploidy, chromosomes counts were obtained for *Cienfuegosia tripartita* ($2n = 20$), *C. hitchcockii* ($2n = 20$), *C. yucatanensis* ($2n = 22$), *T. populnea* ($2n = 26$), *T. thespesioides*

($2n = 26$), and *L. kokioides* ($2n = 26$). These data, in conjunction with counts reported elsewhere (Fryxell, 1979), confirm that all plants included in the study are diploid (Table 1). Accordingly, polyploidy cannot be invoked to explain the dramatic genome size variation among the taxa included in the study.

Directionality of genome size change

The phylogenetic framework provided by Figure 1 permits an heuristic evaluation of the directionality of genome size evolution in the tribe. To accomplish this we superimposed the DNA content estimates on the phylogeny of Figure 1 and used three methods for inferring ancestral states of continuously varying characters: (1) linear or Wagner parsimony (Swofford & Maddison, 1987), as implemented in PAUP*; (2) sum-of-squared-changes parsimony analysis (Maddison, 1991), also as implemented in PAUP*; and (3) a generalized least squared method (Martins & Hansen, 1997) implemented using the computer program COMPARE (Martins, 1999). These methods make no assumptions about the relative likelihood of genome size contraction or expansion, but are 'agnostic' (Bennetzen & Kellogg, 1997a) in this respect. Because there may be many equally parsimonious reconstructions under linear parsimony, the full range of values was used to explore the directionality of genome size change.

As shown in Figure 1, all analytical methods lead to the inference that genome size has fluctuated widely during the radiation of the Gossypieae, and that both genome size expansion and contraction have occurred repeatedly. Genome size is unambiguously inferred to have decreased at least four times, once each in the ancestors of *G. raimondii, C. tripartita*, the common ancestor of *T. lampas* and *T. thespesioides*, and the common ancestor of *G. kirkii* and *K. drynarioides*. Similarly, genome size is suggested to have increased in the lineages leading to *G. herbaceum*, *H. appendiculata*, and *T. populnea*. In some cases it is unclear when these genome size changes took place. For example, it is possible that the relatively large genome of *H. appendiculata* arose is its immediate ancestor (as shown in Figure 1), or earlier in the evolution of the genus (only one species was sampled), or even before the divergence of *Hampea* from *Thespesia* and *Lebronnecia*. This last scenario, suggested as a possibility only by an extreme value from linear parsimony analysis, would entail at least two genome size contractions, once each in the ancestor of *Lebronnecia* and *Thespesia*. Regardless of these uncertainties, all possible reconstructions lead to the conclusion that when summed across the phylogeny, the frequency of 2C DNA content decrease actually exceeds that of increase. This result holds irrespective of any reasonable adjustment of the phylogeny, which, as discussed above, is relatively robust.

Discussion

Bidirectional evolution of genome size

Recent years have witnessed an explosion in investigations of plant genome structure, culminating recently in the near-complete sequencing of the *Arabidopsis* genome (The *Arabidopsis* Genome Initiative, 2000). This effort has led to a greatly enhanced understanding of the important role that transposable elements, particularly retrotransposable elements, play in plant genome evolution (Wessler et al., 1995; Wessler, 1998; Bennetzen, 2000; Federoff, 2000). Especially significant has been the realization that retroelements may proliferate on a massive scale in plant genomes (Bennetzen, 1996; SanMiguel et al., 1996; Tikhonov et al., 1999; Bennett et al., 2000), in the process forming retroelement 'landing pads' that over evolutionary time lead to a hierarchical archeology of nested retroelements (SanMiguel et al., 1998). The notion that emerges from this literature is that plant genomes experience an inexorable increase in genome size over evolutionary time as a consequence of retroelement accumulation. Superimposed on this process is a second phenomenon acting in the same direction, namely, polyploidy. Numerous comparative genomic studies have increased our awareness of the prevalence of polyploidy in plants, as well as its episodic and cyclical nature (Soltis & Soltis, 1999; Otto & Whitton, 2000; Paterson et al., 2000; Wendel, 2000).

Recognition of the dual processes of retroelement accumulation and polyploidization lends *prima facie* credibility to the hypothesis that plant genomes expand over evolutionary time, particularly in the absence of counteracting mechanisms. Indeed, there *may* be a coarse-grained trend toward genome expansion on a global phylogenetic scale in flowering plants (Leitch et al., 1998) or perhaps within individual families, such as the grasses (Bennetzen & Kellogg, 1997a), or genera, such as the orchid genus *Paphiopedilum* (Cox et al., 1998). The phylogenetic evidence gathered to date, however, is circumstantial and subject to alternative interpretations. Nonetheless,

the hypothesis of unidirectional genome size evolution is inherently a phylogenetic statement, and hence phylogenetic analysis represents the first and essential step in critically testing the hypothesis for any particular plant group.

Here we have provided a well-supported phylogenetic framework for cotton and its allies, upon which we superimposed genome sizes to document their phylogenetic distribution. Polyploidy has been removed as a possible confounding factor because we have shown that all of the included taxa are diploid. Hence, the present distribution of genome sizes reflects either genome size increase or decrease, or perhaps both, since these genera and species diverged from their common ancestor. As discussed by Bennetzen and Kellogg (1997a), it is possible to invoke an 'increase-only' scenario, whereby genomes are prohibited from significantly shrinking. If one adopts this strong assumption, the ancestor of the entire tribe is inferred to have had a genome size equal to or smaller than the smallest one known among extant species (in the present case, 2C = 1.2 pg). A corollary of this assumption is that genome size expansion must have occurred in parallel in all lineages, with the possible exception of the one leading to *Gossypioides* and *Kokia*. Under this scenario, the remarkable variation in extant genome sizes would reflect variation only in the rate of genome size growth. Indeed, phylogenetic analysis by itself may be incapable of providing a rigorous refutation of this possibility, as the increase-only scenario may be invoked irrespective of any particular distribution of extant and inferred ancestral genome sizes. For the reasons described in the following paragraphs, however, we view it as more plausible that both expansion and contraction have occurred.

If one conceptually permits both genome expansion and contraction, thereby imposing a less stringent assumption on the origin of the present distribution of genome sizes, one may employ widely applied methods (Swofford & Maddison, 1987; Maddison, 1991; Bennetzen & Kellogg, 1997a; Martins & Hansen, 1997; Cunningham et al., 1998; Martins, 1999; Oakley & Cunningham, 2000) to infer ancestral genome sizes. Of course, one might challenge the quantitative details of the ancestral state reconstruction methodologies, which do not necessarily model biological processes in a realistic manner and which are subject to various biases and large confidence intervals (Cunningham et al., 1998; Oakley & Cunningham, 2000; Polly, 2001). For example, it seems unlikely that genome sizes have evolved in such a fashion that their sum of squared changes has been minimized, or that the total evolutionary change has been minimized, regardless of the estimation method or its precision; mechanisms of genome size change do not lend themselves to such simplistic and regular linearity. Nonetheless, and notwithstanding the inescapable inaccuracy in the details of the quantitative models, permitting the present distribution of genome sizes to reflect both expansion and contraction is less assumption-laden and less restrictive than the alternative, increase-only scenario.

If one accepts that the ancestral genome size reconstruction methodology is conceptually legitimate even if only quantitatively approximate, then one may use the phylogeny to assess the relative frequency of genome size increase and decrease. From Figure 1 it is evident that it is unlikely that there has been a unidirectional and parallel genome size expansion during the evolutionary history of the tribe. Instead, several genome size expansions and contractions are implicated. It seems likely, for example, that episodes of genome contraction have occurred in the ancestry of *Kokia* + *Gossypioides*, *Cienfuegosia*, *Lebronnecia*, *Gossypium*, and *T. lampas* + *T. thespesioides*. Equally probable are genome size expansions in *T. populnea* and *Hampea* and *Gossypium*. In *Gossypium* the largest genomes occur in a phylogenetically derived (Seelanan et al., 1999) group of species from tropical NW Australia; these species, with 2C contents approximating 7.0 pg (J.McD. Stewart, pers. comm.), would clearly be diagnosed as representing an additional example of genome size expansion. Similarly, sampling additional species in the other polytypic genera (*Hampea*, *Thespesia*, and *Cienfuegosia*) would likely reveal additional genome size variation and hence episodes of expansion and contraction.

Our results for the Gossypieae provide a phylogenetic suggestion of episodic genome size contraction, and thereby contribute to an increasing recognition that 'obese' plant genomes may occasionally go 'on a diet' (Rabinowicz, 2000). Much of the evidence for this dieting is circumstantial but collectively the data are compelling. First, our results for the Gossypieae are writ large on a phylogenetic scale: plants possessing small or large genomes are phylogenetically widespread throughout the angiosperms (Leitch et al., 1998; Stace, 2000). Genome size data exist for only about 1.4% of angiosperm species (Bennett et al., 1997; Bennett et al., 2000), yet the available database documents remarkable variation in genome size within numerous genera and families. *A. thaliana* is

often cited as having the smallest angiosperm genome, although *Cardamine amara* from the same family (Brassicaceae) has a genome size half again as small (2C = 0.11; Bennett & Smith, 1991). The distantly related horse-chestnut, *Aesculus hippocastanum* (Hippocastanaceae *sensu strictu*, or Sapindaceae *sensu lato*), has a genome as small as that of *Arabidopsis*, even though it clearly is paleopolyploid ($2n = 40$). In a recent phylogenetic analysis of angiosperm genome sizes (Leitch et al., 1998) small genomes were evident in all orders and larger genomes were widely distributed as well. The range of variation in C-values is remarkable within phylogenetically disparate lineages, including the Asterids (0.3–24.8 pg), Rosids (0.1–16.5 pg), Ranunculales (0.5–25.1 pg), and monocots (0.3–127.4 pg). These data may only be reconciled with unidirectional genome size evolution by the most stringent assumptions, including long-term retention of tiny genomes in phylogenetically disparate lineages as well as rampant genome size expansion in others. A more parsimonious interpretation, in our view, is that the foregoing data indicate that genome size is evolutionary labile and subject to both increase and decrease. It also seems likely, in our opinion, that patterns such as those reported here will be increasingly observed as more genera and families are subjected to joint phylogenetic analysis and genome size determinations.

A second source of evidence bearing on the likelihood of genome size contraction derives from comparative genomics, particularly comparative mapping studies (Paterson et al., 2000), which have informed us about the 'deep history' of angiosperm genomes. As noted earlier, polyploidy is an active, ongoing process in angiosperms (Masterson, 1994; Leitch & Bennett, 1997; Soltis & Soltis, 1999; Otto & Whitton, 2000; Wendel, 2000). Because genome doubling has been continuing since angiosperms first appeared definitively in the Cretaceous, most if not all angiosperm genomes have experienced one or more episodes of polyploidization at various times in the past. Thus, the modern angiosperm genome typically consists of a series of nested duplications of varying antiquity, only some of which descend to the present relatively unscathed by subsequent evolutionary disruptions. This paleopolyploidy often is revealed by detailed comparative mapping studies or other approaches (Reinisch et al., 1994; Moore et al., 1995; Lagercrantz & Lydiate, 1996; Shoemaker et al., 1996; Bennetzen & Freeling, 1997; Gaut & Doebley, 1997; Gómez et al., 1998; Kellogg, 1998; Lagercrantz, 1998; Muravenko et al., 1998; Sossey-Alaouni et al., 1998; Brubaker et al., 1999; Wilson et al., 1999; Paterson et al., 2000). Even *Arabidopsis* appears to have experienced one and possible several cycles of polyploidy during its evolutionary history (The *Arabidopsis* Genome Initiative, 2000; Grant et al., 2000; Vision et al., 2000), yet its genome usually is considered a model of streamlining. Clearly then, the streamlining process has entailed massive loss of redundant genomic sequence.

A third body of evidence pertinent to the possibility of genome size contraction stems from recent studies that increased our awareness of the significance of deletional mechanisms during genome evolution (Petrov, 2001). These mechanisms are many and varied, and sometimes the specific molecular events responsible are rather poorly understood. In *Drosophila*, for example, examination of 'dead-on- arrival' non-LTR retroelements revealed a spontaneous rate of DNA loss many times higher than that exhibited by mammalian pseudogenes (Petrov et al., 1996). Moreover, the efficiency of this mechanism may correlate with genome size; *Laupala* crickets, with a genome size 11-fold higher than *Drosophila*, lose DNA through this process 40 times more slowly (Petrov et al., 2000). Many of the deleted regions are associated with short direct repeats, suggesting loss through recombination or replication slippage (Petrov & Hartl, 1997). These studies were recently extended to *Podisma* grasshoppers, which have a genome that is 10 times larger than *Laupala*; rates of deletion in 'dead- on-arrival' nuclear pseudogenes are much lower in the former than the latter, relative to rates of point substitution (Bensasson et al., 2001). Rates of insertion or deletion in introns also may correlate with genome size, such that animals with larger genomes may possess larger introns (Hughes & Hughes, 1995; Ogata et al., 1996; Moriyama et al., 1998; Deutsch & Long, 1999; Vinogradov, 1999).

In plants, recent research involving experimentally induced double-stranded breaks showed that deletions were larger and insertions rarer in *A. thaliana* than in *Nicotiana tabacum*, which has a much larger genome (Kirik et al., 2000). Analysis of the recently published *Arabidopsis* genome sequence suggests that unequal crossing over may lead to deletion of genes or genomic segments (The *Arabidopsis* Genome Initiative, 2000). Recombinational mechanisms are also likely important in genome size contraction. This has elegantly been shown by the work of Schulman and collaborators using natural and experimental populations of barley (Vicient et al., 1999; Kalendar et al.,

2000; Shirasu et al., 2000). In one particular 66 kb region a minimum of 15 distinct retrotransposon insertion events were identified, yet most long terminal repeats (LTRs) exist in single copy (Shirasu et al., 2000). Presumably the responsible mechanism is homologous recombination between retroelement LTRs leading to internal deletions. An excess of solo LTRs has been observed throughout the genus *Hordeum* (Vicient et al., 1999), and remarkably, correlations between full-length elements and LTR copy number exist on local ecological scales in *H. spontaneum* from Israel (Kalendar et al., 2000). The latter example underscores the possibility of rapid change in genome size, mediated through homologous recombination. Perhaps larger deletions involving redundant chromosome segments or blocks of repetitive DNA could be deleted by similar mechanisms, as speculated by Bennetzen and Kellogg (1997a).

Concluding remarks

Much remains to be learned regarding deletional processes and their significance in genome evolution, but the foregoing attests to an increasing recognition that plant genome sizes are evolutionary labile and that deletional mechanisms may play a more prominent role than previously believed. This suggests, in conjunction with the phylogenetic and comparative mapping evidence discussed above, that genome size contraction is a real and perhaps common evolutionary phenomenon. Our results for the single Malvaceous tribe Gossypieae implicate one or more of these or other mechanisms of genome contraction, and hence contribute to the growing awareness that obese plant genomes frequently go 'on a diet' (Rabinowicz, 2000). Accordingly, present genome sizes may most appropriately be viewed as reflecting a dynamic balance between opposing forces of expansion and contraction, as recently outlined by Petrov (2001).

The foregoing discussion highlights both the evolutionary lability of genome size and that the process entails both expansion and contraction. This recognition may constitute a step toward understanding the evolutionary significance of genome size change, but perhaps this is only a rather small step. The pattern described is global in scope, that is, *total* DNA content per cell, and while genome size itself may have biological significance or be visible to natural selection (Bennett, 1985; Bennett, 1987; Price, 1988; Gregory & Hebert, 1999; Petrov, 2001), it may be that DNA content represents the net effect of myriad finer-scale changes, each with their own fitness consequences. Thus, an important avenue for future research is to simultaneously assess the directionality of change for multiple and different genomic components within a well-understood phylogenetic framework, using models like the one exemplified in the present study. It may be, for example, that overall genome size change primarily reflects copy-number increase or decrease for one or several families of repetitive elements, whose net effect quantitatively overwhelms other genomic components evolving in a countervailing direction, each with their own effects on fitness. Alternatively, one might envision cases where genome size itself is particularly responsive to selection, thereby imposing a unidirectional response for most genomic constituents. At present this is relatively unexplored terrain in the field of genome evolution. By combining the tools of phylogenetics and genomics, insights into these and related questions will undoubtedly soon be forthcoming.

Acknowledgements

We thank D. Petrov for helpful discussion and T. Haselkorn and George Hodnett for technical assistance. This work was supported by the National Science Foundation (USA) and the Texas Agric. Expt. Station.

References

The *Arabidopsis* Genome Initiative, 2000. Analysis of the genome sequence of the flowering plant, *Arabidopsis thaliana*. Nature 408: 796–815.

Bennett, M.D., 1985. Intraspecific variation in DNA amount and the nucleotypic dimension in plant genetics, pp. 283–302 in Plant Genetics, edited by M. Freeling. A. R. Liss, NY.

Bennett, M.D., 1987. Variation in genomic form and its ecological implications. New Phytol. 106: 177–200.

Bennett, M.D., P. Bhandol & I.J. Leitch, 2000. Nuclear DNA amounts in angiosperms and their modern uses – 807 new estimates. Ann. Bot. 86: 859–909.

Bennett, M.D., A.V. Cox & I.J. Leitch, 1997. Angiosperm DNA c-values database. http://www.rbgkew.org.uk/cval/database1.html

Bennett, M.D. & I.J. Leitch, 1995. Nuclear DNA amounts in angiosperms. Ann. Bot. 76: 113–176.

Bennett, M.D. & J.B. Smith, 1991. Nuclear DNA amounts in angiosperms. Phil. Trans. R. Soc. London B 334: 309–345.

Bennetzen, J.L., 1996. The contributions of retroelements to plant genome organization, function and evolution. Trends Microbiol. 4: 347–353.

Bennetzen, J.L., 2000. Transposable element contributions to plant gene and genome evolution. Plant Mol. Biol. 42: 251–269.

Bennetzen, J.L. & M. Freeling, 1997. The unified grass genome: synergy in synteny. Genome Res. 7: 301–306.

Bennetzen, J.L. & E.A. Kellogg, 1997a. Do plants have a one-way ticket to genomic obesity? Plant Cell 9: 1509–1514.

Bennetzen, J.L. & E.A. Kellogg, 1997b. Reply. Plant Cell 9: 1901–1902.

Bensasson, D., D.A. Petrov, D.-X. Zhang, D.L. Hartl & G.M. Hewitt, 2001. Genomic gigantism: DNA loss is slow in mountain grasshoppers. Mol. Biol. Evol. 18: 246–253.

Brubaker, C.L., A.H. Paterson & J.F. Wendel, 1999. Comparative genetic mapping of allotetraploid cotton and its diploid progenitors. Genome 42: 184–203.

Cox, A.V., G.J. Abdelnour, M.D. Bennett & I.J. Leitch, 1998. Genome size and karyotype evolution in the slipper orchids (Cypripedioideae: Orchidaceae). Amer. J. Bot. 85: 681–687.

Cronn, R., R.L. Small & J.F. Wendel, 1999. Duplicated genes evolve independently following polyploid formation in cotton. Proc. Natl. Acad. Sci. USA 96: 14406–14411.

Cunningham, C.W., K.E. Omland & T.H. Oakley, 1998. Reconstructing ancestral character states: a critical reappraisal. Trends Ecol. Evol. 13: 361–366.

Deutsch, M. & M. Long, 1999. Intron-exon structure of eukaryotic model organisms. Nucl. Acids Res. 27: 3219–3228.

Endrizzi, J.E., E.L. Turcotte & R.J. Kohel, 1985. Genetics, cytogenetics, and evolution of *Gossypium*. Adv. Genet. 23: 271–375.

Farris, J.S., 1996. Parsimony jackknifing outperforms neighbor-joining. Cladistics 12: 99–124.

Federoff, N., 2000. Transposons and genome evolution in plants. Proc. Natl. Acad. Sci. USA 97: 7002–7007.

Fryxell, P.A., 1979. The Natural History of the Cotton Tribe. Texas A&M University Press, College Station, TX.

Fryxell, P.A., 1992. A revised taxonomic interpretation of *Gossypium* L. (Malvaceae). Rheedea 2: 108–165.

Gaut, B.S. & J.F. Doebley, 1997. DNA sequence evidence for the segmental allotetraploid origin of maize. Proc. Natl. Acad. Sci. USA 94: 6808–6814.

Gómez, M.I., M.N. Islam-Faridi, M.S. Zwick, D.G. Czeschin, G.E. Hart, R.A. Wing, D.M. Stelly & H.J. Price, 1998. Tetraploid nature of *Sorghum bicolor* (L.) Moench. J. Hered. 89: 188–190.

Grant, D., P. Cregan & R.C. Shoemaker, 2000. Genome organization in dicots: genome duplication in *Arabidopsis* and synteny between soybean and *Arabidopsis*. Proc. Natl. Acad. Sci. USA 97: 4168–4173.

Gregory, T.R. & P.D.N. Hebert, 1999. The modulation of DNA content: proximate causes and ultimate consequences. Genome Res. 9: 317–324.

Hanson, R.E., M.N. Islam-Faridi, E.A. Percival, C.F. Crane, Y. Ji, T.D. McKnight, D.M. Stelly & H.J. Price, 1996. Distribution of 5S and 18S-28S rDNA loci in a tetraploid cotton (*Gossypium hirsutum* L.) and its putative diploid ancestors. Chromosoma 105: 55–61.

Hughes, A.L. & M.K. Hughes, 1995. Small genomes for better flyers. Nature 377: 391.

Jewell, D.C. & M.N. Islam-Faridi, 1994. Details of a technique for somatic chromosome preparation and C-banding of maize, pp. 484–493 in The Maize Handbook, edited by M. Freeling & V. Walbot. Springer, New York.

Johnson, J.D., D.G. Higgins & T.J. Gibson, 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucl. Acids Res. 22: 4673–4680.

Johnston, J.S., M.D. Bennett, A.L. Rayburn, D.W. Galbraith & H.J. Price, 1999. Reference standards for determination of DNA content of plant nuclei. Amer. J. Bot. 86: 609–613.

Kalendar, R., J. Tanskanen, S. Immonen, E. Nevo & A.H. Schulman, 2000. Genome evolution in wild barley (*Hordeum spontaneum*) by *BARE*-1 retrotransposon dynamics in response to sharp microclimatic divergence. Proc. Natl. Acad. Sci. USA 97: 6603–6607.

Kellogg, E.A., 1998. Relationships of cereal crops and other grasses. Proc. Natl. Acad. Sci. USA 95: 2005–2010.

Kirik, A., S. Salomon & H. Puchta, 2000. Species-specific double-strand break repair and genome evolution in plants. EMBO J. 2000: 5562–5566.

LaDuke, J.C. & J.F. Doebley, 1995. A chloroplast DNA based phylogeny of the Malvaceae. Syst. Bot. 20: 259–271.

Lagercrantz, U., 1998. Comparative mapping between *Arabidopsis thaliana* and *Brassica nigra* indicates that *Brassica* genomes have evolved through extensive genome replication accompanied by chromosome fusions and frequent recombinations. Genetics 150: 1217–1228.

Lagercrantz, U. & D.J. Lydiate, 1996. Comparative genome mapping in *Brassica*. Genetics 144: 1903–1910.

Leitch, I.J. & M.D. Bennett, 1997. Polyploidy in angiosperms. Trends Plant Sci. 2: 470–476.

Leitch, I.J., M.W. Chase & M.D. Bennett, 1998. Phylogenetic analysis of DNA C-values provides evidence for a small ancestral genome size in flowering plants. Ann. Bot. (suppl. A) 82: 85–94.

Maddison, W.P., 1991. Squared-change parsimony reconstructions of ancestral states for continuous-valued characters on a phylogenetic tree. Syst. Zool. 40: 304–314.

Martins, E.P., 1999. COMPARE, Version 4.3. Computer Programs for the Statistical Analysis of Comparative Data. Distributed by the author via the WWW at http://darkwing.uoregon.edu/~compare4/. Department of Biology, University of Oregon, Eugene, Oregon.

Martins, E.P. & T.F. Hansen, 1997. Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. Amer. Nat. 149: 646–667.

Masterson, J., 1994. Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. Science 264: 421–424.

Miege, J. & N. Josserand, 1972. Nombres chromosomiques d'especes africaines et malgaches. Candollea 27: 283–292.

Moore, G., K.M. Devos & Z. Wang, 1995. Grasses, line up and form a circle. Curr. Biol. 5: 737–739.

Moriyama, E.N., D.A. Petrov & D.L. Hartl, 1998. Genome size and intron size in *Drosophila*. Mol. Biol. Evol. 15: 770–773.

Muravenko, O., A.R. Fedotov, E.O. Punina, L.I. Federova, V.G. Grif & A.V. Zelenin, 1998. Comparison of chromosome BrdU-Hoechst-Giemsa banding patterns of the A_1 and $(AD)_2$ genomes of cotton. Genome 41: 616–625.

Oakley, T.H. & C.W. Cunningham, 2000. Independent contrasts succeed where ancestor reconstruction fails in a known bacteriophage phylogeny. Evolution 54: 397–405.

Ogata, H., W. Fujibuchi & M. Kanehisa, 1996. The size differences among mammalian introns are due to the accumulation of small deletions. FEBS Lett. 390: 99–103.

Otto, S.P. & J. Whitton, 2000. Polyploid incidence and evolution. Annu. Rev. Genet. 34: 401–437.

Paterson, A.H., J.E. Bowers, M.D. Burow, X. Draye, C.G. Elsik, C.-X. Jiang, C.S. Katsar, Y.-R. Lan, R. Ming & R.J. Wright, 2000. Comparative genomics of plant chromosomes. Plant Cell 12: 1523–1539.

Petrov, D.A., 1997. Slow but steady: reduction to genome size through biased mutation. Plant Cell 9: 1900–1901.

Petrov, D.A., 2001. Evolution of genome size: new approaches to an old problem. Trends Genet. 17: 23–28.

Petrov, D.A. & D.L. Hartl, 1997. Trash DNA is what gets thrown away: high rate of DNA loss in *Drosophila*. Gene 205: 279–289.

Petrov, D.A., E.R. Lozovskaya & D.L. Hartl, 1996. High intrinsic rate of DNA loss in *Drosophila*. Nature 384: 346–349.

Petrov, D.A., T.A. Sangster, J.S. Johnston, D.L. Hartl & K.L. Shaw, 2000. Evidence for DNA loss as a determinant of genome size. Science 287: 1060–1062.

Polly, P.D., 2001. Paleontology and the comparative method: ancestral node reconstructions versus observed node values. Amer. Nat. 157: 596–609.

Price, H.J., 1988. Nuclear DNA content variation within angiosperm species. Evol. Trends Plants 2: 53–60.

Rabinowicz, P.D., 2000. Are obese plant genomes on a diet? Genome Res. 10: 893–894.

Reinisch, A.J., J. Dong, C.L. Brubaker, D.M. Stelly, J.F. Wendel & A.H. Paterson, 1994. A detailed RFLP map of cotton, *Gossypium hirsutum* $\times$ *G. barbadense*: chromosome organization and evolution in a disomic polyploid genome. Genetics 138: 829–847.

SanMiguel, P., B.S. Gaut, A. Tikhonov, Y. Nakajima & J.L. Bennetzen, 1998. The paleontology of intergene retrotransposons of maize: dating the strata. Nat. Gen. 20: 43–45.

SanMiguel, P., A. Tikhonov, Y.K. Jin, N. Motchoulskaia, D. Zakharov, A. Melake-Berhan, P.S. Springer, K.J. Edwards, M. Lee, Z. Avramova & J.L. Bennetzen, 1996. Nested retrotransposons in the intergenic regions of the maize genome. Science 274: 765–768.

Seelanan, T., C.L. Brubaker, J.M. Stewart, L.A. Craven & J.F. Wendel, 1999. Molecular systematics of Australian *Gossypium* section *Grandicalyx* (Malvaceae). Syst. Bot. 24: 183–208.

Seelanan, T., A. Schnabel & J.F. Wendel, 1997. Congruence and consensus in the cotton tribe. Syst. Bot. 22: 259–290.

Shirasu, K., A.H. Schulman, T. Lahaye & P. Schulze-Lefert, 2000. A contiguous 66-kb barley DNA sequence provides evidence for reversible genome expansion. Genome Res. 10: 908–915.

Shoemaker, R.C., K. Polzin, J. Labate, J. Specht, E.C. Brummer, T. Olson, N. Young, V. Concibido, J. Wilcox, J.P. Tamulonis, G. Kochert & H.R. Boerma, 1996. Genome duplication in soybean (*Glycine* subgenus *soja*). Genetics 144: 329–338.

Soltis, D.E. & P.S. Soltis, 1999. Polyploidy: origins of species and genome evolution. Trends Ecol. Evol. 9: 348–352.

Sossey-Alaouni, K., H. Serieys, M. Tersac, P. Lambert, E. Schilling, Y. Griveau, F. Kaan & A. Bervillé, 1998. Evidence for several genomes in *Helianthus*. Theor. Appl. Genet. 97: 422–430.

Stace, C.A., 2000. Cytology and cytogenetics as a fundamental taxonomic resource for the 20th and 21st century. Taxon 49: 451–477.

Swofford, D.L., 2001. PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods), Version 4. Sinauer Associates, Sunderland, MA.

Swofford, D.L. & W.P. Maddison, 1987. Reconstructing ancestral character states under Wagner parsimony. Math. Biosci. 87: 199–229.

Thomas, C.A., 1971. The genetic organisation of chromosomes. Annu. Rev. Genet. 5: 237–256.

Tikhonov, A., P. SanMiguel, Y. Nakajima, N.M. Gorenstein, J.L. Bennetzen & Z. Avramova, 1999. Colinearity and its exceptions in orthologous *adh* regions of maize and sorghum. Proc. Natl. Acad. Sci. USA 96: 7409–7414.

Vicient, C.M., A. Suoniemi, K. Anamthawat-Jonsson, J. Tanskanen, A. Beharav, N.E. & A.H. Schulman, 1999. Retrotransposon *BARE*-1 and its role in genome evolution in the genus *Hordeum*. Plant Cell 11: 1769–1784.

Vinogradov, A.E., 1999. Intron-genome size relationship on a large evolutionary scale. J. Mol. Evol. 49: 376–384.

Vision, T.J., D.G. Brown & S.D. Tanksley, 2000. The origins of genomic duplications in *Arabidopsis*. Science 290: 2114–2117.

Wendel, J.F., 2000. Genome evolution in polyploids. Plant Mol. Biol. 42: 225–249.

Wendel, J.F. & V.A. Albert, 1992. Phylogenetics of the cotton genus (*Gossypium* L.): character-state weighted parsimony analysis of chloroplast DNA restriction site data and its systematic and biogeographic implications. Syst. Bot. 17: 115–143.

Wendel, J.F., R.L. Small, R.C. Cronn & C.L. Brubaker, 1999. Genes, jeans, and genomes: reconstructing the history of cotton. Plant evolution in man-made habitats, pp. 133–161 in Proc. VIIth Int. Symp. Int. Org. Plant Biosystemat. Amsterdam.

Wessler, S.R., 1998. Transposable elements associated with normal plant genes. Physiol. Plantarum 103: 581–586.

Wessler, S.R., T.E. Bureau & S.E. White, 1995. LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. Curr. Opin. Genet. Dev. 5: 814–821.

Wilson, W.A., S.E. Harrington, W.L. Woodman, M. Lee, M.E. Sorrells & S.R. McCouch, 1999. Inferences on the genome structure of progenitor maize through comparative analysis of rice, maize and the domesticated Panicoids. Genetics 153: 453–473.