# The Dynamic Ups and Downs of Genome Size Evolution in *Brassicaceae*

*Martin A. Lysak,*[1] *Marcus A. Koch,† Jeremy M. Beaulieu,‡ Armin Meister,§ and Ilia J. Leitch‖*

*Department of Functional Genomics and Proteomics, Institute of Experimental Biology, Masaryk University, Brno, Czech Republic; †Heidelberg Institute of Plant Sciences, Biodiversity and Plant Systematics, University of Heidelberg, Heidelberg, Germany; ‡Department of Ecology and Evolutionary Biology, Yale University; §Leibniz-Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany; and ‖Jodrell Laboratory, Royal Botanic Gardens, Kew, United Kingdom

Crucifers (*Brassicaceae*, *Cruciferae*) are a large family comprising some 338 genera and c. 3,700 species. The family includes important crops as well as several model species in various fields of plant research. This paper reports new genome size (GS) data for more than 100 cruciferous species in addition to previously published C-values (the DNA amount in the unreplicated gametic nuclei) to give a data set comprising 185 *Brassicaceae* taxa, including all but 1 of the 25 tribes currently recognized. Evolution of GS was analyzed within a phylogenetic framework based on gene trees built from five data sets (*mat*K, *chs*, *adh*, *trn*LF, and ITS). Despite the 16.2-fold variation across the family, most *Brassicaceae* species are characterized by very small genomes with a mean 1C-value of 0.63 pg. The ancestral genome size (ancGS) for *Brassicaceae* was reconstructed as $^{anc}1C = 0.50$ pg. Approximately 50% of crucifer taxa analyzed showed a decrease in GS compared with the ancGS. The remaining species showed an increase in GS although this was generally moderate, with significant increases in C-value found only in the tribes Anchonieae and Physarieae. Using statistical approaches to analyze GS, evolutionary gains or losses in GS were seen to have accumulated disproportionately faster within longer branches. However, we also found that GS has not changed substantially through time and most likely evolves passively (i.e., a tempo that cannot be distinguished between neutral evolution and weak forms of selection). The data reveal an apparent paradox between the narrow range of small GSs over long evolutionary time periods despite evidence of dynamic genomic processes that have the potential to lead to genome obesity (e.g., transposable element amplification and polyploidy). To resolve this, it is suggested that mechanisms to suppress amplification and to eliminate amplified DNA must be active in *Brassicaceae* although their control and mode of operation are still poorly understood.

## Introduction

The mustard family (*Brassicaceae*, *Cruciferae*) comprises about 338 genera and more than 3,700 species (Al-Shehbaz et al. 2006). The flowers usually have four petals resembling a cross, four sepals, and a tetradynamous androecium (four long inner and two short outer stamens, respectively) making them easily distinguishable. *Brassicaceae* holds an important position among angiosperm families for two reasons. First, although not as economically important as cereal species in *Poaceae*, the family includes a wide variety of important vegetable, forage, and oil seed crops such as cabbage, cauliflower, broccoli, kohl rabi (all cultivars of *Brassica oleracea*), rapeseed (*Brassica napus*), as well as several medicinal and ornamental species. The second, equally significant reason lies in the unparalleled role *Arabidopsis thaliana* has played as a model system in plant genetic and genomic research. *Arabidopsis* was proposed as a model species by Friedrich Laibach as early as 1943 (Laibach 1943). Among other favored parameters of *Arabidopsis*, he listed a short generation time of only two months. Later, Bennett (1972) showed that the short life cycle of *Arabidopsis* was correlated with its extremely small genome size (GS) (1C = 157 Mb, Bennett et al. 2003). Indeed, its small GS was one of the reasons why *Arabidopsis* was selected as the first plant to have its genome completely sequenced (Meyerowitz 1999; Federspiel 2000; The Arabidopsis Genome Initiative 2000).

The availability of annotated sequence data for *Arabidopsis* through internet-based databases has played a crucial role in comparative genomics and proteomics of *Brassicaceae*. Recently, genomic research has begun to move out from an *Arabidopsis*-centered perspective toward family-wide comparative studies (e.g., Clauss and Koch 2006; Lysak and Lexer 2006; Schranz et al. 2006; Bomblies and Weigel 2007; Schranz et al. 2007), and for this, knowledge of GS variation and evolution within *Brassicaceae* has been invaluable. For example, GS data are essential 1) for several ongoing and planned sequencing projects of crucifer species such as *Arabidopsis lyrata*, *Brassica rapa*, *Capsella rubella*, and *Thellungiella halophila* (e.g., Yang et al. 2005; Lysak and Lexer 2006; Bomblies and Weigel 2007; Schranz et al. 2007) to know how much DNA to sequence for complete coverage of the genome, 2) for the construction of large-insert (bacterial artificial chromosome [BAC]) libraries to know how many BACs will be needed, and 3) for the accurate quantification of repetitive elements within a genome. However, currently available GS data for *Brassicaceae* species (i.e., from the Plant DNA C-values database; Bennett and Leitch 2005; and http://www.kew.org/genomesize/homepage.html) and other published data not yet incorporated in the database (i.e., GS data for 72 taxa, in 14 tribes, ~1.9% of all crucifers) are biased toward several close relatives of *Arabidopsis* (tribe Camelineae) and economically important species from tribe Brassiceae such as *Brassica*, *Raphanus*, *Sinapis*, and others. This contrasts with current systematic and phylogenetic treatments of *Brassicaceae* that recognize at least 25 tribes (Al-Shehbaz et al. 2006).

Considering the importance of *Arabidopsis* and its relatives as prominent model species and their extremely small genomes (mean 1C = 0.93 pg, Bennett and Leitch 2005), we need to ask if these GS data are the exception or the rule

[1] The present paper is dedicated to the memory of our colleague and friend Dr Leoš Klimeš, a Czech botanist and ecologist who made a significant contribution to the knowledge of the endemic crucifer flora of the Himalayan region. He provided plant material for several studies on evolution and systematics of *Brassicaceae* including the present study. Leoš was reported missing in August 2007 while carrying out field work in Ladakh, India. He has not been seen since.

Key words: *Brassicaceae*, *Cruciferae*, *Arabidopsis*, genome size evolution, chromosomes, phylogenetics, polyploidy.

E-mail: lysak@sci.muni.cz.

*Mol. Biol. Evol.* 26(1):85–98. 2009
doi:10.1093/molbev/msn223
Advance Access publication October 8, 2008

among c. 3,700 crucifer species and what is the overall extent of GS variation and patterns of evolution across *Brassicaceae*. In addition, given the extensive genomic resources available for an increasing number of *Brassicaceae* species, together with increasing knowledge on the diversity of *Brassicaceae* repetitive DNA elements (e.g., Devos et al. 2002; Berr et al. 2006; Hall et al. 2006; Lim et al. 2007), epigenetic regulation (e.g., Gendrel and Colot 2005; Josefsson et al. 2006; Nasrallah et al. 2007), chromosomal and karyotype evolution (Lysak et al. 2005; Lysak and Lexer 2006; Schranz et al. 2006), and comparative genetic mapping (Koch and Kiefer 2005; Parkin et al. 2005; Schranz et al. 2006), there is the potential to view dynamics and evolution of GS diversity in the family from numerous, diverse perspectives. Such an opportunity is not available for many other angiosperm families where genomic data are much scarcer.

The first attempts to gain an overview of GS evolution across *Brassicaceae* within an internal transcribed spacer (ITS)-based phylogenetic framework were made by Johnston et al. (2005). They noted that the GS of 34 analyzed species varied 8-fold (0.16–1.31 pg/1C) and estimated that the ancestral genome size (ancGS) for the family was approximately 0.2 pg/1$C_x$ (1$C_x$ = monoploid GS corresponding to the DNA amount in the basic haploid chromosome complement and is calculated by dividing the 2C-value by ploidy level, cf. Greilhuber et al. 2005) with both increases and decreases detected in different phylogenetic branches. However, this study analyzed only 0.9% of crucifer species from just 8 of the 25 tribes now recognized, and the phylogeny had been inferred using ITS markers which have been shown, in many cases, to be influenced by concerted evolution (Koch et al. 2003). Furthermore, the authors argued that the mode of GS evolution was not known and so inferred ancGS values simply as 1C-values common to members at each branch of the clade (Johnston et al. 2005). Similar results were obtained by Oyama et al. (2008) focusing on microevolutionary patterns of closer relatives of *Arabidopsis* and demonstrating a 4.4-fold variation in GS variation among 26 species analyzed. However, the taxon sampling was biased in a similar way as described above. Here, the underlying phylogenetic hypothesis was generated by utilizing nuclear encoded markers such as *chs* (chalcone synthase) and ITS and two plastid markers, namely the *mat*K and the *trn*L intron. The most important conclusion drawn was that GS is free to increase or decrease without a directional bias.

To address these concerns and extend our knowledge of GS evolution in *Brassicaceae*, this paper reports new GS data for more than 100 cruciferous species selected to extend the phylogenetic coverage of the family to include all but 1 of the 25 tribes of *Brassicaceae* following the most comprehensive and accurate phylogenetic hypotheses available (Bailey et al. 2006; Beilstein et al. 2006; Koch et al. 2007). These data were then combined with previously published C-value estimates to give a data set comprising 185 taxa representing the most comprehensive list of GS data currently available for *Brassicaceae*. The data were analyzed using statistical approaches to 1) track the distribution and extent of GS variation across *Brassicaceae*, 2) reconstruct the ancGS of *Brassicaceae*, and 3) examine the mode and tempo of GS evolution within an increasingly robust phylogenetic framework.

## Materials and Methods

### Species Analyzed

A full list of the species analyzed for GS is given in supplementary table S1 (Supplementary Material online). Information of DNA data has been provided earlier (Koch et al. 2007), and more information is given in supplementary table S2 (Supplementary Material online) which gives an overview of ITS1–ITS2 sequences used together with the ITS alignment used for phylogenetic reconstruction (supplementary tables S2 and S3, Supplementary Material online).

### GS Estimation by Flow Cytometry

Small amounts of leaf tissue of a sample and standard were cochopped in 1 ml of Galbraith buffer (Galbraith et al. 1983) supplemented with 50 μg/ml propidium iodide (Molecular Probes, Eugene, OR) and DNase-free RNase (Boehringer, Heidelberg, Germany) and filtered and analyzed by a FACStar$^{Plus}$ (Becton Dickinson, Franklin Lakes, NJ) flow cytometer (514 nm/500 mW excitation, 630 nm emission). Data evaluation was carried out using the CellQuest analysis program.

*Arabidopsis thaliana* (Columbia) with a 1C-value of 0.16 pg (Bennett et al. 2003) was used as a primary standard. Soybean (*Glycine max* "Cina 5202"; 1.13 pg/1C) and radish (*Raphanus sativus* "Voran"; 0.53 pg/1C) obtained from the Genebank of Leibniz-Institute for Plant Genetics and Crop Plant Research (IPK), Gatersleben, were used as secondary standards. C-values of *G. max* and *R. sativus* were calculated from the pairwise measurements with *Arabidopsis* as an internal standard.

### Chromosome Counts

Mitotic chromosome counts were recorded from floral or root tip tissues fixed in 3:1 ethanol:acetic acid and digested in a 0.3% mixture of pectolyase, cytohelicase, and cellulase (all Sigma, St Louis, MO) in citric buffer (pH 4.8). Anthers or entire flower buds were dissected and spread on a microscopic slide as described by Lysak et al. (2005). Root tips were pretreated in 4 $\mu$M amiprophos-methyl (Duchefa, Haarlem, the Netherlands), digested, and squashed in a drop of 50% acetic acid under a coverslip and frozen on dry ice. After removing the coverslip, the slide was dehydrated and air dried. All slides were stained by 4′,6-diamidino-2-phenylindole (DAPI) (1 or 2 μg/ml) in Vectashield (Vector Laboratories, Burlingame, CA), and chromosome images were captured with a Zeiss Axioplan 2 epifluorescence microscope equipped with a Spot 2e CCD camera. The images were processed for contrast and sharpness in Adobe Photoshop 6.0 software.

### Data Used for Statistical Analysis of GS Evolution in *Brassicaceae*

Prior to the statistical analysis of GS evolution careful consideration was given as to whether 1C-or 1$C_x$-values should be analyzed (i.e., whether or not to take into account GS differences due to ploidy). Chromosome numbers vary extensively in our species set (2n = 8–56), indicative of the

polyploid origin of many taxa analyzed. Although $C_x$-values could theoretically be calculated by dividing absolute 2C-values by (often only estimated) ploidy level, this approach is problematic given the paleopolyploid history of many crucifer taxa including bona fide diploids such as *A. thaliana* (The Arabidopsis Genome Initiative 2000). Recently, it was shown that species of the tribe Brassiceae underwent two or three ancient whole-genome duplications (The Arabidopsis Genome Initiative 2000; De Bodt et al. 2005) shared with the lineage leading to *Arabidopsis*, followed by a more recent whole-genome triplication event associated with extensive chromosome number variation (Lysak et al. 2005; Parkin et al. 2005). This type of genome evolution has also been directly or indirectly observed in other crucifer groups (Lysak and Lexer 2006; T. Mandáková and M.A. Lysak, unpublished data). For example, large C-values found in some *Physaria* species (tribe Physarieae) with low chromosome numbers (2n = 8, e.g., *Physaria bellii* 2C = 4.7 pg; Lysak and Lexer 2006) suggest that these taxa underwent polyploidization followed by subsequent reduction of chromosome numbers. Given these ongoing paleopolyploidy–diploidization cycles, inferring ploidy levels and hence $C_x$-values can be very subjective. To avoid complications arising from this issue, GS data for statistical analysis were input as C-values for all species (unless a species had more than one cytotype at different ploidy levels in which case the lowest C-value was used). It is realized that this approach may lead to an overestimation of GS parameters in the analysis, but it is considered the most objective approach available.

### Phylogenetic Relationships in *Brassicaceae* Used for Analysis of GS Evolution

Phylogenetic hypotheses within the *Brassicaceae* family are still lacking significant resolution of deeper nodes and splits of lineages (Bailey et al. 2006). As yet, it is still unclear if this is simply because of limited number of mutations resolving deeper nodes unambiguously (which might be easily overcome by increased data sets) or if evolution of the whole family is characterized by consecutive radiation events (which cannot be resolved in simple dichotomous phylograms). A recent study (Koch et al. 2007) favored the latter scenario. Nevertheless, the most recently presented molecular studies (Bailey et al. 2006; Beilstein et al. 2006; Koch et al. 2007) support the taxonomic treatment of 25 crucifer tribes given by Al-Shehbaz et al. (2006). However, future phylogenetic–systematic research will introduce some additional tribes in order to avoid paraphyletic groups (Koch and Al-Shehbaz 2008). To resolve evolutionary relationships among the newly defined tribes, new molecular markers (enlarged sets of single-copy genes) and/or more advanced tools to analyze phylogenetic data are needed.

Recently, we successfully applied the method of Huson et al. (2004) which allows the reconstruction of phylogenetic relationships based on analyzing gene trees with overlapping but not necessarily identical taxon sets (Koch et al. 2007). The final result is a "SuperNetwork." For the present work, the SuperNetwork (supplementary fig. S1, Supplementary Material online), using gene trees built from five data sets (*mat*K, *chs*, *adh*, *trn*LF, and ITS; supplementary table S3, Supplementary Material online), was used. This is based largely on the data used by Koch et al. (2007) except for *trn*LF data (Lysak et al. 2005) and ITS data which was increased by the addition of 48 new sequences to include more sister taxa and to add those species for which GS estimates were available (supplementary table S2, Supplementary Material online). In a first study (Koch et al. 2001) comparing phylogenetic hypothesis derived from single-copy genes from the chloroplast and the nuclear genome, namely *mat*K and *chs*, some incongruencies were shown. Subsequently, minor incongruencies between phylogenies derived from genes form the various genomes have been detected later on (Lysak et al. 2005; Koch et al. 2007; Franzke et al. 2008). Furthermore, ITS-derived phylogenies might be biased by the special mode of concerted evolution also creating either unresolved or contradicting phylogenetic hypothesis. At this stage, we do not know enough about individual marker gene evolution (e.g., *trn*F pseudogene evolution in the *trn*LF region: Dobeš et al. 2007; Schmickl et al. 2008; concerted evolution with recombination of the ITS region: Koch et al. 2003). Consequently, we followed a strategy to calculate a SuperNetwork with all markers simultaneously and collapse incongruent splits subsequently. This approach may have benefits if the five genes have different histories due to recombination or partial reintegration into the genome.

For this study, five strict consensus trees (*adh*, *chs*, *mat*K, *trn*LF, and ITS; input file: supplementary table S3, Supplementary Material online) were used to calculate a SuperNetwork using the Z-closure option in Splitstree version 4beta26 with the following assumptions: splitstransfom = EqualAngle; SplisPostProcess filter = dimension value = 4 (Huson and Bryant 2006; for further details, see Koch et al. 2007). The resulting SuperNetwork combines weighted splits from the single trees. Branch lengths were weighted using information from partial splits in the source trees. However, because bracket notations have been provided without branch length, each tree contributed equally to each branch length. To minimize conflicting phylogenetic signals, we used the option COLLAPSE SPLITS. This final collapsed SuperNetwork tree has been translated into bracket notation, which served as input format for "pruning" using the program Phyutility (Smith and Dunne 2008) to match the taxa in our GS data set. The resulting tree contained polytomies. To deal with this issue, we simultaneously resolved all polytomies at random 100 times in Mesquite (Maddison WP and Maddison DR 2007) to obtain a distribution of all possible resolutions of these phylogenetic uncertainties. We then used these 100 randomly resolved trees as topological constraints to calculate maximum likelihood estimates of branch length using the ITS sequence data in PAUP* (Swofford 2000). Maximum likelihood estimates were calculated assuming a general time reversible + $\Gamma$ model of molecular evolution.

### Statistical Analysis of GS Evolution

#### *A Priori Transformation of GS Data*

We applied a $\log_{10}$ transformation prior to all analyses to ensure "proportional" GS evolution. Under a Brownian

motion model of evolution (as we utilize here), a given trait should have an equal probability of either increasing or decreasing in the same magnitude given its current state. However, this assumption is violated when traits, such as GS, are constrained to be nonzero. For example, given a 0.25 pg genome, an increase or decrease of 0.50 pg is not likely to occur in equal probability. Rather, in this case, change would be better expressed as a proportion, where the probability of an increase or decrease of say, 20%, is likely to occur regardless of the initial GS at speciation. Thus, it is generally acknowledged that GS evolution may be better represented as proportional change through an a priori $\log_{10}$ transformation (O'Meara et al. 2006; Oliver et al. 2007)

### *Bayesian Inference of ancGS and Its Mode and Tempo of Evolution*

There was uncertainty in the topology of our underlying phylogenetic hypothesis. To utilize this uncertainty, we used a generalized least squares (GLS) model in a Bayesian framework to test for further deviations from a Brownian motion model of evolution (apart from proportional effects), to calculate an estimate of the ancGS in *Brassicaceae*, and to characterize additional parameters involving branch length transformation ($\kappa$, $\delta$, and $\lambda$, as described below) to describe the mode and tempo of trait evolution. This was carried out using the BayesTraits implementation of Markov chain Monte Carlo (MCMC) methods. The GLS method incorporates nonindependence of species values by scaling regression parameters according to a variance–covariance matrix that describes the shared path length among all species (described by the phylogeny). For single trait analyses, such as GS, each trait observation is treated as a prediction of the trait value regressed against its total path length from root to species (Pagel 1997). The slope term is analogous to the expected variance of evolutionary change or the Brownian motion rate parameter. The *y* intercept ($\alpha$) of this regression is an estimate of the root of the tree (or ancestral value). The GLS method can also incorporate an additional parameter to account for directional trends in trait evolution. The detection of a directional trend can result in ancestral estimates that are outside the range of extant taxa trait values.

A Markov chain with uniform priors was allowed to run through 21,000,000 iterations with a burn-in of 1,000,000. This large run ensured that the Markov chain converged on the posterior distribution given our large GS data set. The rate deviation was kept to the default of 2 to force a Markov chain with an acceptance rate range between 20% and 40%. The rate deviation parameter allows for proper mixing when the chain reaches the stationary distribution. High acceptance rates indicate that the Markov chain has accepted nearly all proposed parameter values resulting in a high autocorrelation between successive iterations of the chain. Sampling every 500th iteration of the chain also circumvented the problem of autocorrelation. Thus, the posterior distributions from our Markov chain resulted in a sample size of 40,000. For hypothesis testing, a Bayes factor was calculated to compare the harmonic mean between three models of evolution. We first compared a "random-walk" model and a random-walk model with a directional parameter, where both simultaneously estimated the branch length scaling parameters. The best model from this comparison was then used to compare against a null model of pure Brownian motion where all scaling parameters were set to 1. The Bayes factor test statistic is calculated as 2(log [harmonic mean (improved model)]−log[harmonic mean (null model)]. A Bayes factor $>5$ indicates strong evidence and $>10$ indicates very strong evidence.

The branch scaling parameters $\kappa$, $\delta$, and $\lambda$ can improve the fit of the data to the model. The parameter $\kappa$ detects trait evolutionary rate heterogeneity by differentially scaling individual branch lengths. A value of $\kappa > 1$ suggests that long branches contribute disproportionately to trait evolution than shorter ones, indicative of stasis in shorter branches, whereas $\kappa < 1$ suggests that shorter branches contribute disproportionately to trait evolution, indicating stasis on longer branches. The parameter $\delta$ scales the overall path lengths (shared path lengths between related species and the total distance from root to species) between species to test shifts in rates of evolution through time. A value of $\delta > 1$ suggests that longer path lengths (longer distance from root to species) have contributed disproportionately to trait evolution and are associated with more recent radiations. A value of $\delta < 1$ suggests that shorter paths have contributed disproportionately to trait evolution, which is associated with rapid evolution early on with subsequent stasis (i.e., adaptive radiation). The parameter $\lambda$ describes whether the phylogenetic tree correctly predicts the pattern of covariance among species. A $\lambda$ near 1 indicates strong phylogenetic signal, whereas a $\lambda$ near 0 indicates that a star phylogeny (i.e., all species diverged from a single ancestor) better describes the relationship among trait values. For both $\kappa$ and $\delta$, we tested whether the posterior distribution was significantly different from 1. The *P* values for these tests were calculated by determining the proportion of the posterior distribution of the scaling parameters that crossed 1. For heuristic purposes, we present the mean and the 95% highest posterior density (HPD) for each parameter. The proper interpretation of the 95% HPD is that it is a credible set that contains 95% of the sampled values. Summary statistics from the MCMC runs were analyzed in Tracer (V. 1.4).

### *Phylogenetically Corrected t-Tests*

Using the approaches outlined above eight clades were identified (Fig. 1*A*) and differences in the range and mean genome sizes for each are illustrated in Fig. 1*B*. However, without considering phylogenetic history, it is not possible to know whether mean trait differences could have evolved by chance or through selective processes. We performed a series of phylogenetically corrected *t*-tests (Organ et al. 2007) to test for significant nonrandom differences in GS between clades. This was carried out using GLS methods in BayesTraits in MCMC mode (Organ et al. 2007). Different clades of interest were pruned from the 100 phylogenies using the multiple tree pruning algorithm in Phyutility (Smith and Dunne 2008). Each clade was given a binary code of either 1 or 0. For example, testing for differences in clades 1 and 2, clade 1 was arbitrarily coded as a 1 and clade 2 coded as 0. For differences in clade 3 and clades 1 + 2, clade 3 was coded with a 1 and clades 1 + 2 were coded

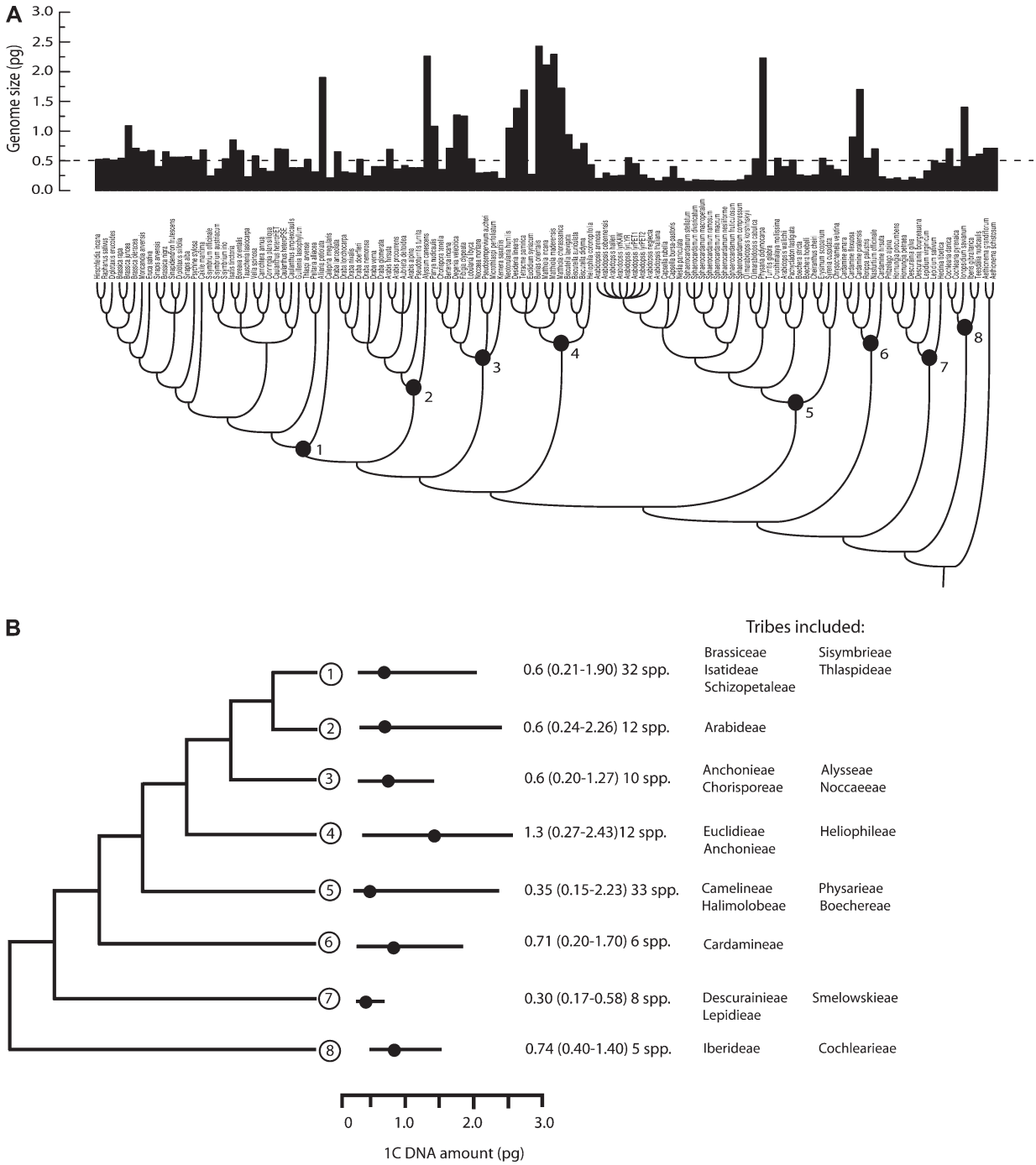


FIG. 1.—(*A*) The distribution of 1C-values for 120 taxa superimposed on our supertree phylogeny. The dashed line is the estimated ancGS for *Brassicaceae* ($^{anc}1C = 0.504$ pg) based on the mean of the posterior distribution of the ancGS using MCMC methods (see text). Numbers in the tree denote the crown of the eight monophyletic clades referred to in the text. (*B*) Range of 1C-values in the eight groups identified in (*A*) shown as a line connecting the minimum and maximum C value, with the mean shown as a dot. The mean followed by the minimum and maximum C-values in picograms is given on the right together with the number of taxa and the tribes analyzed in each group.

as 0 and so on. A phylogenetically corrected regression between GS and the binary grouping variable was determined by weighting the regression coefficient by the variance–covariance matrix describing the shared path length among all species. The posterior distribution was obtained for regression coefficients from an MCMC using the default settings, except we sampled every 500th iteration and the rate deviation was set at the appropriate value to ensure an acceptance rate of 20%. *P* values were calculated by determining the proportion of the posterior distribution of the regression coefficients that crossed zero. A nonsignificant phylogenetic *t*-test would indicate that given the limited sample the observed differences and variability between clades could have arisen by chance (Organ et al. 2007).

**Table 1**
**Minimum (Min.), Maximum (Max.), and Mean 1C-values for 170 *Brassicaceae* Species Grouped into Tribes Recognized by Al-Shehbaz et al. (2006). The percentage representation of species with C-value data in each tribe is also given. C-value data taken from supplementary table S1, Supplementary Material online. (N.B. the tribe assignment for 15 species is currently unclear and they were excluded from this analysis)**

| Tribe | Min. (pg) | Max. (pg) | Mean (pg) | Number of Species with C-Values | Approximate Number of Species in Tribe[a] | Approximate Percentage Representation |
|---|---|---|---|---|---|---|
| Aethionemeae | 0.71 | 0.71 | 0.71 | 2 | 57 | 3.5 |
| Alysseae | 0.53 | 2.26 | 1.07 | 8 | Unclear | — |
| Anchonieae | 1.08 | 2.43 | 1.94 | 8 | 130 | 6.1 |
| Arabideae | 0.24 | 1.46 | 0.45 | 21 | 460 | 4.6 |
| Boechereae | 0.24 | 0.24 | 0.24 | 2 | 110 | 1.8 |
| Brassiceae | 0.23 | 1.31 | 0.70 | 27 | 230 | 11.7 |
| Camelineae | 0.16 | 0.83 | 0.37 | 24 | 240 | 10.0 |
| Cardamineae | 0.20 | 1.70 | 0.79 | 9 | 340 | 2.6 |
| Chorisporeae | 0.35 | 0.35 | 0.35 | 1 | 12 | 8.3 |
| Cochlearieae | 0.40 | 1.40 | 0.81 | 4 | 21 | 19.0 |
| Descurainieae | 0.17 | 0.23 | 0.21 | 8 | 60 | 13.3 |
| Euclidieae | 0.23 | 1.69 | 0.75 | 7 | 150 | 4.7 |
| Eutremeae | 0.32 | 0.32 | 0.32 | 1 | 25 | 4.0 |
| Halimolobeae | 0.15 | 0.18 | 0.16 | 8 | 40 | 20 |
| Heliophileae | 0.38 | 0.43 | 0.41 | 2 | 82 | 2.4 |
| Iberideae | 0.56 | 0.61 | 0.58 | 3 | 27 | 11.0 |
| Isatideae | 0.29 | 0.58 | 0.39 | 4 | 90 | 4.4 |
| Lepidieae | 0.33 | 1.04 | 0.58 | 4 | 240 | 1.7 |
| Noccaeeae | 0.24 | 1.00 | 0.38 | 9 | 85 | 10.6 |
| Physarieae | 0.26 | 2.34 | 1.19 | 5 | 150 | 3.3 |
| Schizopetaleae | 0.38 | 0.70 | 0.54 | 4 | 230 | 1.7 |
| Sisymbrieae | 0.24 | 0.53 | 0.38 | 3 | 40 | 7.5 |
| Smelowskieae | 0.46 | 0.46 | 0.46 | 1 | 25 | 4 |
| Thlaspideae | 0.31 | 1.90 | 0.82 | 5 | 26 | 19.2 |

[a] The approximate number of species in each tribe was taken from Al-Shehbaz et al. (2006).

## Results

### Molecular Systematics and Evolutionary Framework

Following "pruning" of the larger tree, eight clades were identified (see fig. 1*A*). These were assigned numbers and are referred to in the text below. However, the results from phylogenetic analysis prior to pruning (supplementary fig. S2, Supplementary Material online) corresponded well with our present-day knowledge on phylogenetic relationships among crucifers (Bailey et al. 2006; Beilstein et al. 2006; Koch et al. 2007). Taxon sampling covered 24 out of 25 tribes recognized by Al-Shehbaz et al. (2006). The one missing tribe (Hesperideae) comprises only one genus and c. 46 species (Warwick et. al 2006), and its absence did not affect our phylogenetic framework. Hesperideae has already been shown to be closely related to the probably polyphyletic tribe Anchonieae (Warwick et al. 2007). However, phylogenetic relationships among the eight clades detected here still needs further data because clades 5–7 (closely related) appear paraphyletic in our analysis, although they have been combined to one single lineage in two independent analyses (Beilstein et al. 2006; Koch et al. 2007).

Several tribes in our analysis appeared polyphyletic: For Camelineae and Alysseae, these findings are consistent with previous studies (Bailey et al. 2006; Warwick et al. 2007). Additional studies revealed also that tribes Euclidieae and Anchonieae (Al-Shehbaz and Warwick 2007; Warwick et al. 2007) need to be redefined, which are fully consistent with our presented data here (supplementary fig. S2, Supplementary Material online). However, here Brassiceae was also shown to be polyphyletic, a finding that conflicts with previous analyses (Lysak et al. 2005; Warwick and Sauder 2005; Bailey et al. 2006; Koch et al. 2007), and future research is necessary to resolve evolutionary processes in this particular tribe characterized by an ancient genome triplication (Lysak et al. 2005)

### GS Variation across the *Brassicaceae*

This paper reports new C-values for 114 taxa from 22 tribes (supplementary table S1, Supplementary Material online). These data were combined with an additional 71 C-values taken either from the Plant DNA C-values database (Bennett and Leitch 2005) or from more recent publications not yet incorporated in the database to give a data set comprising 185 *Brassicaceae* taxa in 24 out of the 25 tribes currently recognized in *Brassicaceae* (Al-Shehbaz et al. 2006, see supplementary table S1, Supplementary Material online). A summary of the phylogenetic spread of data and the minimum, maximum, and mean C-values for each tribe are given in table 1 and figures 1 and 3.

### Bayesian Inference of ancGS in *Brassicaceae*

A Bayes factor of 2 indicated that there was not strong evidence to support a model with an additional parameter for directional evolution. Therefore, a random-walk model that incorporated the branch length transformation parameters was fit to our GS data. This model was very strongly supported by a Bayes factor of 1,056 when compared with the null hypothesis of simple Brownian motion, where all scaling parameters were set to 1.

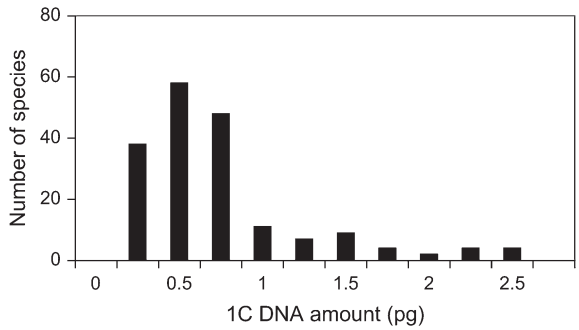


FIG. 2.—Histogram showing the distribution of GSs in 185 *Brassicaceae* taxa.

The mean $\kappa$ estimate was 1.72 (95% confidence interval [CI] ± 0.004; fig. 4*A*). The posterior distribution was significantly greater than 1 ($P = 0.012$), indicating that GS evolved faster within longer branches. For $\delta$, the posterior distribution centered on 1.01 (95% CI ± 0.005) and was not significantly different from 1 ($P = 0.455$; fig. 4*B*). This suggests that across the entire *Brassicaceae* phylogeny, the rate of GS evolution has been constant. The mean $\lambda$ estimate was 0.843 (95% CI ± 0.001; fig. 4*C*), indicating that GS among members of *Brassicaceae* exhibits strong phylogenetic signal. By incorporating these parameters into the evolutionary model, the back-transformed mean 1C-value for the ancGS was calculated from the posterior distribution of the log ancestral estimate ($\alpha$) and shown to be 0.504 pg (95% CI = 0.502–0.507; fig. 4*D*).

### Phylogenetically Corrected *t*-tests

From examining the mean of each major clade within *Brassicaceae* (see fig. 1*B*), it is clear that clade 4 has a greater mean than the rest of *Brassicaceae*. However, the phylogenetically corrected *t*-test results suggested that the observed differences between all major clades could have evolved randomly (passively) rather than due to selective processes (*P* values ranged from 0.185 to 0.465).

## Discussion

### GS Diversity in *Brassicaceae*

The smallest 1C-values reported to date are found in *A. thaliana* (1C = 0.16 pg, 2n = 10; Bennett et al. 2003) and three *Sphaerocardamum* species (1C = 0.15–0.16 pg, 2n = 16; Bailey 2001) endemic to northern and central America. At the other end of the scale, the largest 1C-value so far is found in *Bunias orientalis* (1C = 2.43 pg, 2n = 14). This value corresponds well with the 1C-value of 2.64 pg reported by Greilhuber and Obermayer (1999) for the same species. Nevertheless, it is clear from the histogram shown in figure 2 that although the GS data range 16.2-fold, most *Brassicaceae* species are characterized by small genomes with a mean, median, and modal 1C-value for 185 taxa of 0.63, 0.46, and 0.40 pg, respectively. Overall, considering the GS categories of Leitch et al. (1998), all *Brassicaceae* species analyzed so far have very small (≤1.4 pg) or small (≤3.5 pg/1C) genomes. Interestingly, this narrow range of GS contrasts with some other eudicot families of high species diversity and >100 GS estimates (e.g., *Asteraceae*—c. 23,000 species, 355 GS estimates, 1C-values range 65-fold; *Solanaceae*—c. 2,600 species, 175 GS estimates, 1C-values range 25.5-fold; and *Ranunculaceae*—c. 1,750 species, 155 GS estimates, 1C-values range 42-fold).

Out of the 185 GS estimates, the three largest genomes were found in species with low diploid chromosome numbers belonging to *Bunias*, *Matthiola*, and *Physaria*. The possible significance of this is discussed below. Other

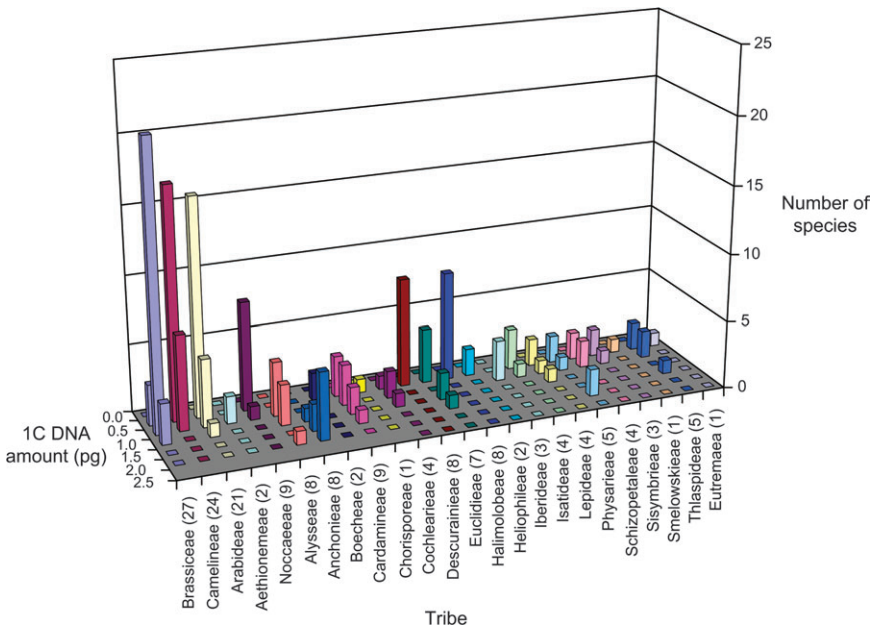


FIG. 3.—Histogram showing the distribution of 1C DNA amounts in each of the 24 tribes of *Brassicaceae* identified by Al-Shehbaz et al. (2006) although not all these tribes are considered to be monophyletic (e.g., Alysseae—see text). The number following the tribal name represents the number of species with GS data.

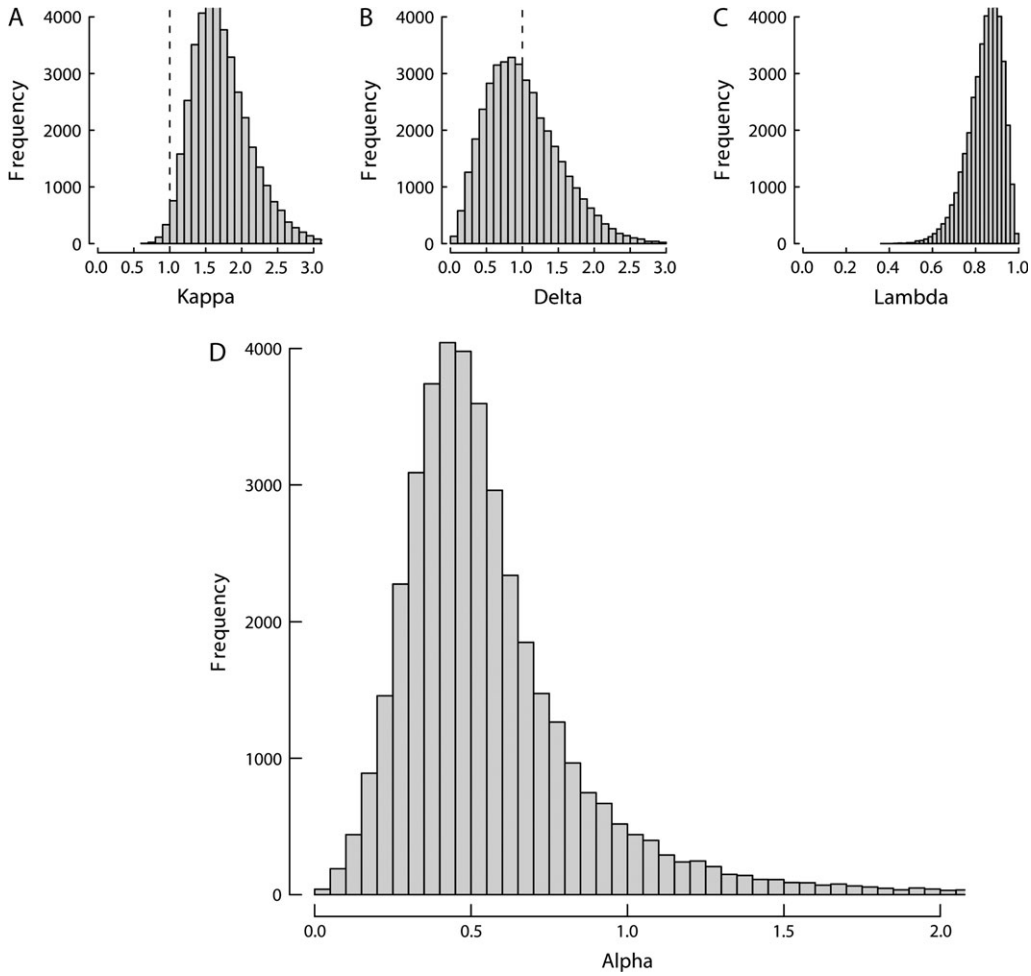


FIG. 4.—The posterior distributions of the MCMC run through 21,00,000 iterations for (*A*) $\kappa$, (*B*) $\delta$, (*C*) $\lambda$, and (*D*) the back-transformed log ancGS estimate. For $\kappa$ (*A*), the posterior distribution was significantly greater than 1 (indicated by the dashed line; $P = 0.012$), indicating greater GS evolution in longer branches. For $\delta$ (*B*), the posterior distribution was not significantly less than 1 indicating constant GS evolution throughout the *Brassicaceae* phylogeny. For $\lambda$ (*C*), the distribution is near 1 indicating strong phylogenetic signal. (*D*) After incorporating these parameters, the back-transformed posterior distribution of the $\log_{10}$ ancGS centered on 0.504 (95% CI 0.502–0.507).

species with large genomes are either clearly polyploid (e.g., *Ptilotrichum canescens*, 2.26 pg/1C) or most likely represent the result of polyploidy events, although the lack of chromosome data prevents firm conclusions being drawn. For example, the 1C-value of *Cardamine pratensis* (Cardamineae) (1.7 pg/1C) most likely corresponds to a polyploid cytotype of this species, well known for its immense karyological variability with several ploidy levels and a continuous series of aneuploid numbers (Lövkvist 1956; Warwick and Al-Shehbaz 2006). The large GS of *Alliaria petiolata* (Thlaspideae) (1.9 pg/1C) compared with other members of clade 1 (fig. 1*A*; where 1C-values range from 0.21 to 1.09 pg) probably reflects the hexaploid (2n = 6x = 42) status of this species.

## Statistical Analysis of GS Evolution

Over the years, there have been a number of different approaches to analyzing GS evolution within a phylogenetic framework. Some of the first studies superimposed GS data directly onto phylogenetic trees to make inferences of ancGS and directions of GS change (e.g., Cox et al. 1998; Leitch et al. 1998). More recently, various statistical approaches have also been used to provide insights into the dynamics of GS evolution and to reconstruct ancGS. These include programs such as MacClade (Maddison WP and Maddison DR 1992), which can use parsimony or squared-change parsimony (e.g., Soltis et al. 2003; Albach and Greilhuber 2004; Kellogg and Bennetzen 2004; Caetano-Anolles 2005; Leitch et al. 2005), and Compare (Martins 2004) and Continuous (Pagel 1997, 1999), which use maximum likelihood and GLS methods (e.g., Wendel et al. 2002; Albach and Greilhuber 2004; Jakob et al. 2004; Weiss-Schneeweiss et al. 2005; Leitch et al. 2007). Phylogenetic independent contrasts (Felsenstein 1985; Harvey and Pagel 1991) implemented in various statistical packages such as Analysis of Traits module (Ackerly 2006) of Phylocom (Webb et al. 2006) have also been applied (e.g., Leitch et al. 2007).

The picture emerging from these studies is that GS evolution is dynamic with both increases and decreases being reported. Such patterns have been observed at all taxonomic scales from whole groups such as angiosperms and land plants (Leitch et al. 1998; Soltis et al. 2003; Leitch

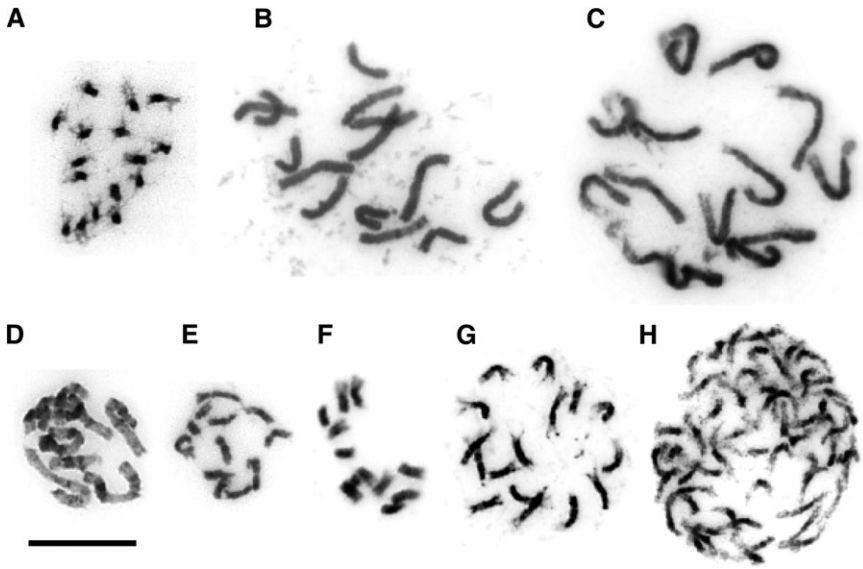


FIG. 5.—Mitotic chromosomes of *Capsella rubella* (Camelineae, *A*) and species from tribes Buniadeae (*B*), Anchonieae (*C*), and Physarieae (*D–H*) taken at the same magnification. (*A*) *Capsella rubella* (2n = 16, 0.22 pg/1C; F), (*B*) *Bunias orientalis* (2n = 14, 2.43 pg/1C; R), (*C*) *Matthiola sinuata* (2n = 14, 1C = 2.25 pg/1C; R), (*D*) *Physaria bellii* (2n = 8, 2.34 pg/1C; F), (*E*) *Physaria gracilis* (2n = 12, 0.26 pg/1C; F), (*F*) *Physaria ovalifolia* (2n = 12, 0.43 pg/1C; R), (*G*) *Physaria arctica* (2n = 16, 0.69 pg/1C; F), and (*H*) *Physaria didymocarpa* (2n = 56, 2.23 pg/1C; F). Chromosome preparations were prepared from root tips (R) or young flower buds (F). Chromosomes were counterstained by DAPI and the images inverted for better contrast. Scale = 5 μm.

et al. 2005) down to the level of the genus (e.g., *Gossypium*, *Hordeum*, and *Orobanche*; Wendel et al. 2002; Jakob et al. 2004; Weiss-Schneeweiss et al. 2005). As noted in the introduction, such dynamic patterns were also reported in *Brassicaceae* on a small and taxonomically restricted data set (Johnston et al. 2005). Using the increasingly robust phylogenetic tree together with a significantly larger data set encompassing the phylogenetic breadth of *Brassicaceae*, further evidence of dynamic GS evolution has been uncovered.

In the present work, the use of Bayesian MCMC methods allowed the incorporation of uncertainty into our model of GS evolution. In our case, there was uncertainty in the topology of our underlying phylogenetic hypothesis. The estimation of GLS parameters for the best model was therefore integrated over a distribution of trees. Through this approach, we have gained robust insights into the ancGS, mode, and tempo of GS evolution in *Brassicaceae*.

## Reconstruction of the ancGS for *Brassicaceae*

The ancGS for *Brassicaceae* inferred in the present study was 1C = c. 0.5 pg compared with 0.2 pg found by Johnston et al. (2005). Moreover, the 0.2 estimate found by Johnston et al. (2005) significantly differed from our posterior distribution of the most likely ancestral values ($P = 0.039$). Johnston et al. noted that their estimate was tentative, considering the small size of their data set and the lack of statistical analysis. Given the improved representation of GS data used in the present study (both phylogenetically and number of species) and the statistical approaches taken, the present result is considered more robust and highlights the dynamic nature of GS evolution in *Brassicaceae* with 52% of species showing evidence of GS decreases and the remaining species showing increases during their evolution (fig. 1*A*).

Perhaps surprisingly, the ancGS estimate of 0.5 pg is somewhat different from the 1C-values estimated for two species of *Aethionema* (0.71 pg/1C; 2n = 48 in *Aethionema schistosum*, 0.71 pg/1C in *Aethionema grandiflorum*), a genus considered sister to all other *Brassicaceae* (Al-Shehbaz et al. 2006; Beilstein et al. 2006). Possible explanations include 1) GS data are available for just 2 of the 56 *Aethionema* species currently recognized, many of which are polyploid including *A. schistosum* reported here (Al-Shehbaz et al. 2006; Warwick and Al-Shehbaz 2006), hence the data may not be representative, and 2) recent sequence data suggest some *Aethionema* species with low (diploid) chromosome numbers should be placed into the genus *Eunomia* (Al-Shehbaz et al. 2006) leaving a redefined genus comprising just polyploid species with small cruciferlike chromosomes (fig. 5*A*). Given that multiple rounds of polyploidy will mask the ancGS for the genus, insights into the ancGS of *Brassicaceae* through a study of GS variation in *Aethionema* is unlikely to be informative.

The complexity of assessing the inferred ancGS is also underlined by the palaeohexaploid genome structure uncovered in *Cleomaceae* (Schranz and Mitchell-Olds 2006), the family considered sister to *Brassicaceae* (Hall et al. 2002, 2004). Although a few GS estimates (all in the region of 1C = 0.3 pg) have been reported for *Cleome* (the largest genus in the family), the higher rate of gene loss following polyploidy in *Cleomaceae* compared with *Brassicaceae* (Schranz and Mitchell-Olds 2006) suggests that insights into the ancGS for *Brassicaceae* may also not be gleaned from a study of GS in *Cleomaceae*.

## Mode and Tempo of GS Evolution

Our analysis showed that GS evolution did not follow a purely random-walk model (Brownian motion). Instead,

there was strong evidence for a model showing GS evolution as gradual and branch length dependent. The posterior distribution for $\kappa$ was significantly greater than 1 (see Results, $P = 0.012$), indicating that evolutionary gains or losses in GS accumulate disproportionately faster within longer branches than within shorter branches. This result is novel for plants. Previous analyses across various angiosperm groups have consistently shown GS evolution to be punctuated and uncorrelated with branch length (*Veronica*, Albach and Greilhuber 2004; *Orobanche*, Weiss-Schneeweiss et al. 2005; *Liliaceae*, Leitch et al. 2007). However, for both *Veronica* and *Orobanche*, GLS analyses were carried out across phylogenies using equal branch lengths (i.e., all branches set to 1). This assumes that trait evolution occurred at the time of speciation, which would inevitably lead to a punctuated interpretation (see Martins and Garland 1991). The only other reported case of punctuated evolution that did not assume equal branch lengths was for *Liliaceae* (Leitch et al. 2007), a family that contains species with the largest GSs so far reported for angiosperms (1C-values range 23-fold from 3.4 to 77.4 pg; Leitch et al. 2007).

It is intriguing to consider that perhaps the difference between our results and those of *Liliaceae* represents a GS-dependent shift in the mode of evolution, given the fact that *Brassicaceae* and *Liliaceae* are characterized by C-values lying at the extreme poles of GS distribution in angiosperms. As GS increases, larger gains or losses are possible and manifest as larger changes per unit branch length than would be possible in species of smaller GS. This leads to the expectation that with increasing GS, evolutionary change becomes more unrelated to branch length and is thus viewed as punctuated. Support for this hypothesis comes from comparative studies indicating that some underlying genomic processes may be GS dependent and/or that the relative rates of activity of particular mechanisms may be different for species with different GS. For example, studies on mutation and recombination processes in *Arabidopsis* (0.16 pg/1C) and tobacco (*Nicotiana tabacum* c. 5.0 pg/1C) highlighted marked differences, such as an inverse relationship between deletion size and GS, which may have contributed to their c. 30-fold difference in DNA amount (Kirik et al. 2000; Filkowski et al. 2004; Puchta 2005). Similarly, studies on insects have revealed how the rate and size spectrum of deletions are correlated with a c. 100-fold difference in GS (from 180 to 16,000 Mb; Bensasson et al. 2001; Petrov 2002; Vitte and Bennetzen 2006). Whether these differences in patterns are a cause or consequence of GS change is, however, currently unknown and clearly, more work is needed not only to assess the biological significance of our results but also to provide a better understanding of the underlying mode of GS evolution across all higher plants.

The statistical analysis also suggests that across *Brassicaceae* GS is not strongly influenced by selection. This conclusion is based on the posterior distribution of the $\lambda$ parameter (fig. 4*C*) which showed that GS exhibited a strong phylogenetic signal, very near the expectation for a stochastic process (i.e., closely related species are more similar in GS than more distantly related lineages). If selective pressures had directly influenced GS, then species-specific responses would obscure phylogenetic signal (Blomberg et al. 2003). Furthermore, analysis of the observed differences in GS among the eight major clades (fig. 1*B*) showed that they could not be distinguished from having evolved by chance. No significant shifts among these major clades suggest that GS has not changed substantially through time. Thus, the narrow range of small GS that characterizes *Brassicaceae* is most likely due to a passive tempo of GS evolution rather than one where selection has played a prominent role.

## The Paradox of the Small Dynamic Genomes in *Brassicaceae*

From these insights into the GS profile of *Brassicaceae*, a key question remains—why is *Brassicaceae* genomically so dynamic and yet characterized by a narrow range of small GS? This observation is perhaps particularly surprising given that processes which are known to play major roles in generating GS increases, namely polyploidy and amplification of repetitive DNAs (especially long terminal repeat [LTR] retrotransposon), have been active in *Brassicaceae*.

Polyploidy is clearly an important evolutionary process in the family. About 37% of species are reported to be polyploid based on chromosome numbers that range from n = 4 to 128 (Warwick and Al-Shehbaz 2006), and this percentage increases considerably if paleopolyploidy events, such as those revealed in *Arabidopsis* and several other eudicot lineages, are also taken into account (The Arabidopsis Genome Initiative 2000; De Bodt et al. 2005; Cui et al. 2006). Although initially this will lead to an increase in DNA amount, the comparatively narrow range of GS in *Brassicaceae* suggests that mechanisms which lead to the diploidization of the genome must include efficient methods to eliminate DNA.

Indeed, many detailed comparative genomic studies have provided evidence of extensive DNA loss following polyploidy in *Brassicaceae* (e.g., Song et al. 1995; O'Neill and Bancroft 2000; Parkin et al. 2005; Town et al. 2006; Gaeta et al. 2007). In contrast, reports of genome expansion following polyploidy are rare in angiosperms as a whole (Leitch et al. 2008), and in *Brassicaceae*, this has so far only been tentatively suggested for the hexaploid *Cardamine asarifolia* (2n = 6x = 48) based on DNA sequence analysis (Lihova et al. 2006). The mechanisms leading to DNA loss are poorly understood but likely to include various types of recombination including unequal crossing-over and illegitimate recombination (Comai 2000; Devos et al. 2002; Ma et al. 2004; Bennetzen et al. 2005; Gaut et al. 2007). Diploidization may also be accompanied by large-scale chromosome rearrangements, and these can lead not only to a reduction in chromosome number toward a diploid-like state but can also result in DNA loss (Lysak and Lexer 2006). For example, chromosomal rearrangements via unequal reciprocal translocations accompanied by DNA losses were recently demonstrated in the evolution of the derived karyotype of the paleopolyploid *Arabidopsis* (2n = 10) and other closely related species with reduced chromosome numbers (Lysak et al. 2006; Schranz et al. 2006). The more recent whole-genome triplication in the

tribe Brassiceae (7.9–14.6 Ma) was shown to have been followed by taxon- and lineage-specific chromosome rearrangements resulting in chromosome number reductions toward lower and sometimes diploid-like numbers (Lysak et al. 2005; Lysak et al. 2007). Also homeologous nonreciprocal transposition is suggested as a mechanism mediating DNA fragment losses (restriction fragment length polymorphism and simple sequence repeat) in resynthesized lines of *B. napus* (Gaeta et al. 2007).

Nevertheless, our data indicate that reductions in chromosome number are not always accompanied by genome downsizing. This is particularly evident in three genera which have among the largest genomes so far reported for *Brassicaceae*: *Physaria* (Physarieae), *Bunias*, and *Matthiola* (Anchonieae) (see supplementary table S1, Supplementary Material online).

1. The tribe Physarieae (c. 150 spp. in seven genera, Al-Shehbaz et al. 2006) is characterized by a huge diversity of chromosome numbers (2n = 8–140, e.g., fig. 5*D–H*, Warwick and Al-Shehbaz 2006) including species with the lowest chromosome number reported for *Brassicaceae* (2n = 8). In the present study, the large GS reported in diploid *P. bellii* (2n = 8; 2.34 pg/1C, fig. 5*D*) with only four chromosome pairs is similar to that in the polyploid species *Physaria didymocarpa* (2n = 56; 2.23 pg/1C, fig. 5*H*) (Lysak and Lexer 2006) but considerably larger than other species with low chromosome numbers (i.e., *Physaria gracilis* 2n = 12, 0.26 pg/1C; *Physaria ovalifolia*, 2n = 12, 0.43 pg/1C; *Physaria arctica*, 2n = 16, 0.69 pg/1C, fig. 5*E–G*, respectively). Figure 5*D–H* shows chromosome size variation in all *Physaria* species analyzed including the largest chromosomes found in *P. bellii* (2n = 8). These observations suggest that the large genome of *P. bellii* arose via one or more polyploidization events followed by a series of chromosome rearrangements leading to chromosome number reduction. Nevertheless, the scarcity of data prevents firm conclusions, and further data are clearly needed to elucidate the modes of genome evolution in this lineage (S. Fuentes-Soriano and I.A. Al-Shehbaz, unpublished data).
2. The tribe Anchonieae comprises c. 130 species in 12 genera (Al-Shehbaz et al. 2006), however, the genus *Bunias* was recently recognized as an independent tribe Buniadeae (Al-Shehbaz and Warwick 2007). Anchonieae *s. l.* includes species with the largest genomes so far reported in *Brassicaceae* (i.e., *B. orientalis* 2.43 pg/1C and five species of *Matthiola* 1.60–2.29 pg/1C; see supplementary table S1, Supplementary Material online). Once again, these large genomes are striking considering that the *Bunias* and *Matthiola* species studied are characterized by low diploid-like chromosome numbers (2n = 10–16 in *Matthiola* and 2n = 14 in *Bunias*; Warwick and Al-Shehbaz 2006; Warwick et al. 2007). However, their chromosomes are larger than the average chromosomes of most crucifer taxa (compare fig. 5*A* with fig. 5*B* and *C*) (present study and Manton 1932). This karyotype structure may have arisen in two ways: 1) the diploid-like number and large size of the chromosomes may have arisen from the diploidization of ancient polyploidy events involving chromosome number reduction mediated by chromosome fusions as noted above but not accompanied by extensive DNA loss. Alternatively, 2) retrotransposon amplification may have occurred leading to an increase in GS and hence chromosome size within a diploid karyotype framework. The latter scenario has not been reported in *Brassicaceae* to date, but evidence for this mode of evolution has been documented, for example, in species of *Oryza* (*Poaceae*) differing 3-fold in GS (Zuccolo et al. 2007). Further genomic and cytogenetic analyses are needed to resolve the mode of karyotype evolution in these genera. The phylogenetic relationships within the tribe have been studied by Warwick et al. (2007) who distinguished two major lineages—Anchonieae I and II. It is anticipated that other taxa within these two Anchonieae lineages which have relatively low chromosome numbers (e.g., *Clausia*, x = 7; *Dontostemon*, x = 7) might also have large GS.

Similar trends may also be observed in other genera although GS data are currently lacking. For example, Manton (1932) noted that some species with low diploid-like chromosome numbers had surprisingly large chromosomes. Genera studied included *Hesperis* (2n = 12, 14), some *Iberis* species (2n = 14), and *Menonvillea* (2n = 22), and such observations suggest that large genomes may be found in these genera too. Negative correlation between GS and chromosome number was also observed in noncrucifer taxa such as the tribe Cardueae (Asteraceae) (Garnatje et al. 2004).

In the present study (see supplementary table S1, Supplementary Material online), several other taxa with diploid-like chromosome numbers (2n = 14, 16) exhibited increased GS and larger chromosomes (data not shown). Three species from the tribe Alysseae (*Degenia velebitica*, 2n = 16, 1.25 pg/1C; two *Fibigia* species, 2n = 16, 1.25 and 1.29 pg/1C) (Al-Shehbaz et al. 2006), *Parrya nudicaulis* (2n = 14, 1.08 pg/1C) assigned to Chorisporeae (Warwick et al. 2007), *Desideria linearis* (2n = 14, 1.38 pg/1C) from Euclidieae (Warwick et al. 2007), and *Christolea crassifolia* (2n = 14, 1.41 pg/1C, without tribal assignment) showed increased C-values (supplementary table S1, Supplementary Material online) accompanied by relatively large chromosomes (data not shown). The significance of this pattern remains elusive and perhaps restricted to individual genera as in all three tribes some diploid species of different genera possess smaller genomes (supplementary table S1, Supplementary Material online) and average-sized chromosome complements (Alysseae: *Berteroa*; Chorisporeae: *Chorispora*; and Euclidieae: *Euclidium*).

Apart from polyploidy and diploidization processes discussed above, GS differences in plants are largely associated with differences in the amount of repetitive DNA they contain (Flavell et al. 1974). Amplification of repetitive elements, particularly LTR retrotransposons which account for much of the nuclear DNA in plants, is considered to be a major force behind GS increase and plant genome obesity. The narrow range of small genomes observed in *Brassicaceae* suggests two possible explanations. Either plants with small genomes possess an as-yet unidentified mechanisms (or constraint?) impeding amplification of

repetitive elements and/or they have an efficient mechanism of continuous removal of amplified sequences to counteract their proliferation. Recent research data provide evidence that both mechanisms are operating in *Brassicaceae*.

Mechanisms to eliminate repetitive DNA especially LTR retrotransposons, while still poorly understood, include various recombination processes as mentioned above (Gaut et al. 2007). These include different types of intra- and interchromosomal recombination involving both mitotic and meiotic chromosomes with resulting products being potentially transmitted to the offspring. Unequal homologous crossing-over between repeats (e.g., LTRs of LTR retrotransposons) within the same chromatid (intrastrand crossing-over), between sister chromatids or homologous chromosomes cause deletions (e.g., solo-LTRs) and duplications. Similarly, deletions can arise via illegitimate recombination (nonallelic homologous or ectopic recombination), that is, recombination between nonhomologous sequences with a limited nucleotide homology. Deletions, caused by illegitimate recombination (manifest as truncated LTR retrotransposons and solo-LTRs), have been shown to be a driving force behind genome contraction in *Arabidopsis* where it was suggested that this process was at least 5-fold more effective than unequal homologous recombination in removing DNA as it could act on a larger part of the genome (Devos et al. 2002).

Evidence that mechanisms exist to suppress the amplification of repetitive DNA and hence reduce their impact on GS has been provided by several studies. In *Brassicaceae*, Zhang and Wessler (2004) made a comparative study of the diversity and evolutionary dynamics of LTR retrotransposons in *A. thaliana* and *B. oleracea* (Zhang and Wessler 2004). These *Brassicaceae* species belonging to two different major crucifer phylogenetic clades differ c. 4-fold in GS (*A. thaliana*: tribe Camelineae, clade 5 in fig. 1*B*, 0.16 pg/1C and *B. oleracea*: tribe Brassiceae, clade 1 in fig. 1*B*, 0.71 pg/1C). The authors looked to see whether the larger *Brassica* genome had experienced lineage-specific GS expansion or if an ancestral genome shared by both species has been reduced in *Arabidopsis* and not in *B. oleracea*. Both species were shown to possess virtually the same repertoire of transposable elements (TEs), but in *B. oleracea*, the copy numbers of certain repeats were significantly higher. The most parsimonious explanation is therefore that the 4-fold GS difference between the two species is due to a more limited proliferation of LTR retrotransposons in *Arabidopsis* rather than to their extensive elimination. It can only be speculated that the inhibited amplification is due to the high gene density in the compact *Arabidopsis* genome; where each TE insertion is potentially deleterious. Conversely, the whole-genome triplication in *Brassica* (and other polyploid genomes) "would have produced numerous safe havens for TE insertions because of functional redundancy" (Zhang and Wessler 2004).

## Supplementary Material

Supplementary tables S1–S3 and figures S1 and S2 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

## Acknowledgments

The work was supported by a grant from the Grant Agency of the Czech Academy of Science (KJB601630606) awarded to M.A.L. The support of the Czech Ministry of Education (grant no. MSM0021622415) is acknowledged.

## Literature Cited

Ackerly DD. 2006. Analysis of traits (AOT): a module of phylocom. Version 3.1. Available from: http://www.phylodiversity.net/phylocom.

Albach DC, Greilhuber J. 2004. Genome size variation and evolution in *Veronica*. Ann Bot. 94:897–911.

Al-Shehbaz IA, Beilstein MA, Kellogg EA. 2006. Systematics and phylogeny of the Brassicaceae (Cruciferae): an overview. Plant Syst Evol. 259:89–120.

Al-Shehbaz IA, Warwick SI. 2007. Two new tribes (Dontostemoneae and Malcolmieae) in the Brassicaceae (Cruciferae). Harv Pap Bot. 12:429–433.

Bailey CD. 2001. Systematics of *Sphaerocardamum* (Brassicaceae) and related genera. PhD thesis, Cornell University, Ithaca, New York.

Bailey CD, Koch MA, Mayer M, Mummenhoff K, O'Kane SL, Warwick SI, Windham MD, Al-Shehbaz IA. 2006. Toward a global phylogeny of the Brassicaceae. Mol Biol Evol. 23:2142–2160.

Beilstein MA, Al-Shehbaz IA, Kellogg EA. 2006. Brassicaceae phylogeny and trichome evolution. Am J Bot. 93:607–619.

Bennett MD. 1972. Nuclear DNA content and minimum generation time in herbaceous plants. Proc R Soc Lond B Biol Sci. 181:109–135.

Bennett MD, Leitch IJ. 2005. Plant DNA C-values database (release 4.0, Oct. 2005) [Internet]. Available from: http://www.kew.org/genomesize/homepage.

Bennett MD, Leitch IJ, Price HJ, Johnston JS. 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. 91: 547–557.

Bennetzen JL, Ma J, Devos KM. 2005. Mechanisms of recent genome size variation in flowering plants. Ann Bot. 95: 127–132.

Bensasson D, Petrov DA, Zhang DX, Hartl DL, Hewitt GM. 2001. Genomic gigantism: DNA loss is slow in mountain grasshoppers. Mol Biol Evol. 18:246–253.

Berr A, Pecinka A, Meister A, Kreth G, Fuchs J, Blattner FR, Lysak MA, Schubert I. 2006. Chromosome arrangement and nuclear architecture but not centromeric sequences are conserved between *Arabidopsis thaliana* and *Arabidopsis lyrata*. Plant J. 48:771–783.

Blomberg SP, Garland T, Ives AR. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. Evolution. 57:717–745.

Bomblies K, Weigel D. 2007. *Arabidopsis*—a model genus for speciation. Curr Opin Genet Dev. 17:500–504.

Caetano-Anolles G. 2005. Evolution of genome size in the grasses. Crop Sci. 45:1809–1816.

Clauss MJ, Koch MA. 2006. Poorly known relatives of *Arabidopsis thaliana*. Trends Plant Sci. 11:449–459.

Comai L. 2000. Genetic and epigenetic interactions in allopolyploid plants. Plant Mol Biol. 43:387–399.

Cox AV, Abdelnour GJ, Bennett MD, Leitch IJ. 1998. Genome size and karyotype evolution in the slipper orchids (Cypripedioideae: Orchidaceae). Am J Bot. 85:681–687.

Cui L, Wall PK, Leebens-Mack JH, et al. (13 co-authors). 2006. Widespread genome duplications throughout the history of flowering plants. Genome Res. 16:738–749.
De Bodt S, Maere S, Van de Peer Y. 2005. Genome duplication and the origin of angiosperms. Trends Ecol Evol. 20:591–597.
Devos KM, Brown JKM, Bennetzen JL. 2002. Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. Genome Res. 12: 1075–1079.
Dobeš C, Kiefer C, Kiefer M, Koch M. 2007. *Trn*F pseudogenes in North American genus *Boechera* (Brassicaceae): mechanistic aspects of evolution. Plant Biol. 9:1–14.
Federspiel N. 2000. Deciphering a weed. Genomic sequencing of Arabidopsis. Plant Physiol. 124:1456–1459.
Felsenstein J. 1985. Phylogenies and the comparative method. Am Nat. 125:1–15.
Filkowski J, Kovalchuk O, Kovalchuk I. 2004. Dissimilar mutation and recombination rates in Arabidopsis and tobacco. Plant Sci. 166:265–272.
Flavell RB, Bennett MD, Smith JB, Smith DB. 1974. Genome size and proportion of repeated nucleotide sequence DNA in plants. Biochem Genet. 12:257–269.
Franzke A, German D, Al-Shehbaz IA. 2008. Arabidopsis family ties: molecular phylogeny and age estimates in the Brassicaceae. Taxon. (forthcoming).
Gaeta RT, Pires JC, Iniguez-Luy F, Leon E, Osborn T. 2007. Genomic changes in resynthesized *Brassica napus* and their effect on gene expression and phenotype. Plant Cell. 19: 3403–3417.
Galbraith DW, Harkins KR, Maddox JM, Ayres NM, Sharma DP, Firoozabady E. 1983. Rapid flow cytometric analysis of the cell cycle in intact plant tissues. Science. 220:1049–1051.
Garnatje T, Vallès J, Garcia S, Hidalgo O, Sanz M, Canela MÁ, Siljak-Yakovlev S. 2004. Genome size in *Echinops* L. and related genera (Asteraceae, Cardueae): karyological, ecological and phylogenetic implications. Biol Cell. 96:117–124.
Gaut BS, Wright SI, Rizzon C, Dvorak J, Anderson LK. 2007. Recombination: an underappreciated factor in the evolution of plant genomes. Nat Rev Genet. 8:77–84.
Gendrel A-V, Colot V. 2005. *Arabidopsis* epigenetics: when RNA meets chromatin. Curr Opin Plant Biol. 8:142–147.
Greilhuber J, Obermayer R. 1999. Cryptopolyploidy in *Bunias* (Brassicaceae) revisited—a flow-cytometric and densitometric study. Plant Syst Evol. 218:1–4.
Greilhuber J, Dolešel J, Lysak MA, Bennett MD. 2005. The origin, evolution and proposed stabilization of the terms 'Genome size' and 'C-value' to describe nuclear DNA contents. Ann of Bot 95:255-260
Hall AE, Kettler GC, Preuss D. 2006. Dynamic evolution at pericentromeres. Genome Res. 16:355–364.
Hall JC, Iltis HH, Sytsma KJ. 2004. Molecular phylogenetics of core Brassicales, placement of orphan genera *Emblingia*, *Forchhammeria*, *Tirania*, and character evolution. Syst Bot. 29:654–669.
Hall JC, Sytsma KJ, Iltis HH. 2002. Phylogeny of Capparaceae and Brassicaceae based on chloroplast sequence data. Am J Bot. 89:1826–1842.
Harvey PH, Pagel MD. 1991. The comparative method in evolutionary biology. Oxford series in ecology and evolution. Oxford: Oxford University Press.
Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol. 23:254–267.
Huson DH, Dezulian T, Kloepper T, Steel M. 2004. Phylogenetic super-networks from partial trees. IEEE/ACM Trans Comput Biol Bioinform. 1:151–158.
Jakob SS, Meister A, Blattner FR. 2004. Considerable genome size variation of *Hordeum* species (Poaceae) is linked to phylogeny, life form, ecology, and speciation rates. Mol Biol Evol. 21:860–869.
Johnston JS, Pepper AE, Hall AE, Chen ZJ, Hodnett G, Drabek J, Lopez R, Price HJ. 2005. Evolution of genome size in Brassicaceae. Ann Bot. 95:229–235.
Josefsson C, Dilkes B, Comai L. 2006. Parent-dependent loss of gene silencing during interspecies hybridization. Curr Biol. 16:1322–1328.
Kellogg EA, Bennetzen JL. 2004. The evolution of nuclear genome structure in seed plants. Am J Bot. 91:1709–1725.
Kirik A, Salomon S, Puchta H. 2000. Species-specific double-strand break repair and genome evolution in plants. EMBO J. 19:5562–5566.
Koch M, Dobes C, Mitchell-Olds T. 2003. Multiple hybrid formation in natural populations: concerted evolution of the internal transcribed spacer of nuclear ribosomal DNA (ITS) in North American *Arabis divaricarpa* (Brassicaceae). Mol Biol Evol. 20:338–350.
Koch M, Haubold B, Mitchell-Olds T. 2001. Molecular systematics of the Brassicaceae: evidence from coding plastome *matK* and nuclear *Chs* sequences. Am J Bot. 88:534–544.
Koch MA, Al-Shehbaz IA. 2008. Forthcoming. 2008. Molecular systematics and evolution of "wild" crucifers (Brassicaceae or Cruciferae). In: Gupta SK, editor. Biology and breeding of crucifers. Taylor and Francis Group.
Koch MA, Dobes C, Kiefer C, Schmickl R, Klimes L, Lysak MA. 2007. Supernetwork identifies multiple events of plastid $trnF_{(GAA)}$ pseudogene evolution in the Brassicaceae. Mol Biol Evol. 24:63–73.
Koch MA, Kiefer M. 2005. Genome evolution among cruciferous plants: a lecture from the comparison of the genetic maps of three diploid species—*Capsella rubella*, *Arabidopsis lyrata* subsp. *petraea*, and *A. thaliana*. Am J Bot. 92:761–767.
Laibach F. 1943. *Arabidopsis thaliana* (L.) Heynh. als Object für genetische und Entwicklungsphysiologische Untersuchungen. Bot Archiv. 44:439–455.
Leitch IJ, Beaulieu JM, Cheung K, Hanson L, Lysak M, Fay MF. 2007. Punctuated genome size evolution in Liliaceae. J Evol Biol. 20:2296–2308.
Leitch IJ, Chase MW, Bennett MD. 1998. Phylogenetic analysis of DNA C-values provides evidence for a small ancestral genome size in flowering plants. Ann Bot. 82:85–94.
Leitch IJ, Hanson L, Lim KY, Kovarik A, Chase MW, Clarkson JJ, Leitch AR. 2008. The ups and downs of genome size evolution in polyploid species of *Nicotiana* (Solanaceae). Ann Bot. 101:805–814.
Leitch IJ, Soltis DE, Soltis PS, Bennett MD. 2005. Evolution of DNA amounts across land plants (Embryophyta). Ann Bot. 95:207–217.
Lihova J, Shimizu KK, Marhold K. 2006. Allopolyploid origin of *Cardamine asarifolia* (Brassicaceae): incongruence between plastid and nuclear ribosomal DNA sequences solved by a single-copy nuclear gene. Mol Phylogenet Evol. 39:759–786.
Lim K-B, Yang T-J, Hwang Y-J, et al. (15 co-authors). 2007. Characterization of the centromere and peri-centromere retrotransposons in *Brassica rapa* and their distribution in related *Brassica* species. Plant J. 49:173–183.
Lövkvist B. 1956. The *Cardamine pratensis* complex. Outline of its cytogenetics and taxonomy. Symb Bot Ups. 14:1–131.
Lysak MA, Berr A, Pecinka A, Schmidt R, McBreen K, Schubert I. 2006. Mechanisms of chromosome number reduction in *Arabidopsis thaliana* and related Brassicaceae species. Proc Natl Acad Sci USA. 103:5224–5229.
Lysak MA, Cheung K, Kitschke M, Bures P. 2007. Ancestral chromosomal blocks are triplicated in Brassiceae species with varying chromosome number and genome size. Plant Physiol. 145:402–410.

Lysak MA, Koch MA, Pecinka A, Schubert I. 2005. Chromosome triplication found across the tribe Brassiceae. Genome Res. 15:516–525.
Lysak MA, Lexer C. 2006. Towards the era of comparative evolutionary genomics in Brassicaceae. Plant Syst Evol. 259:175–198.
Ma JX, Devos KM, Bennetzen JL. 2004. Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. Genome Res. 14:860–869.
Maddison WP, Maddison DR. 1992. MacClade: analysis of phylogeny and character evolution. Sunderland (MA): Sinauer Associates, Inc.
Maddison WP, Maddison DR. 2007. Mesquite: a modular system for evolutionary analysis. Version 2.01 [Internet]. Available from: http://mesquiteproject.org.
Manton I. 1932. Introduction to the general cytology of the Cruciferae. Ann Bot. 46:509–556.
Martins EP. 2004. COMPARE, version 4.6b. Computer programs for the statistical analysis of comparative data. Distributed by the author. Bloomington (IN): Department of Biology, Indiana University.
Martins EP, Garland T Jr. 1991. Phylogenetic analyses of the correlated evolution of continuous characters: a simulation study. Evolution. 45:534–557.
Meyerowitz EM. 1999. Today we have the naming of the parts. Nature. 402:731–732.
Nasrallah JB, Liu P, Sherman-Broyles S, Schmidt R, Nasrallah ME. 2007. Epigenetic mechanisms for breakdown of self-incompatibility in interspecific hybrids. Genetics. 175:1965–1973.
Oliver MJ, Petrov D, Ackerly D, Falkowski P, Schofield OM. 2007. The mode and tempo of genome size evolution in eukaryotes. Genome Res. 17:594–601.
O'Meara BC, Ane C, Sanderson MJ, Wainwright PC. 2006. Testing for different rates of continuous trait evolution using likelihood. Evolution. 60:922–933.
O'Neill CM, Bancroft I. 2000. Comparative physical mapping of segments of the genome of *Brassica oleracea* var. *alboglabra* that are homoeologous to sequenced regions of chromosomes 4 and 5 of *Arabidopsis thaliana*. Plant J. 23:233–243.
Organ CL, Shedlock AM, Meade A, Pagel M, Edwards SV. 2007. Origin of avian genome size and structure in non-avian dinosaurs. Nature. 446:180–184.
Oyama RK, Clauss MJ, Formanová N, Kroymann J, Schmid KJ, Vogel H, Weniger K, Windsor AJ, Mitchell-Olds T. 2008. The shrunken genome of *Arabidopsis thaliana*. Plant Syst Evol. 273:257–271.
Pagel M. 1997. Inferring evolutionary processes from phylogenies. Zool Scr. 26:331–348.
Pagel M. 1999. Inferring the historical patterns of biological evolution. Nature. 401:877–884.
Parkin IAP, Gulden SM, Sharpe AG, Lukens L, Trick M, Osborn TC, Lydiate DJ. 2005. Segmental structure of the *Brassica napus* genome based on comparative analysis with *Arabidopsis thaliana*. Genetics. 171:765–781.
Petrov DA. 2002. Mutational equilibrium model of genome size evolution. Theor Popul Biol. 61:531–544.
Puchta H. 2005. The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution. J Exp Bot. 56:1–14.
Schmickl R, Kiefer C, Dobeš C, Koch MA. Forthcoming. 2008. Evolution of *trn*F(GAA) pseudogenes in cruciferous plants. Plant Syst Evol. doi:10.1007/s00606-008-0030-2.
Schranz ME, Lysak MA, Mitchell-Olds T. 2006. The ABC's of comparative genomics in the Brassicaceae: building blocks of crucifer genomes. Trends Plant Sci. 11:535–542.
Schranz ME, Mitchell-Olds T. 2006. Independent ancient polyploidy events in the sister families Brassicaceae and Cleomaceae. Plant Cell. 18:1152–1165.
Schranz ME, Song BH, Windsor AJ, Mitchell-Olds T. 2007. Comparative genomics in the Brassicaceae: a family-wide perspective. Curr Opin Plant Biol. 10:168–175.
Smith SA, Dunne CW. 2008. Phyutility: a phyloinformatics tool for trees, alignments, and molecular data. Bioinformatics. 24:715-716.
Soltis DE, Soltis PS, Bennett MD, Leitch IJ. 2003. Evolution of genome size in the angiosperms. Am J Bot. 90:1596–1603.
Song KM, Lu P, Tang KL, Osborn TC. 1995. Rapid genome change in synthetic polyploids of *Brassica* and its implications for polyploid evolution. Proc Natl Acad Sci USA. 92:7719–7723.
Swofford DL. 2000. PAUP* phylogenetic analysis using parsimony (*and other methods). Version 4. Sunderland (MA): Sinauer Associates.
The Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature. 408:796–815.
Town CD, Cheung F, Maiti R, et al. (13 co-authors). 2006. Comparative genomics of *Brassica oleracea* and *Arabidopsis thaliana* reveal gene loss, fragmentation, and dispersal after polyploidy. Plant Cell. 18:1348–1359.
Vitte C, Bennetzen JL. 2006. Analysis of retrotransposon structural diversity uncovers properties and propensities in angiosperm genome evolution. Proc Natl Acad Sci USA. 103:17638–17643.
Warwick SI, Al-Shehbaz IA. 2006. Brassicaceae: chromosome number index and database on CD-Rom. Plant Sys Evol. 259:237–248.
Warwick SI, Francis A, Al-Shehbaz IA. 2006. Brassicaceae: Species checklist and database on CD-Rom. Pl. Syst. Evol. 259:249-258.
Warwick SI, Sauder CA. 2005. Phylogeny of tribe Brassiceae (Brassicaceae) based on chloroplast restriction site polymorphisms and nuclear ribosomal internal transcribed spacer and chloroplast *trn*L intron sequences. Can J Bot. 83:467–483.
Warwick SI, Sauder CA, Al-Shehbaz IA, Jacquemoud F. 2007. Phylogenetic relationships in the tribes Anchonieae, Chorisporeae, Euclidieae, and Hesperideae (Brassicaceae) based on nuclear ribosomal its DNA sequences. Ann Missouri Bot Gard. 94(1):56–78.
Webb CO, Ackerly DD, Kembel SW. 2006. Phylocom: software for the analysis of community phylogenetic structure and character evolution c3.40 [Internet]. Available from: http://phylodiversity.net/phylocom.
Weiss-Schneeweiss H, Greilhuber J, Schneeweiss GM. 2005. Genome size evolution in holoparasitic *Orobanche* (Orobanchaceae) and related genera. Am J Bot. 93:148–156.
Wendel JF, Cronn RC, Johnston JS, Price HJ. 2002. Feast and famine in plant genomes. Genetica. 115:37–47.
Yang T-J, Kim J-S, Lim K-B, et al. (12 co-authors). 2005. The Korea *Brassica* genome project: a glimpse of the *Brassica* genome based on comparative genome analysis with *Arabidopsis*. Comp Funct Genomics. 6:138–146.
Zhang X, Wessler SR. 2004. Genome-wide comparative analysis of the transposable elements in the related species *Arabidopsis thaliana* and *Brassica oleracea*. Proc Natl Acad Sci USA. 101:5589–5594.
Zuccolo A, Sebastian A, Talag J, Yu Y, Kim H, Collura K, Kudrna D, Wing RA. 2007. Transposable element distribution, abundance and role in genome size variation in the genus *Oryza*. BMC Evol Biol. 7:1–15.

Barbara Holland, Associate Editor

Accepted September 30, 2008