

Research Article

On the Tempo of Genome Size Evolution in Angiosperms

Jeremy M. Beaulieu,¹ Stephen A. Smith,² and Ilia J. Leitch³

¹Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06520-8106, USA

²National Evolutionary Synthesis Center, 2024 W. Main St. A200, Durham, NC 27705-4667, USA

³Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3AD, UK

Correspondence should be addressed to Jeremy M. Beaulieu, jeremy.beaulieu@yale.edu

Received 4 January 2010; Accepted 17 April 2010

Academic Editor: Jan Suda

Copyright © 2010 Jeremy M. Beaulieu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Broadly sampled phylogenies have uncovered extreme deviations from a molecular clock with the rates of molecular substitution varying dramatically within/among lineages. While growth form, a proxy for life history, is strongly correlated with molecular rate heterogeneity, its influence on trait evolution has yet to be examined. Here, we explore genome size evolution in relation to growth form by combining recent advances in large-scale phylogeny construction with model-based phylogenetic comparative methods. We construct phylogenies for *Monocotyledonae* (monocots) and Fabaceae (legumes), including all species with genome size information, and assess whether rates of genome size evolution depend on growth form. We found that the rates of genome size evolution for woody lineages were consistently an order of magnitude slower than those of herbaceous lineages. Our findings also suggest that growth form constrains genome size evolution, not through consequences associated with the phenotype, but instead through the influence of life history attributes on the tempo of evolution. Consequences associated with life history now extend to genomic evolution and may shed light on the frequently observed threshold effect of genome size variation on higher phenotypic traits.

1. Introduction

The concept of a “molecular clock” predicts that nucleotide substitution rates should scale linearly with time and therefore be equal among lineages. However, rarely do datasets conform to a molecular clock (e.g., [1]) and broadly sampled phylogenies have clearly documented dramatic lineage-specific molecular rate heterogeneity across the Tree of Life (e.g., [2–6]). Life history, or more specifically, generation time, is a strong correlate of among/within lineage rate heterogeneity in both animals and plants [6–8]. In plants, molecular rates are consistently more variable and typically higher in herbaceous species when compared to “woody” (i.e., trees/shrubs) species. Generation time may play a role in this pattern as herbaceous species typically have shorter generation times than woody species, and hence a greater capacity to accumulate nucleotide substitutions per unit time. Implicit in these results is a renewed appreciation for the link between microevolutionary process and macroevolutionary pattern [9].

The consistent pattern of life history influences on rates of molecular evolution across several loci [5, 6, 10] implies this pattern may manifest at the whole genome-level—the first phenotypic scale above molecules. The size of any given genome is determined by rates of DNA accumulation (e.g., retrotransposition and polyploidy) and deletions (e.g., via unequal crossing over and illegitimate recombination). The rate of genome size evolution is therefore set by the interplay between selection and drift promoting and eliminating these mutational changes [11–13]. Indeed, several phylogenetic studies have revealed increases and decreases in genome size [14–17].

Extant angiosperms exhibit a growth form dependent distribution in genome size. Woody angiosperms are characterized by small genome sizes with lower overall variance compared to herbaceous species [18, 19]. This asymmetry in genome size variance among growth forms has been interpreted as an indication of large increases in DNA content negatively impacting woody species [19, 20]. However, when viewed in the context of microevolutionary processes,

the growth form dependent distribution of genome size could also be explained in part by consequences associated with life history. For example, woody angiosperms take many years to reach reproductive maturity [21]. For genome size, this may allow fewer opportunities for insertion/deletions to occur per unit time. Therefore, in terms of generation time, the smaller and lower variance in genome size exhibited by woody species need not be explained only by functional constraints on the phenotype [19, 22].

Here, we test for growth-form dependent rates of genome size evolution between woody and herbaceous lineages. Specifically, we test whether woody species exhibit slower rates of genome size evolution than related herbaceous species. To explore genome size evolution in relation to growth form, we combine recent advances in large-scale phylogeny construction [23] with model-based phylogenetic comparative methods [24]. We focus our analyses on two major branches of the angiosperms that are well represented in the Plant DNA C-value database [25]: the *Monocotyledonae* (monocots; [26]) and the Fabaceae (*Leguminosae* or legumes). The monocots are a large clade of mainly herbaceous angiosperms that also contain a few clades of predominately woody species, including the palms (Arecales; [27]). The legumes are the third largest family of angiosperms and exhibit a wide range of growth habits throughout the clade. It is worth noting that unlike the woody legumes, “woody” monocots do not produce true “wood”. In this context, however, we generally define the tree/shrub or “woody” category as simply large plants with long generation times, for example, [6].

2. Methods and Materials

2.1. Genome Size Data. The amount of DNA in the unrepliated gametic nucleus (i.e., pollen or egg) is referred to as the 1C DNA amount or holoploid genome size, regardless of ploidy level [28]. However, since many angiosperms undergo polyploidy, the monoploid genome size, or 1Cx value, is also often reported and analyzed. The monoploid genome size represents the amount of DNA in the unrepliated monoploid chromosome set and is calculated by dividing the 2C DNA amount by ploidy. Because rates of evolution can be inflated due to polyploidy, we compare and contrast evolutionary rates between the two measures (see below). We compiled genome size estimates for legumes and monocot species where both the 1C amount and the ploidy level were known. Data from the Plant DNA C-values database [25] were combined with additional genome size estimates not yet listed in the database but published in the literature, resulting in an initial list of 1659 and 565 monocot and legume species, respectively, to search GenBank (see below).

2.2. Mega-Phylogeny Construction. We constructed a *mega-phylogeny* of legumes and monocots using the procedures described in [23]. The *mega-phylogeny* method applies orthology tests, sequence saturation analyses, and multiple profile-to-profile alignment methodology to user-specified gene regions. Sequence saturation is detected by calculating the median absolute deviation (MAD) assessed on the

one-dimensional Euclidean distance between the raw and Jukes-Cantor corrected pair-wise sequence distances. For a given gene region, if the most inclusive grouping of these sequences is saturated (MAD > 0.01) then the group is broken up into less inclusive groups using the next level in the NCBI (National Center for Biotechnology Information) taxonomic hierarchy. After every sequence has been placed in an alignment, the individual alignments are then “profile aligned” into a larger alignment. Profile-to-profile alignment combines separate alignments, while preserving the structural elements that are highly conserved between them [29, 30]. We employed a guide tree based on the phylogeny of the NCBI taxonomy to carry out profile alignments.

For the monocots, we specified *atpB*, *matK*, *ndhF*, *rbcl*, *rps16*, *trnL-F*, and ITS as our gene regions of interest. For the legumes we specified *matK*, *psbA-trnH*, *rbcl*, *trnL-F*, ITS, and ETS. However, instead of compiling all possible monocot and legume taxa for a given gene region, we limited our GenBank search to only return sequences for taxa represented in our genome size dataset. The mega-phylogeny matrix construction pipeline was carried out in Python (Ver. 2.5) with the BioPython (Ver. 1.48) module using the BioSQL (Ver. 1.0.1) database schema. Each phylogeny was inferred from the resulting matrix using RAxML (Ver. 7.0.4; [31]), partitioning each gene region and applying a GTRMIX model of rate substitution. For monocots, the maximum likelihood tree was rooted with Acorales (*sensu* [32]) and the legumes were rooted with the tribe Cercideae (*sensu* [33]). In both cases, due to synonymy and errors in Genbank, the trees were further pruned to match our genome size data sets (for a complete list see supplementary materials).

2.3. Time Calibrating the Mega-Phylogeny. We time-calibrated the legume mega-phylogeny using the nonparametric rate smoothing method (NPRS; [34]) with the Powell algorithm in r8s (Ver. 1.71; [35]). The NPRS analysis was restarted three times with different starting values to ensure convergence to a global optimum. We selectively assigned five age constraints from age estimates inferred by Lavin et al. [36]. These included the *Umtzia* crown group (54.0 million years ago, Mya), the *Hologalegina* crown (50.6 Mya), the *Vigna-Phaseolus* split (8.0 Mya), and one assigned to crown Fabaceae (59.0 Mya). We also assigned a constraint within the dalbergioid clade that corresponded to a node in our tree (49.1 Mya).

For the monocots, we selectively assigned eight age constraints using the mean absolute age estimates from Smith et al. [37]. Six age constraints corresponded to the crown age estimates for major clades of monocots (Asparagales, 99.8 Mya; Arecales, 70.9 Mya; Poales, 74.8 Mya; Zingiberales, 88.5 Mya; Commelinales, 76.8 Mya), two corresponded to deep divergences (Liliales + Asparagales, 121.3 Mya; crown Commelinids, 114.9 Mya), and one was assigned to crown monocots (163.5 Mya). We initially used the same procedure to date the monocot tree as above, but the nonparametric rate smoothing analysis did not run to completion. To deal with this problem, we reduced the dataset to 200 tips and reran the NPRS analysis to completion. We obtained the estimated ages for all nodes in the reduced dataset and placed

them in the full dataset. We then used the nonparametric dating method PATHD8 [38] to infer ages for the remaining uncalibrated nodes. PATHD8 uses mean path lengths from the node to tips and deals with substitution rate variation by smoothing rates locally.

2.4. Comparative Analyses. To test for differences in the rate of genome size evolution (1C and 1Cx DNA content) among woody and herbaceous lineages, we compared the fit of single- and two-rate models of Brownian motion evolution. Any phenotypic trait found to accumulate evolutionary change in proportion to time is best described by Brownian motion [39]. The time-independent parameter, σ^2 , or the variance of phenotypic evolution, describes the rate at which this process proceeds. The single-rate model assumes that all analyzed branches accumulate evolutionary changes in genome size at the same rate, σ^2 , while the multiple-rate model assigns a separate rate to each lineage that differs in a particular discrete character state (e.g., $\sigma_{\text{woody}}^2, \sigma_{\text{herb}}^2$). We carried out the single- versus two-rate model comparisons using the “nuncensored” approach in BROWNIE (Ver. 2.1; [24]). Because the “nuncensored” approach assumes the discrete character state of internal branches are known, we used a procedure implemented in BROWNIE that estimates the likeliest growth form state (e.g., woody or herbaceous) across all branches in a given tree based on character codings at the tips. Evaluating the best-fit model between the single- and two- rate models was based on the sample size corrected Akaike Information Criterion (AICc; [40]). The “best” fit model was chosen based on a slightly modified ΔAICc . Because we are only comparing two models, we always calculated ΔAICc as AICc obtained from the single rate model minus the AICc from the two-rate model. A ΔAICc of <2 was taken as evidence for the single-rate model, whereas a ΔAICc >2 indicated considerable evidence for the two-rate model.

We also tested for mean differences in genome size among extant woody and herbaceous species in both our monocot and legume datasets. However, many types of evolutionary processes could have produced the observed trait differences, including Brownian motion. Therefore, we assessed genome size differences among growth form and compared the results of a conventional ANOVA to a null distribution based on ANOVA results obtained from simulations of Brownian motion evolution [41]. This was used to test whether significant species differences between growth forms were larger than would be expected given a random model of Brownian motion evolution. We used the R [42] package GEIGER [43] to generate 1000 Monte Carlo simulations using our input tree topology and time-calibrated branch lengths. We compared the observed F -statistic calculated using an ordinary ANOVA to a null distribution of F -statistics obtained from the Monte Carlo simulations to test for significance. If the observed F -statistic was greater than 95% of the null distribution, then trait differences were greater than expected based on a model of Brownian motion evolution. We carried out this test within each clade separately, using both 1C and 1Cx DNA content.

We \log_{10} transformed the genome size data prior to all analyses to ensure the data minimally conformed to Brownian motion evolution [23, 44]. Under a simple Brownian motion model of evolution (as we employ throughout), a given trait should have an equal probability of increasing or decreasing in the same magnitude given its current state. However, this assumption is inherently violated when traits, such as genome size, are constrained to be non-zero. For example, given a genome size of 0.25 pg, 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, 50%, is likely to occur regardless of the initial genome size at speciation. Thus, it is generally acknowledged that genome size evolution may be better represented as proportional change through an *a priori* \log_{10} transformation [23, 44].

3. Results

3.1. Mega-Phylogeny. Our final matrices for the *Monocotyledonae* (monocots) and Fabaceae (legumes) consisted of 495 and 250 species, respectively. The combined matrix for the legumes comprised 60 woody species from 20 genera and 190 herbaceous species from 21 genera. The woody species were mostly confined to the clades corresponding to the Cercideae, Mimosoideae, and Caesalpinioideae, with additional occurrences found within the Papilionoideae. For monocots, the matrix comprised 213 genera belonging to 9 of the 10 orders of monocots recognized by the Angiosperm Phylogeny Group [27]. Slow growing, tall and/or “woody” genera have been described in several different monocot families, including Arecaceae, (e.g., *Cordyline*, *Dasyllirion*, *Draacaena*, *Nolina*), Bromeliaceae (e.g., *Puya*), Dasypogonaceae (*Dasypogon*, *Kingia*), Pandanaceae (*Pandanus*), Strelitziaceae (e.g., *Ravenala*), Velloziaceae (*Vellozia*), Xanthorrhoeaceae (e.g., *Aloe* and *Xanthorrhoea*), and the woody bamboo genera in the tribe Bambuseae of Poaceae (e.g., *Phyllostachys*, *Sasa*, *Semiarundinaria*). However, due to the absence of genome size and/or sequence data for many of these genera the effect of growth form analyses were restricted to comparisons between (i) *Dasypogon* (Dasypogonaceae; 1 species), the “woody” palms (Arecaceae; 34 species), and the “woody” *Aloe* (Xanthorrhoeaceae; 5 species), (ii) the remaining species which were classified as herbaceous (452 species).

The combined matrix for the monocots contained 10,922 sites and 74.5% gaps or missing sequence, while the legume matrix had a total length of 8221 sites that contained 80.4% gaps or missing sequence. In both cases, the majority of the sequence data came from ITS (Table 1). Additionally, the degree of saturation varied among gene regions, ranging from profiling broad clades (e.g., *rbcL*) to profiling mostly tribes and genera (e.g., ITS; Table 1). Interestingly, of the all genes sampled, only *rbcL* did not require some degree of profile alignment (Table 1). It is worth noting that the degree of saturation was not related to whether or not the gene was protein coding. For example, in both the legume and monocot data set, the noncoding *trnL-F* regions required as much profile aligning as the coding *matK* (Table 1).

TABLE 1: Gene regions specified in the mega-phylogeny construction of *Monocotyledonae* (monocots) and Fabaceae (legumes). The median absolute deviation (MAD) was used to assess sequence saturation and to parse sequences into separate files based on NCBI taxonomy and brought together again using NCBI-based guide tree and profile-to-profile alignment methodology (see Methods and Materials).

Phylogeny	Gene region	Description	<i>N</i>	MAD	Profiles
<i>Monocotyledonae</i>	<i>atpB</i>	Atp synthase beta chain	128	0.006	none
<i>Monocotyledonae</i>	ITS	Internal transcribed spacer 1, 5.8S ribosomal RNA, and internal transcribed spacer 2	596	0.108	mostly to tribe and genus
<i>Monocotyledonae</i>	<i>matK</i>	Maturase K	310	0.068	mostly to family
<i>Monocotyledonae</i>	<i>ndhF</i>	NADH-plastoquinone oxidoreductase	203	0.021	mostly to order
<i>Monocotyledonae</i>	<i>rbcL</i>	Ribulose biphosphate carboxylase	270	0.005	none
<i>Monocotyledonae</i>	<i>rps16</i>	Ribosomal protein S16 intron	88	0.014	mostly to order
<i>Monocotyledonae</i>	<i>trnL-trnF</i>	<i>trnL-trnF</i> intergenic spacer	472	0.033	mostly to family
Fabaceae	ETS	External transcribed spacer and 18S ribosomal RNA	18	0.064	mostly to tribe and genus
Fabaceae	ITS	Internal transcribed spacer 1, 5.8S ribosomal RNA, and internal transcribed spacer 2	293	0.031	mostly to tribe and genus
Fabaceae	<i>matK</i>	Maturase K	119	0.014	mostly to tribe and genus
Fabaceae	<i>psbA-trnH</i>	<i>psbA-trnH</i> intergenic spacer	31	0.005	none
Fabaceae	<i>rbcL</i>	Ribulose biphosphate carboxylase	66	0.001	none
Fabaceae	<i>trnL-trnF</i>	<i>trnL-trnF</i> intergenic spacer	81	0.040	mostly to tribe

MAD scores in bold italics indicate the gene region was saturated across the most inclusive taxonomic-level and broken up into profiles of various taxonomic levels.

N indicates the number of sequences in GenBank returned according to our input search list; however, due to synonymy and errors in GenBank the final tree was pruned to exactly match our genome size data set.

TABLE 2: Parameter estimates from comparisons of single- versus two-rate models of Brownian motion (BM) and applied to both 1C DNA and 1Cx DNA content separately.

Clade	1C DNA content				1Cx DNA content			
	Single-rate	Two-rate		ΔAICc	Single-rate	Two-rate		ΔAICc
	$\sigma^2(\text{My}^{-1})$	$\sigma_{\text{woody}}^2(\text{My}^{-1})$	$\sigma_{\text{herb}}^2(\text{My}^{-1})$		$\sigma^2(\text{My}^{-1})$	$\sigma_{\text{woody}}^2(\text{My}^{-1})$	$\sigma_{\text{herb}}^2(\text{My}^{-1})$	
<i>Monocotyledonae</i>	0.0132	0.0018	0.0142	44.8	0.0055	0.0018	0.0058	17.7
Fabaceae	0.0444	0.0043	0.0552	85.4	0.0338	0.0041	0.0417	74.0

ΔAICc is calculated as the AICc obtained from the single rate model minus the AICc obtained from the two-rate model, where a $\Delta\text{AICc} < 2$ was taken as evidence for the single-rate model, whereas a $\Delta\text{AICc} > 2$ indicated strong evidence for the two-rate model.

3.2. Rates of Genome Size Evolution. In both the monocots and legumes, we found that the genome size data were best fit by a two-rate model of Brownian motion evolution, which inferred a separate rate for woody and herbaceous lineages (Table 2). For legumes, the two-rate model applied to the 1C DNA content was strongly supported ($\Delta\text{AICc} = 85.4$) and woody lineages were inferred to accumulate changes in genome size an order of magnitude slower than related herbaceous lineages. Even when testing 1Cx DNA content, the disparity in rates between woody and herbaceous legumes remained (Table 2). In monocots, the two-rate model was also strongly favored ($\Delta\text{AICc} = 44.8$) with accumulated changes in 1C DNA content occurring at a rate that was also an order of magnitude slower than related herbaceous lineages. Although a significant difference in rates was still detected ($\Delta\text{AICc} = 17.7$), the discrepancy in inferred rates was somewhat reduced between growth forms when testing 1Cx DNA content (Table 2).

Across monocots, there were no significant differences in genome size among woody and herbaceous species ($F_{1,493} = 0.533, P = .904$). For the legumes, mean genome size was significantly smaller in woody species than herbaceous species ($F_{1,248} = 27.5, P < .001$). However, the phylogenetically informed ANOVA suggested that the mean values between woody and herbaceous species of legumes were significantly different, but no more different than would be expected under a model of gradual Brownian motion ($P = .253$). In other words, the observed mean differences among growth form could have arisen by chance alone.

4. Discussion

Our analyses demonstrated that the tempo of genome size evolution is strongly influenced by growth form. In both monocots and legumes, the best fitting model of evolution for genome size inferred a separate rate for each growth form,

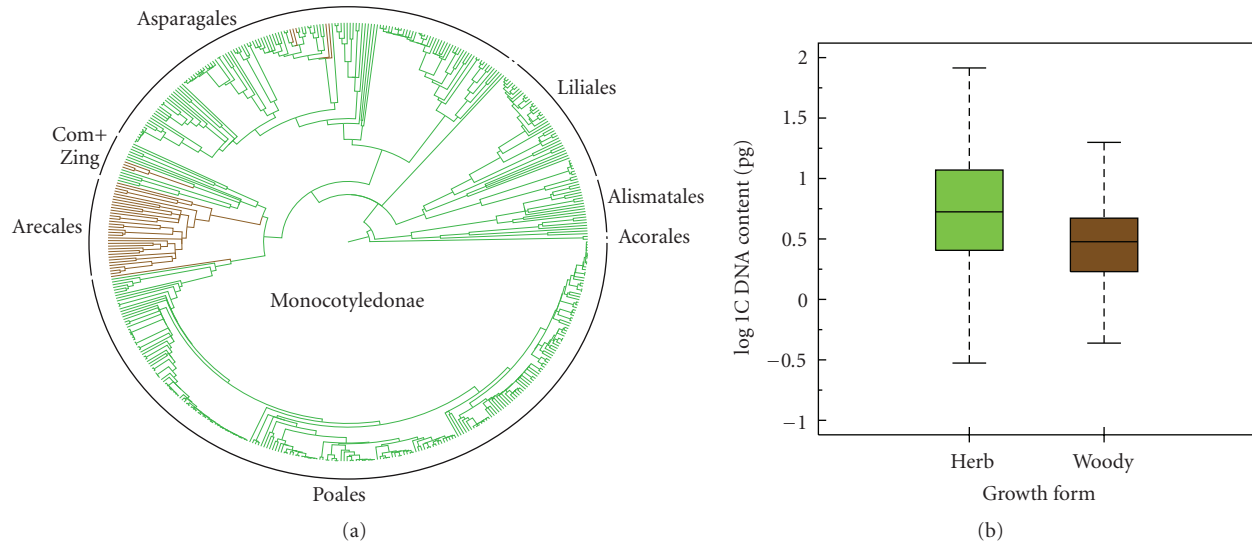


FIGURE 1: (a) Time-calibrated phylogeny of Monocotyledonae (monocots; [26]). Phylogeny is taken from a maximum likelihood analysis of 495 species based on combined analysis *atpB*, ITS, *matK*, *ndhF*, *rbcL*, *rps16*, and *trnL-F*. The major clades of monocots are labeled, and estimates of the likeliest growth form state (woody = brown; herbaceous = green) across all branches in the tree. Com+Zing represents the combined clade of Commelinales and Zingiberales. (b) The distributions of 1C DNA content among growth form, we detect no significant differences among herbaceous (green) and woody (brown) monocots for both 1C DNA and 1Cx DNA content (not shown). The boxplot represents the median (central line), 1st and 3rd quartiles (gray box), and outliers.

with woody lineages accumulating changes in genome size at rates that were consistently an order of magnitude slower than related herbaceous lineages (Table 2). The pattern was consistent across not only two very distinct clades of angiosperms, but also two separate measures of genome size (1C and 1Cx DNA content; Table 2). Therefore, our results suggest that life history alone can impose constraints to the evolution of genome size. These constraints likely reflect the influence of generation time with the longer generation times that characterize woody species [21, 45] providing fewer opportunities for changes in genome sizes to occur per unit time (e.g., [46]).

Plants, unlike animals, do not sequester a germ line early in development, which has the potential for somatic mutations to accumulate throughout growth, particularly for plants with longer generation times. Indeed, there is evidence for greater somatic mutations in longer-lived species compared to annuals on a per generation basis [47–50]. Thus, if an increased number of somatic mutations also involve changes in genome size, then this would complicate any generation time explanation for the observed slower rate of genome size evolution in presumably longer-lived woody species. However, extensive intraindividual and intraspecific variation is not commonly observed for genome size [51–53] suggesting that although there is a potential for a greater number of somatic mutations in longer-lived species to contribute to genome size differences, this may not be a significant factor. The observed excess of new radial cell files in the vascular cambium of trees has been suggested to be one important mechanism for removing somatic mutations from the meristematic population [45, 54]. Likewise,

various plant life cycle characteristics (e.g., pollen tube competition, interovule selection within the same ovary, selective seed/fruit abortion, etc.) have the potential to purge defective genotypes arising from both somatic and gametic mutations without markedly reducing reproductive capacity [47, 55]. Such characteristics may contribute towards explaining the observed reduction in accumulated mutations per unit *time* in woody species despite the potential for more mutations to accumulate on a per generation basis (e.g., [6]).

Additional life history correlates such as effective population size may also play a role, though they are less clearly associated with the observed disparity in rate. Angiosperm trees are reported to have large effective population sizes [45], which would make selection more efficient at removing deleterious mutations and excess DNA [13]. Stronger selection in woody species would be consistent with the suggestion that large increases in DNA content negatively affect woody growth and physiology [19, 20]. However, we found no significant phylogenetic differences in genome size between woody and herbaceous species in either the monocot or legume data sets (Figures 1 and 2), which would be expected if small genome sizes were a requirement for woody species [19]. Moreover, there was no consistent pattern of woody species having smaller genome size in genera consisting of both woody and herbaceous species. For example, within the primarily herbaceous genus *Medicago*, the only woody representative, *M. arborescens*, has a genome size that is nearly twice that of most other species in our dataset. This was also true within the monocot genus *Aloe* (Xanthorrhoeaceae) and is a general observation from

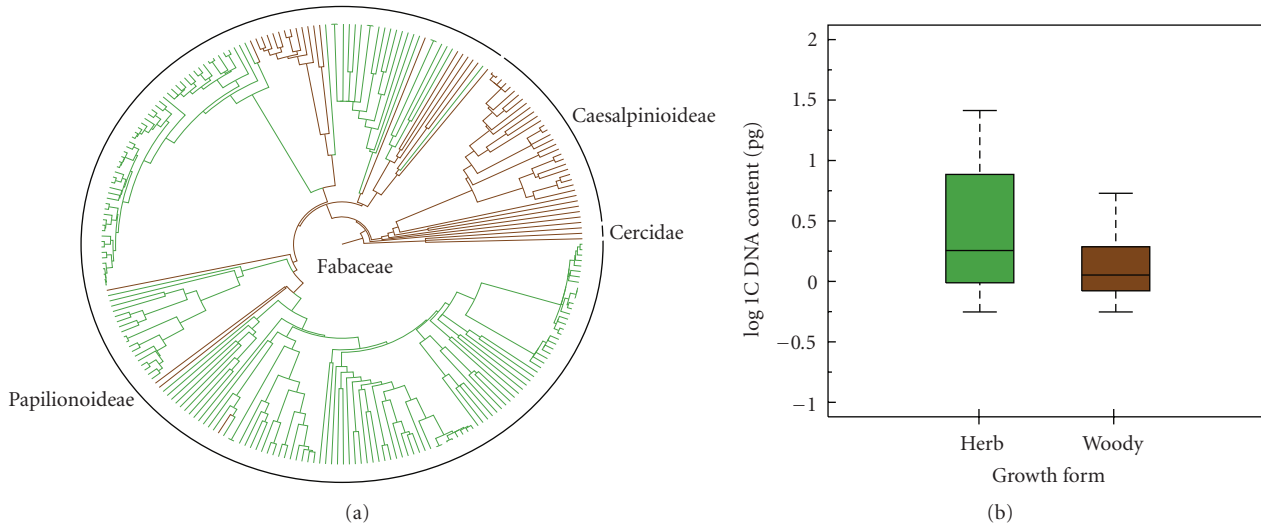


FIGURE 2: (a) Time-calibrated phylogeny of Fabaceae (legumes). Phylogeny is taken from a maximum likelihood analysis of 253 species based on a combined analysis of ETS, ITS, *matK*, *psbA-trnH*, *rbcL*, and *trnL-F*. The major clades of legumes are labeled and estimates of the likeliest growth form state (woody = brown; herbaceous = green) across all branches in the tree. (b) The distributions of 1C DNA content among growth form, where we find significant differences among herbaceous (green) and woody (brown) legumes for both 1C DNA and 1Cx DNA content (not shown), but no more different than could arise by chance (see Methods and Materials). The boxplot represents the median (central line), 1st and 3rd quartiles (gray box), and outliers.

angiosperm genera not included in this study (see [18]). In addition, it is clear that not all woody plants possess small genome sizes, as the completely woody *Acrogyminospermae* (a clade containing the four major lineages of extant “gymnosperms”; [26]) are characterized by much larger genomes that are 12 times the modal value of angiosperms [56]. Nonetheless, the influence of selection and generation time may not be mutually exclusive, but assessing a potential asymmetry in selection due to growth form will require developing models of phenotypic evolution that allow decoupling of the strength of selection (e.g., Ornstein-Uhlenbeck model; [57, 58]) across discrete character states.

Small genome sizes are consistently associated with a large range of phenotypic variation that decreases with increasing genome size. This pattern has been documented for a suite of traits, including climate tolerance [59], leaf mass per unit area (LMA; [60]), maximum height [61], and seed mass [62]. For example, very large genome sizes do not produce small-sized seeds and species with small genome sizes exhibit a range of seed sizes [62]. While genome size may set the minimum seed mass due to size constraints at the cellular level (e.g., large genomes are not contained within small cells; [19]), it remains unclear why the largest seeds are not associated with large genomes. Perhaps, the observed upper constraint does not relate to genome size at all, but instead reflects the constraint imposed on genome size evolution by generation time. Trees and shrubs produce large seeds in comparison to herbaceous species [63, 64], suggesting that the preponderance of seed mass variation at smaller genome sizes may simply reflect a diversity of growth form. Because “woodiness” confers

a marked reduction in the rate of genome size evolution, the decreasing phenotypic variation with increasing genome size may simply be a function of insufficient time having elapsed for woody angiosperm to evolve large genome sizes.

Further analyses should focus on large-scale comparisons of growth form dependent rates of genome size evolution in order to uncover a generality. In addition, more focused studies on specific life forms such as succulents, parasites, geophytes may also help to resolve and refine the interplay and influence of growth form on rates of genome size evolution. Such approaches require increased sampling efforts of genome size estimates for species across a broad range of taxonomic groups. Nevertheless, our results of monocots and legumes suggest that, in addition to molecular substitution rates [2, 5, 6], growth form can also influence the tempo of genome size evolution. Therefore, given the consistency across two scales—molecules and genomes—a logical next step is to examine higher phenotypic traits in relation to growth form. Only through combining trait databases (e.g., Glopnet [65]; SID [66], etc.) with the construction of broadly sampled phylogenies (e.g., [23]) will interesting life history trends continue to be uncovered.

Supplementary Materials

Tables 1 and 2 in the supplementary material list all species and associated growth form designation included in the combined mega-phylogeny matrix of monocots and legumes; see tables 1 and 2 in the supplementary material available online at doi: 10.1155/2010/989152.

References

- [1] M. J. Sanderson, M. F. Wojciechowski, J.-M. Hu, T. S. Khan, and S. G. Brady, "Error, bias, and long-branch attraction in data for two chloroplast photosystem genes in seed plants," *Molecular Biology and Evolution*, vol. 17, no. 5, pp. 782–797, 2000.
- [2] B. S. Gaut, S. V. Muse, W. D. Clark, and M. T. Clegg, "Relative rates of nucleotide substitution at the *rbcL* locus of monocotyledonous plants," *Journal of Molecular Evolution*, vol. 35, no. 4, pp. 292–303, 1992.
- [3] A. P. Martin and S. R. Palumbi, "Body size, metabolic rate, generation time, and the molecular clock," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 90, no. 9, pp. 4087–4091, 1993.
- [4] A. O. Mooers and P. H. Harvey, "Metabolic rate, generation time, and the rate of molecular evolution in birds," *Molecular Phylogenetics and Evolution*, vol. 3, no. 4, pp. 344–350, 1994.
- [5] B. S. Gaut, B. R. Morton, B. C. Mccaig, and M. T. Clegg, "Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcL*," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 93, no. 19, pp. 10274–10279, 1996.
- [6] S. A. Smith and M. J. Donoghue, "Rates of molecular evolution are linked to life history in flowering plants," *Science*, vol. 322, no. 5898, pp. 86–89, 2008.
- [7] M. S. Hafner, P. D. Sudman, F. X. Villablanca, T. A. Spradling, J. W. Demastes, and S. A. Nadler, "Disparate rates of molecular evolution in cospeciating hosts and parasites," *Science*, vol. 265, no. 5175, pp. 1087–1090, 1994.
- [8] L. Bromham, A. Rambaut, and P. H. Harvey, "Determinants of rate variation in mammalian DNA sequence evolution," *Journal of Molecular Evolution*, vol. 43, no. 6, pp. 610–621, 1996.
- [9] B. Charlesworth, R. Lande, and S. Montgomery, "A neo-Darwinian commentary on macroevolution," *Evolution*, vol. 36, pp. 474–498, 1982.
- [10] B. S. Gaut, L. G. Clark, J. F. Wendel, and S. V. Muse, "Comparisons of the molecular evolutionary process at *rbcL* and *ndhF* in the grass family (Poaceae)," *Molecular Biology and Evolution*, vol. 14, no. 7, pp. 769–777, 1997.
- [11] D. A. Petrov, T. A. Sangster, J. S. Johnston, D. L. Hartl, and K. L. Shaw, "Evidence for DNA loss as a determinant of genome size," *Science*, vol. 287, no. 5455, pp. 1060–1062, 2000.
- [12] M. Lynch and J. S. Conery, "The origins of genome complexity," *Science*, vol. 302, no. 5649, pp. 1401–1404, 2003.
- [13] M. Lynch, *The Origins of Genome Architecture*, Sinauer Associates, Sunderland, Mass, USA, 2007.
- [14] I. J. Leitch, M. W. Chase, and M. D. Bennett, "Phylogenetic analysis of DNA C-values provides evidence for a small ancestral genome size in flowering plants," *Annals of Botany*, vol. 82, pp. 85–94, 1998.
- [15] D. E. Soltis, P. S. Soltis, M. D. Bennett, and I. J. Leitch, "Evolution of genome size in the angiosperms," *American Journal of Botany*, vol. 90, no. 11, pp. 1596–1603, 2003.
- [16] I. J. Leitch, J. M. Beaulieu, K. Cheung, L. Hanson, M. A. Lysak, and M. F. Fay, "Punctuated genome size evolution in Liliaceae," *Journal of Evolutionary Biology*, vol. 20, no. 6, pp. 2296–2308, 2007.
- [17] M. A. Lysak, M. A. Koch, J. M. Beaulieu, A. Meister, and I. J. Leitch, "The dynamic ups and downs of genome size evolution in Brassicaceae," *Molecular Biology and Evolution*, vol. 26, no. 1, pp. 85–98, 2009.
- [18] D. Ohri, "Climate and growth form: the consequences for genome size in plants," *Plant Biology*, vol. 7, no. 5, pp. 449–458, 2005.
- [19] J. M. Beaulieu, I. J. Leitch, S. Patel, A. Pendharkar, and C. A. Knight, "Genome size is a strong predictor of cell size and stomatal density in angiosperms," *New Phytologist*, vol. 179, no. 4, pp. 975–986, 2008.
- [20] G. L. Stebbins, "Cytological characteristics associated with different growth habits in dicotyledons," *American Journal of Botany*, vol. 25, pp. 189–198, 1938.
- [21] M. Verdú, "Age at maturity and diversification in woody angiosperms," *Evolution*, vol. 56, no. 7, pp. 1352–1361, 2002.
- [22] C. A. Knight, N. A. Molinari, and D. A. Petrov, "The large genome constraint hypothesis: evolution, ecology and phenotype," *Annals of Botany*, vol. 95, no. 1, pp. 177–190, 2005.
- [23] S. A. Smith, J. M. Beaulieu, and M. J. Donoghue, "Mega-phylogeny approach for comparative biology: an alternative to supertree and supermatrix approaches," *BMC Evolutionary Biology*, vol. 9, no. 1, article 37, pp. 1–12, 2009.
- [24] B. C. O'Meara, A. Cécile, M. J. Sanderson, and P. C. Wainwright, "Testing for different rates of continuous trait evolution using likelihood," *Evolution*, vol. 60, no. 5, pp. 922–923, 2006.
- [25] M. D. Bennett and I. J. Leitch, "Plant DNA C-values database," October 2005, <http://data.kew.org/cvalues/>.
- [26] P. D. Cantino, J. A. Doyle, S. W. Graham et al., "Towards a phylogenetic nomenclature of Tracheophyta," *Taxon*, vol. 56, no. 3, pp. 822–846, 2007.
- [27] B. Bremer, K. Bremer, M. W. Chase et al., "An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II," *Botanical Journal of the Linnean Society*, vol. 141, no. 4, pp. 399–436, 2003.
- [28] J. Greilhuber, J. Doležal, M. A. Lysák, and M. D. Bennett, "The origin, evolution and proposed stabilization of the terms "genome size" and "C-value" to describe nuclear DNA contents," *Annals of Botany*, vol. 95, no. 1, pp. 255–260, 2005.
- [29] N. von Ohlsen, I. Sommer, and R. Zimmer, "Profile-profile alignment: a powerful tool for protein structure prediction," *Pacific Symposium on Biocomputing*, pp. 252–263, 2003.
- [30] R. C. Edgar, "MUSCLE: multiple sequence alignment with high accuracy and high throughput," *Nucleic Acids Research*, vol. 32, no. 5, pp. 1792–1797, 2004.
- [31] A. Stamatakis, "RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models," *Bioinformatics*, vol. 22, no. 21, pp. 2688–2690, 2006.
- [32] M. W. Chase, M. F. Fay, D. S. Devey, et al., "Multigene analyses of monocot relationships: a summary," *Aliso*, vol. 22, pp. 63–75, 2006.
- [33] M. F. Wojciechowski, M. Lavin, and M. J. Sanderson, "A phylogeny of legumes (Leguminosae) based on analysis of the plastid *matK* gene resolves many well-supported subclades within the family," *American Journal of Botany*, vol. 91, no. 11, pp. 1846–1862, 2004.
- [34] M. J. Sanderson, "A nonparametric approach to estimating divergence times in the absence of rate constancy," *Molecular Biology and Evolution*, vol. 14, no. 12, pp. 1218–1231, 1997.
- [35] M. J. Sanderson, "r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock," *Bioinformatics*, vol. 19, no. 2, pp. 301–302, 2003.
- [36] M. Lavin, P. S. Herendeen, and M. F. Wojciechowski, "Evolutionary rates analysis of leguminosae implicates a rapid diversification of lineages during the tertiary," *Systematic Biology*, vol. 54, no. 4, pp. 575–594, 2005.

- [37] S. A. Smith, J. M. Beaulieu, and M. J. Donoghue, "An uncorrelated relaxed-clock analysis suggests an earlier origin for flowering plants," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 13, pp. 5897–5902, 2010.
- [38] T. Britton, C. L. Anderson, D. Jacquet, S. Lundqvist, and K. Bremer, "Estimating divergence times in large phylogenetic trees," *Systematic Biology*, vol. 56, no. 5, pp. 741–752, 2007.
- [39] J. Felsenstein, "Phylogenies and the comparative method," *American Naturalist*, vol. 125, no. 1, pp. 1–15, 1985.
- [40] K. P. Burham and D. R. Anderson, *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, Springer, New York, NY, USA, 2002.
- [41] T. Garland Jr., A. W. Dickerman, C. M. Janis, and J. A. Jones, "Phylogenetic analysis of covariance by computer simulation," *Systematic Biology*, vol. 42, no. 3, pp. 265–292, 1993.
- [42] R Development Core Team, "R: a language and environment for statistical computing," R Foundation for Statistical Computing, Vienna, Austria, 2008, <http://www.R-project.org/>.
- [43] L. J. Harmon, J. T. Weir, C. D. Brock, R. E. Glor, and W. Challenger, "GEIGER: investigating evolutionary radiations," *Bioinformatics*, vol. 24, no. 1, pp. 129–131, 2008.
- [44] M. J. Oliver, D. Petrov, D. Ackerly, P. Falkowski, and O. M. Schofield, "The mode and tempo of genome size evolution in eukaryotes," *Genome Research*, vol. 17, no. 5, pp. 594–601, 2007.
- [45] R. J. Petit and A. Hampe, "Some evolutionary consequences of being a tree," *Annual Review of Ecology, Evolution, and Systematics*, vol. 37, pp. 187–214, 2006.
- [46] E. W. Sinnott, "Comparative rapidity of evolution in various plant types," *American Naturalist*, vol. 50, pp. 466–478, 1916.
- [47] E. J. Klekowski Jr. and P. J. Godfrey, "Ageing and mutation in plants," *Nature*, vol. 340, no. 6232, pp. 389–391, 1989.
- [48] E. J. Klekowski, "Mutation rates in mangroves and other plants," *Genetica*, vol. 102–103, pp. 325–331, 1998.
- [49] S. M. Udupa and M. Baum, "High mutation rate and mutational bias at (TAA)_n microsatellite loci in chickpea (*Cicer arietinum* L.)," *Molecular Genetics and Genomics*, vol. 265, no. 6, pp. 1097–1103, 2001.
- [50] D. G. Scofield and S. T. Schultz, "Mitosis, stature and evolution of plant mating systems: low- Φ and high- Φ plants," *Proceedings of the Royal Society B*, vol. 273, no. 1584, pp. 275–282, 2006.
- [51] J. Greilhuber, "Intraspecific variation in genome size in angiosperms: identifying its existence," *Annals of Botany*, vol. 95, no. 1, pp. 91–98, 2005.
- [52] J. Greilhuber, "Cytochemistry and C-values: the less-well-known world of nuclear DNA amounts," *Annals of Botany*, vol. 101, no. 6, pp. 791–804, 2008.
- [53] J. Doležel and J. Bartoš, "Plant DNA flow cytometry and estimation of nuclear genome size," *Annals of Botany*, vol. 95, no. 1, pp. 99–110, 2005.
- [54] E. J. Mellerowicz, M. Baucher, B. Sundberg, and W. Boerjan, "Unravelling cell wall formation in the woody dicot stem," *Plant Molecular Biology*, vol. 47, no. 1–2, pp. 239–274, 2001.
- [55] E. J. Klekowski, N. Kazarinova-Fukshansky, and L. Fukshansky, "Shoot apical meristems and mutation—stratified meristems and angiosperm evolution," *American Journal of Botany*, vol. 72, pp. 1788–1800, 1985.
- [56] I. J. Leitch, D. E. Soltis, P. S. Soltis, and M. D. Bennett, "Evolution of DNA amounts across land plants (Embryophyta)," *Annals of Botany*, vol. 95, no. 1, pp. 207–217, 2005.
- [57] T. F. Hansen, "Stabilizing selection and the comparative analysis of adaptation," *Evolution*, vol. 51, no. 5, pp. 1341–1351, 1997.
- [58] M. A. Butler and A. A. King, "Phylogenetic comparative analysis: a modeling approach for adaptive evolution," *American Naturalist*, vol. 164, no. 6, pp. 683–695, 2004.
- [59] C. A. Knight and D. D. Ackerly, "Variation in nuclear DNA content across environmental gradients: a quantile regression analysis," *Ecology Letters*, vol. 5, no. 1, pp. 66–76, 2002.
- [60] J. M. Beaulieu, I. J. Leitch, and C. A. Knight, "Genome size evolution in relation to leaf strategy and metabolic rates revisited," *Annals of Botany*, vol. 99, no. 3, pp. 495–505, 2007.
- [61] C. A. Knight and J. M. Beaulieu, "Genome size scaling through phenotype space," *Annals of Botany*, vol. 101, no. 6, pp. 759–766, 2008.
- [62] J. M. Beaulieu, A. T. Moles, I. J. Leitch, M. D. Bennett, J. B. Dickie, and C. A. Knight, "Correlated evolution of genome size and seed mass," *New Phytologist*, vol. 173, no. 2, pp. 422–437, 2007.
- [63] P. J. Grubb, D. A. Coomes, D. J. Metcalfe et al., "Comment on 'a brief history of seed size'," *Science*, vol. 310, no. 5749, p. 783, 2005.
- [64] A. T. Moles, D. D. Ackerly, C. O. Webb et al., "Factors that shape seed mass evolution," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 30, pp. 10540–10544, 2005.
- [65] I. J. Wright, P. B. Reich, M. Westoby et al., "The worldwide leaf economics spectrum," *Nature*, vol. 428, no. 6985, pp. 821–827, 2004.
- [66] S. Flynn, R. M. Turner, and J. B. Dickie, "Seed Information Database," October 2004, <http://data.kew.org/sid/sidsearch.html>.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

