doi:10.1111/j.1558-5646.2010.00967.x

A ROLE FOR NONADAPTIVE PROCESSES IN PLANT GENOME SIZE EVOLUTION?

Kenneth D. Whitney,[1,2] Eric J. Baack,[3] James L. Hamrick,[4] Mary Jo W. Godt,[4] Brian C. Barringer,[5] Michael D. Bennett,[6] Christopher G. Eckert,[7] Carol Goodwillie,[8] Susan Kalisz,[9] Ilia J. Leitch,[6] and Jeffrey Ross-Ibarra[10]

[1]Department of Ecology and Evolutionary Biology, Rice University, 6100 Main St., Houston, Texas 77005

[2]E-mail: kwhitney@rice.edu

[3]Department of Biology, Luther College, Decorah, Iowa 52101

[4]Department of Plant Biology, University of Georgia, Athens, Georgia 30602

[5]Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, New York 14853

[6]Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3DS, United Kingdom

[7]Department of Biology, Queen's University, Kingston, Ontario K7L 3N6, Canada

[8]Department of Biology, East Carolina University, Greenville, North Carolina 27858

[9]Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

[10]Department of Plant Sciences, University of California, Davis, California 95616

Received August 22, 2009
Accepted January 6, 2010

Genome sizes vary widely among species, but comprehensive explanations for the emergence of this variation have not been validated. Lynch and Conery (2003) hypothesized that genome expansion is maladaptive, and that lineages with small effective population size (N_e) evolve larger genomes than those with large N_e as a consequence of the lowered efficacy of natural selection in small populations. In addition, mating systems likely affect genome size evolution via effects on both N_e and the spread of transposable elements (TEs). We present a comparative analysis of the effects of N_e and mating system on genome size evolution in seed plants. The dataset includes 205 species with monoploid genome size estimates (corrected for recent polyploidy) ranging from 2Cx = 0.3 to 65.9 pg. The raw data exhibited a strong positive relationship between outcrossing and genome size, a negative relationship between N_e and genome size, but no detectable N_e × outcrossing interaction. In contrast, phylogenetically independent contrast analyses found only a weak relationship between outcrossing and genome size and no relationship between N_e and genome size. Thus, seed plants do not support the Lynch and Conery mechanism of genome size evolution. Further work is needed to disentangle contrasting effects of mating systems on the efficacy of selection and TE transmission.

KEY WORDS: **Effective population size, genetic diversity, genetic drift, N_e, outcrossing rate, phylogenetically independent contrasts, plant mating systems, selfing, transposable elements.**

Across the web of life, genome size (measured either in base pairs or mass, where 10^9 bp ≈ 1 picogram) spans several orders of magnitude. Even within multicellular lineages, variation is substantial: genome size estimates (2C values) range from 0.04 to 265.6 pg for animals (Gregory 2005a); from 0.014 to 1.62 pg for fungi (Kullman et al. 2005); and from 0.13 to 254.8 pg for seed plants (Bennett and Leitch 2005b; Greilhuber et al. 2006). Attempts to explain this variation have spanned several decades (e.g., Price 1976) and have been the subject of recent comprehensive reviews (Gregory 2005b; Lynch 2007).

Proximal mechanisms for both increases and decreases in DNA content are known (reviewed in Hawkins et al. 2008a). Increases can result from polyploidy, transposable element (TE) proliferation, intron proliferation, segmental duplications, and small-scale insertions, whereas decreases can occur via deletions (via, e.g., illegitimate recombination, unequal crossing over, and DNA replication errors) and chromosome loss (Petrov 2001; Lynch 2007). For example, Zuccolo et al. (2007) showed that long-terminal repeat TEs (in addition to polyploidy) are responsible for much of the genome size variation among 23 species of *Oryza* (rice and relatives); similar patterns were seen in *Gossypium* (Hawkins et al. 2008b). However, why the balance between genome reduction and expansion has resulted in such wide variation in genome size among species and higher taxonomic groups is still an open question. Recently, there have been calls for phylogenetic comparative analyses of genome size across diverse taxa (Charlesworth and Barton 2004; Flowers and Purugganan 2008).

Hypotheses for genome expansion fall into adaptive, neutral, and maladaptive classes. Early explanations focused on the potential adaptive significance of increases in DNA mass: for example, the nucleotypic hypothesis posits that genome size expansion could trigger larger nucleus size, which in turn would lead to larger cell size (Bennett and Leitch 2005a). External environmental factors favoring larger cells would favor larger nuclei, and thus drive increased DNA content. In contrast, a few authors have argued that the process of genome size evolution might be neutral, or nearly so (Petrov 2002). Under this scenario, lower bounds for genome sizes exist because of the minimum genetic complexity necessary for organism functioning. Above this lower bound, the distribution of genome sizes is random, and large genomes are rare by chance, not by selection against them. Neutral hypotheses can rarely be tested directly, but are supported if nonadaptive processes can be identified that produce the observed pattern. An examination of genome sizes across eukaryotes in a phylogenetic context found that rates of genome size change are proportional to genome size (Oliver et al. 2007). Under proportionality, it is hard for small genomes to become large and stay large, but easier for large genomes to become small and stay small. Thus, organisms with large genomes should be rare even if there are no negative fitness consequences of increased genome size (Oliver et al. 2007).

Finally, some argue that larger genomes are maladaptive; for example, large genomes may constrain rates of cell division and thus growth (Bennett & Leitch 2005a). Lynch and Conery (2003) posit that genome expansion generally imposes a fitness cost, and that lineages differ in effective population size (N_e) and, as a result, differ in the efficacy with which natural selection will counteract that expansion. In particular, they cite a pattern in which taxa with large expected N_e (e.g., microbes) tend to have far smaller genomes than taxa with presumably smaller N_e (e.g., multicellular eukaryotes) as support for a central role of genetic drift in genome size evolution. To date, comparative empirical evidence on the Lynch and Conery hypothesis is mixed. First, rare or geographically restricted plant taxa can have larger genomes than more common congeners (Vinogradov 2003), as expected, but other studies have found no such relationship (Grotkopp et al. 2004). Second, the hypothesis has been supported by an analysis of 33 species of ray-finned fish; using microsatellite heterozygosity as a proxy for N_e, Yi and Streelman (2005) detected a negative relationship between putative N_e and genome size. However, this analysis has been challenged as artifactual (Gregory and Witt 2008; see Discussion). Third, at the intraspecific scale, a comparative analysis of *Arabidopsis lyrata* populations found evidence that TEs are removed by purifying selection in a large refugial population, whereas TE copy number appears to be evolving neutrally in smaller populations (Lockton et al. 2008).

Mating systems could also influence genome size trajectories, via potentially contrasting effects on N_e and on TE proliferation rates. On one hand, selfing species should have smaller N_e than outcrossing species (Nordborg 2000; but note that there is extensive variation, see Schoen and Brown 1991), potentially leading to larger genomes via reduced efficacy of selection (Charlesworth and Wright 2001; Lynch and Conery 2003). On the other hand, mobile genetic elements can function analogously to sexually transmitted diseases, in that a given lineage may "catch" a rapidly spreading TE via outcrossing (Wright and Schoen 1999; Arkhipova and Meselson 2000; Morgan 2001). By this reasoning, predominantly selfing lineages should be less likely to incorporate novel TEs and should experience genome expansion via TE proliferation less frequently than outcrossing lineages. Selfing lineages would be expected to have smaller genome sizes due to their isolation from contagious TEs, with no requirement that small genomes have a direct selective benefit.

Several previous studies of vascular plants have concluded that selfing species have smaller genomes than outcrossers (Govindaraju and Cullis 1991; Albach and Greilhuber 2004; Wright et al. 2008). However, inconsistent application of phylogenetic corrections and restricted scope make generalizations from these studies difficult. Govindaraju and Cullis (1991) examined 176 species, but only scored mating system qualitatively, made no adjustments for recent polyploidy, and did not employ phylogenetic corrections. Wright et al. (2008) compared the monoploid genome size of 14 self-pollinating taxa with paired outcrossing congeners in nine genera. Albach and Greilhuber (2004) examined 42 species in the Veroniceae and corrected for both polyploidy and phylogeny. However, it is unclear whether the same factors that affect genome size in their relatively restricted sample (monoploid genome size varied only from 2Cx = 0.62 to 2.7 pg) also explain larger scales of genome size variation.

Here, we assemble the largest dataset to date to examine evidence for effects of N_e and mating system on genome size

evolution in seed plants. We take a comparative approach and compile data relating expected heterozygosity (a proxy for N_e, see Methods) and mating system across 205 plant species whose monoploid genome sizes (2C divided by ploidy; see Greilhuber et al. 2005) range from 2Cx = 0.3 to 65.9 pg. We examine the relationships between genome size, mating system, and N_e using general linear models, first on the raw data and then on phylogenetically independent contrasts (PICs). Our analysis differs from prior mating system/genome size studies because we (1) use multilocus outcrossing rate data in addition to categorical breeding system classifications, (2) correct for phylogenetic relatedness via independent contrasts, and (3) simultaneously examine the effects of N_e.

Methods

GENOME SIZE, MATING SYSTEM, AND EFFECTIVE POPULATION SIZE DATA

Following Leitch and Bennett (2004), we measured genome size as the DNA content of the monoploid chromosome set; e.g., a tetraploid with 2C = 8 pg has a monoploid genome size of 2Cx = 4 pg. In this way, the signal of recent polyploidy was removed from genome size estimates to increase the power to detect an effect of N_e and/or mating system. For comparison, analyses using 2C genome sizes were also run. All 2C values and ploidy levels were derived from the Kew Plant DNA C-values Database (Bennett and Leitch 2005b). For species with multiple ploidy levels in the database, the 2C value used corresponded to the ploidy of the material used in the mating system and H_{es} datasets (see below), as determined from the original source publications. Unclear cases were excluded from the analyses.

Two mating system datasets were compiled, an "Outcrossing Rate Dataset" and an "Outcrossing Index Dataset." The former was based on multilocus outcrossing rate (t) data compiled from the literature by Goodwillie et al. (2005) and Barringer (2007); t varies continuously from 0 to 1 reflecting fully selfing to fully outcrossing, respectively (Goodwillie et al. 2005). After dropping species lacking genome size and/or heterozygosity information, this dataset contained 58 species. The outcrossing index dataset placed species into three ordered mating system categories: 1—selfing, 2—mixed, and 3—outcrossing. Classifications were based on Cruden (1977), Hamrick et al. (1979), Govindaraju (1988), Govindaraju and Cullis (1991), and Fryxell (1957). In general, these sources used data on corolla size and shape, spatial separation of anther and stigma, temporal separation of male and female function, and seed set after self and outcross pollination to classify species as selfing, possessing a mixed system, or outcrossing. Although admittedly a cruder approach, the index is available for a much larger number of species. Additional species from the outcrossing rate dataset were then added by converting outcrossing rates to index values, using the conventions that $t > 0.8$ indicates outcrossing species (index = 3), $0.8 \geq t \leq 0.2$ indicates species with mixed mating systems (index = 2), and $t < 0.2$ indicates selfing species (index = 1; Goodwillie et al. 2005). After dropping species lacking genome size and/or heterozygosity information, the outcrossing index dataset contained 205 species. The dataset comprised 38 gymnosperms (in three families) and 167 angiosperms (comprising 98 eudicots, 68 monocots, and one magnoliid in a total of 25 families).

Species-wide expected heterozygosities (H_{es}, the proportion of individuals expected to be heterozygous if Hardy–Weinberg equilibrium frequencies apply) were taken from a database of allozyme genetic diversities compiled from the literature. This database was the basis of earlier analyses of life-history traits and genetic diversity in seed plants (Hamrick and Godt 1989, 1996) and has since been updated to include 776 H_{es} values for 600 species. For each species, H_{es} values originating from different studies were averaged. Based on Ohta and Kimura (1973), we calculated N_e as follows:

$$N_e = ((1 - H_{es})^{-2} - 1)/(8u),$$

where the mutation rate u is estimated at 10^{-5} electrophoretic changes per locus per generation (Drake et al. 1998). Note that a different estimate for u would change absolute values in the analysis, but not relative values, and so should have no effect on the conclusions. Estimates of N_e for all species in our analyses are given in Appendix S1.

Two issues related to the heterozygosity data deserve mention. First, we chose H_{es} (expected species-wide heterozygosity) over H_{ep} (expected population heterozygosity) because the former better represents the long-term heterozygosity experienced by a lineage, and thus should be most relevant for evolution of genome size. In any case, substitution of H_{ep} for H_{es} had no qualitative effect on the outcome (see Results). Second, we recognize that allozyme variation (and any other nonneutral variation) has limitations when used to estimate absolute values of N_e (see Discussion). However, in the absence of neutral sequence data for large numbers of species, allozymes represent the best available data for large comparative analyses requiring relative values of H_e or N_e.

TESTS OF GENOME SIZE VS. MATING SYSTEM AND N_e

We examined the relationship between genome size and mating system, N_e, and the mating system × N_e interaction for both datasets (outcrossing rate and outcrossing index) using general linear models. Both N_e and mating system were centered (to mean = 0) prior to the calculation of the interaction term to reduce potential multicollinearity between predictor variables

(Aiken and West 2001). All analyses used SAS proc REG (SAS Institute 2003). Outliers were detected using the RSTUDENT and DFFITS options with values > 2 signifying influential outliers (Belsley et al. 1980). No outliers were detected in the analyses of the raw data, but one outlier was detected and excluded in the PIC analysis of the outcrossing index dataset (below). Because residuals were nonnormal in preliminary analyses, significance levels were assessed using a randomization procedure (Cassell 2002) with 10,000 replicates. Partial regression plots (Neter et al. 1996) were used to examine the relationship of an individual predictor variable to genome size by controlling for the effects of the other predictor variable and the interaction term; slopes in these plots are equal to the parameter estimates (b′) in the full model.

SUPERTREE CONSTRUCTION AND PICS

To account for the phylogenetic nonindependence of our observations, we revisited the analyses (previous section) using PICs (Felsenstein 1985). We employed Phylomatic (Webb and Donoghue 2005) using the Davies et al. (2004) angiosperm supertree to construct base family-level phylogenies. Phylogenies were imported to Mesquite version 2.71 (Maddison and Maddison 2009), where gymnosperm families were added by hand, based on their position in the Phylomatic maximally resolved seed plant tree R20070607 (Webb and Donoghue 2005; accessed June, 2007). Resolution at the genus and species level was then added by hand based on a large number of sources (Appendix S2). Phylogenies for the rate and index datasets are presented in Figures 1 and 2, respectively, with genome sizes traced onto the phylogenies using the parsimony ancestral states method of Mesquite (Maddison and Maddison 2009).

We then examined PICs (Felsenstein 1985). As a first step, we confirmed that the traits exhibit phylogenetic signal by calculating λ in BayesTraits (Pagel and Meade 2009). PICs for mating systems, N_e, and genome size were then generated using the PDAP:PDTREE module in Mesquite (Garland et al. 1993; Midford et al. 2002). Actual branch lengths are unknown, but diagnostics indicated that branch lengths conforming to the Nee model were adequate in all analyses. Although some polytomies are present in the phylogenies, theoretical work has shown that soft polytomies do not bias PIC analyses (Garland and Diaz-Uriarte 1999). Standardized contrasts were obtained by dividing the raw contrasts by their standard deviations (Garland et al. 1992). Then, as above, regressions compared contrasts in mating systems and N_e to contrasts in genome size. We used SAS proc REG (SAS Institute 2003), constraining the regressions through the origin (Garland et al. 1992). Sample size (N, number of species) was smaller in the outcrossing index dataset PIC analysis because phylogenetic information was not available for some species. For both the raw and PIC analyses, log-transformation of the variables did not qualitatively change the patterns (results not shown).

Results

Analyses of the raw (phylogenetically uncorrected) data indicated significant relationships between plant mating system and genome size in both datasets (Table 1). In these analyses, plants with higher rates of outcrossing (whether measured by t or the outcrossing index) had larger genomes (Figs. 3A and 4A). There was a significant negative relationship between N_e and genome size in the raw outcrossing index dataset (Fig. 5A), but no relationship in the raw outcrossing rate dataset. Mating system and heterozygosity were weakly correlated ($r = 0.26$) and did not interact in their effects (Table 1).

Traits exhibited strong phylogenetic signal. For the outcrossing rate and index datasets, $\lambda = 0.89$ and 0.81, respectively, and differed from 0.0 in both cases ($P < 0.001$); furthermore λ for the rate dataset did not differ from 1.0 ($P > 0.1$). These results indicate phylogenetic dependence (Pagel 1999; Freckleton et al. 2002). Once phylogeny was incorporated into the regression analyses, explanatory power of the models dropped substantially, from $r^2 = 0.22$–0.46 (raw datasets) to 0.01–0.06 (PIC datasets; Table 1). A single influential outlier, the contrast between *Heuchera grossulariifolia* ($N_e = 9722$, genome size = 1.04 pg) and *Paeonia californica* ($N_e = 0$, genome size = 33.5 pg), was detected and removed. The relationship between plant mating system and genome size largely disappeared from both datasets after phylogenetic correction (Table 1; Figs. 3B and 4B), although there was marginal support ($P = 0.07$) for a mating system/genome size relationship in the outcrossing rate dataset. Similarly, the relationship between N_e and genome size disappeared in the PIC analyses (Fig. 5B). In summary, the PIC analyses found no evidence of strong relationships between genome size and either N_e or mating system.

These differences between the raw and PIC analyses reflect the concentration of extreme trait values in certain lineages. For example, in the outcrossing rate dataset, most species with large genomes and high outcrossing rates are in the genus *Pinus* (Fig 1). The same is true of the outcrossing index dataset, where *Pinus* is represented by 26 highly outcrossing species, all with large genomes (mean 2Cx = 49.5 pg) relative to the mean for the dataset (13.1 pg; Fig. 2). The second largest genus in this dataset (*Amaranthus*, seven species) contains mostly selfing species with relatively small genomes (mean 2Cx = 1.25 pg; Fig. 2). The PIC analyses account for the phylogenetic nonindependence of these observations and prevent them from driving an overall significant relationship between mating system and genome size.

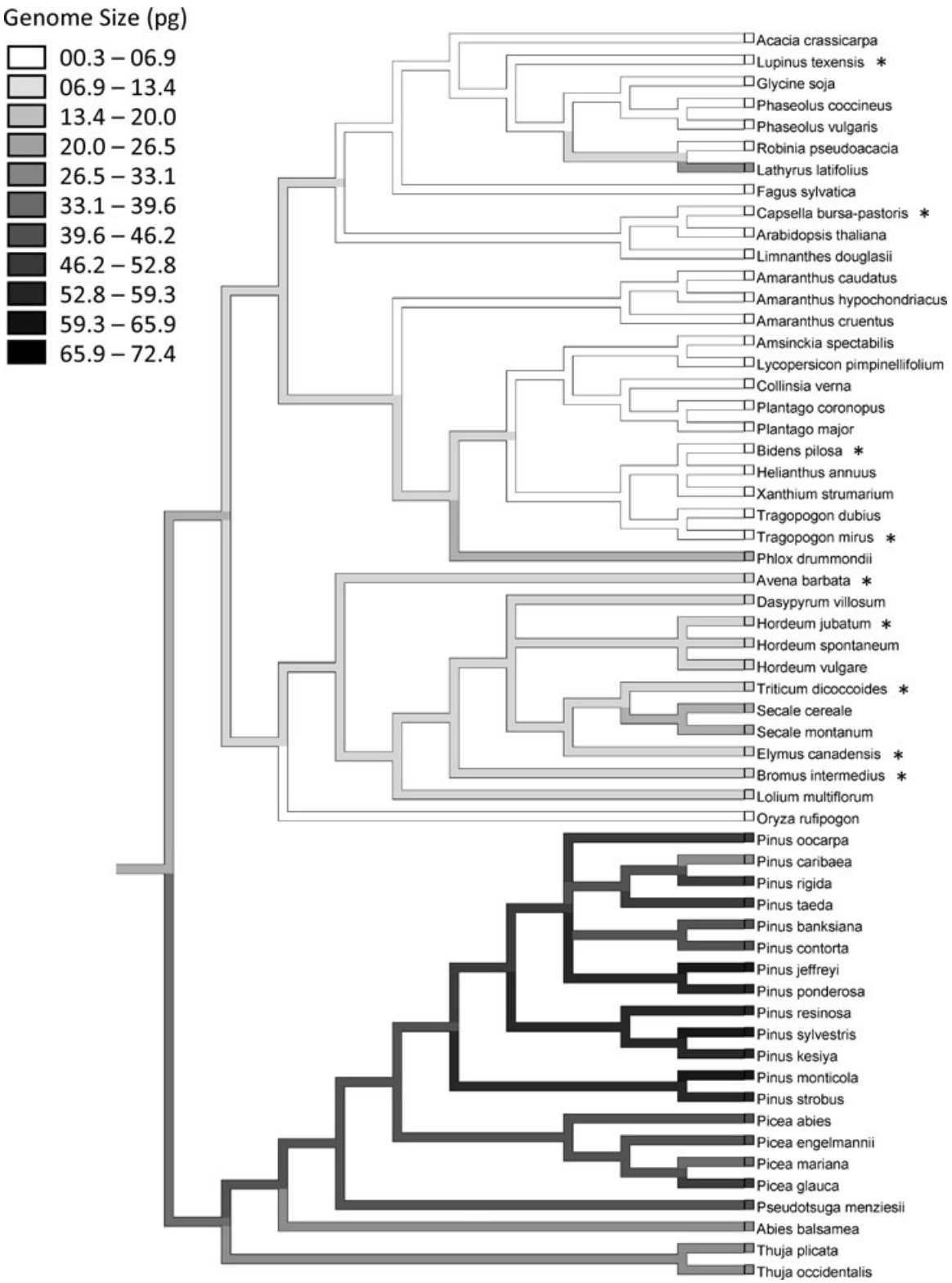

Figure 1. **Phylogeny for the outcrossing rate dataset, with a reconstruction of monoploid genome sizes. Recent polyploids are indicated with an asterisk.**

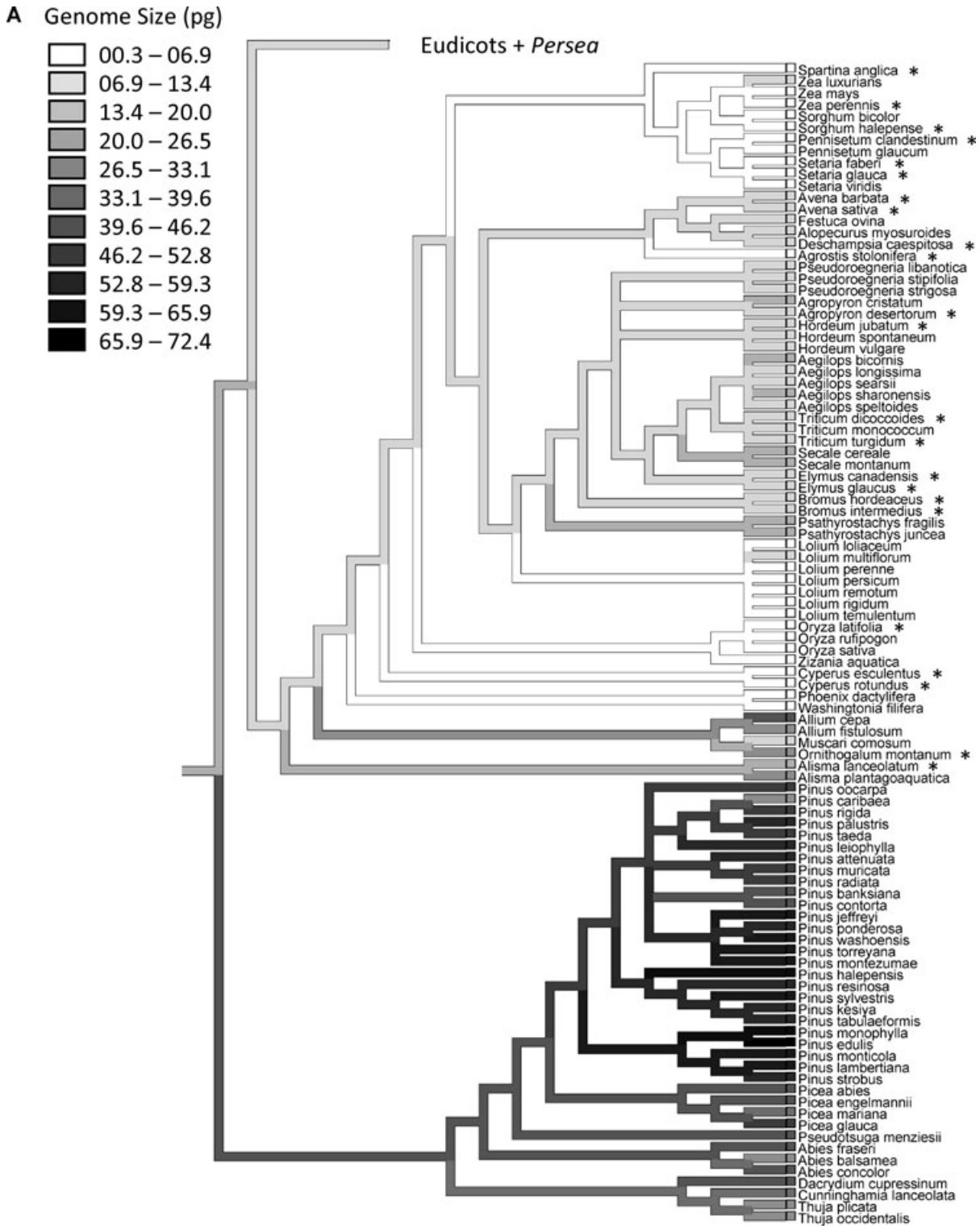

Figure 2. **(A) Phylogeny for the outcrossing index dataset, with a reconstruction of monoploid genome sizes. Recent polyploids are indicated with an asterisk. (A) Base phylogeny. (B) Detail for the Eudicot + *Persea* clade.**

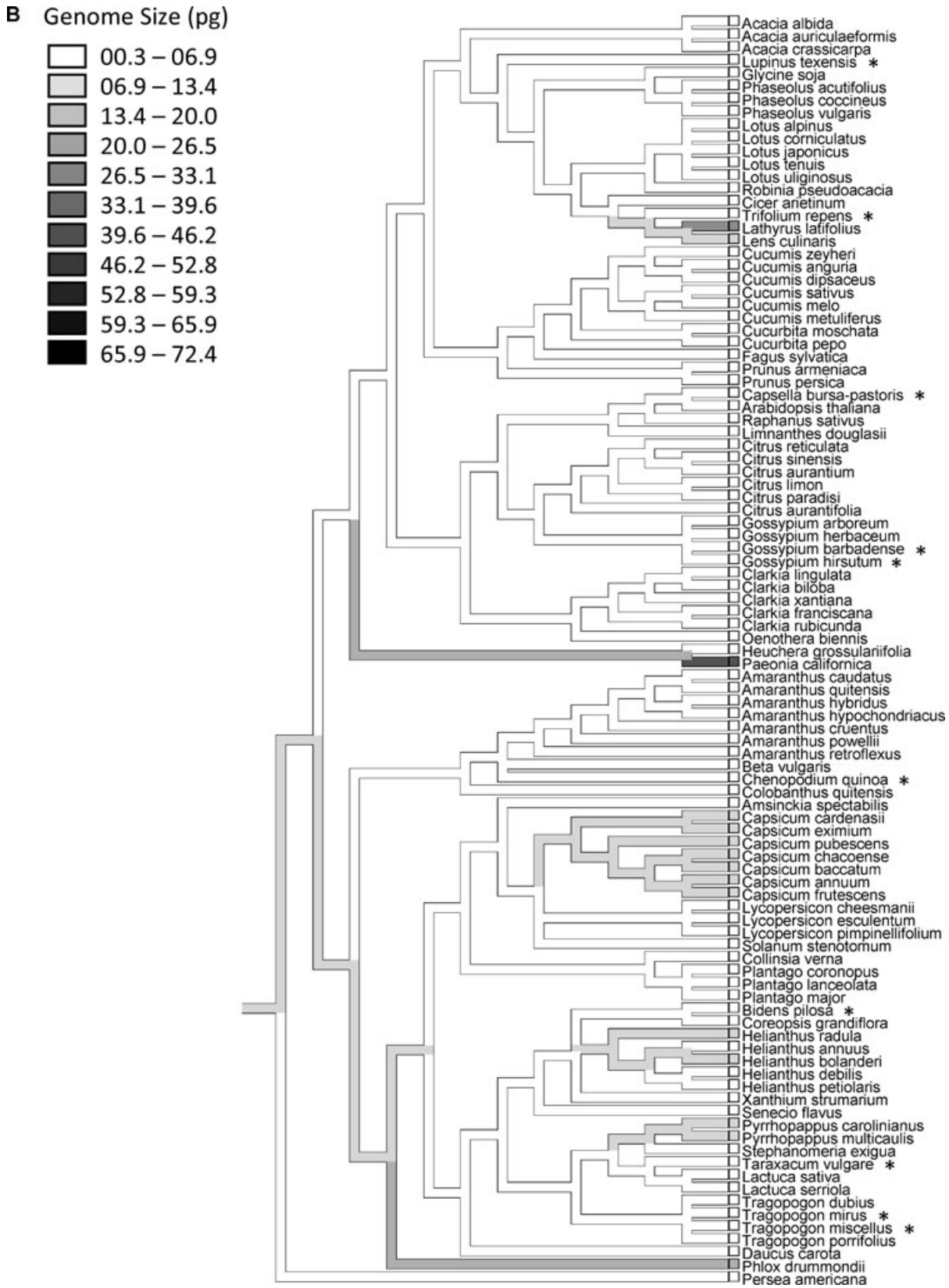

Figure 2. Continued.

Table 1. Results of multiple regression analyses of plant genome sizes on N_e and measures of outcrossing. Monoploid genome sizes were used to take into account recent polyploidization events (see Methods). Note that species in the outcrossing rate dataset are a subset of those in the outcrossing index dataset. *P* values significant at $P < 0.05$ are in bold. *b*′ partial regression coefficient

Dataset/predictor variables	Raw data				Phylogenetically independent contrasts			
	N	*b*′	r^2	*P*	*N*	*b*′*	r^2	*P*
Outcrossing rate dataset								
Model	58		0.46	**<0.0001**	57		0.06	0.311
Outcrossing rate (*t*)		30.75		**<0.0001**		6.61		0.066
N_e		0.000		0.144		−0.000		0.751
$N_e \times t$		−0.002		0.075		0.000		0.681
Outcrossing index dataset								
Model	205		0.22	**<0.0001**	198		0.01	0.621
Outcrossing index†		8.86		**<0.0001**		0.710		0.204
N_e		−0.000		**0.020**		−0.000		0.769
$N_e \times$ Outcrossing index		−0.000		0.189		0.000		0.627

*constrained through (0,0)

†1=selfing, 2=mixed, 3=outcrossing.

We also examined whether using H_{ep}, the expected population heterozygosity, rather than species-wide expected heterozygosity (H_{es}) would alter our results. Substituting H_{ep} (transformed to N_e, as above) resulted in very minor changes to *P*-values and no qualitative shifts in the patterns reported in Table 1 (data not shown). Finally, we examined whether use of uncorrected 2C genome sizes would alter our results. Substituting 2C for 2Cx genome sizes for the 35 recent polyploids in the dataset resulted in very minor changes to *P*-values and no qualitative shifts in the patterns reported in Table 1 (data not shown).

Discussion

We examined multiple factors potentially affecting genome size in seed plants. When phylogeny was taken into account, we found no relationship between genome size and N_e in either dataset, and a weak, marginally significant relationship between genome size and mating system in one of the two datasets. Our findings differ from previous studies reporting strong associations between N_e and genome size in fish (Yi and Streelman, 2005) and across kingdoms (Lynch and Conery 2003), as well as studies finding associations between mating system and genome size in plants (Albach and Greilhuber 2004, and studies reviewed therein).

EFFECTIVE POPULATION SIZE, GENETIC DRIFT, AND GENOME SIZE

Lynch and Conery's (2003) conclusion that genetic drift allows for maladaptive genome size expansion was based on an examination of N_e and genome size that ignored phylogenetic nonindependence of species. Although we found a similar negative relationship in one of our raw datasets, it was driven by phylogenetic nonindependence and disappeared after PICs were employed, despite a large sample size, a substantial range of N_e values ($\approx 10^0$–10^5), and a large range of genome sizes (0.3–65.9 pg). Possibly, our results may differ from those of Lynch and Conery (2003) because our range of N_e estimates, while spanning a similar number of orders of magnitude, was lower in absolute magnitude than theirs (roughly 10^0–10^5 vs. 10^4–10^8). However, the Lynch and Conery (2003) analysis contained only a small range of genome sizes ($\approx$0.002–3.0 pg) relative to the very large range of monoploid genome sizes in our dataset (0.3–65.9 pg), suggesting similar power to detect a N_e—genome size relationship. Thus, N_e does not appear to explain genome sizes in plants. It will be important to assess these patterns in other major eukaryotic groups, bacteria and archaea using corrections for phylogenetic nonindependence.

Apart from the current study, there are few large-scale phylogenetic comparative analyses of the impacts of N_e and/or drift on genome size. Results from such studies are mixed. In accord with the Lynch and Conery (2003) predictions, Yi and Streelman (2005) found a significant negative relationship between genome size and microsatellite-based N_e in ray-finned fish. Although reduced mating-system variation in ray-finned fish relative to plants could conceivably increase the ability to detect the consequences of N_e, the observed correlation could also reflect a number of methodological issues, including historical effects on microsatellite diversity and the confounding effects of polyploidy (Gregory and Witt 2008). In contrast, Kuo et al. (2009) analyzed 42 paired bacterial genomes, using the efficiency of purifying selection in coding regions to quantify genetic drift. Bacterial taxa experiencing greater levels of genetic drift—implying a smaller

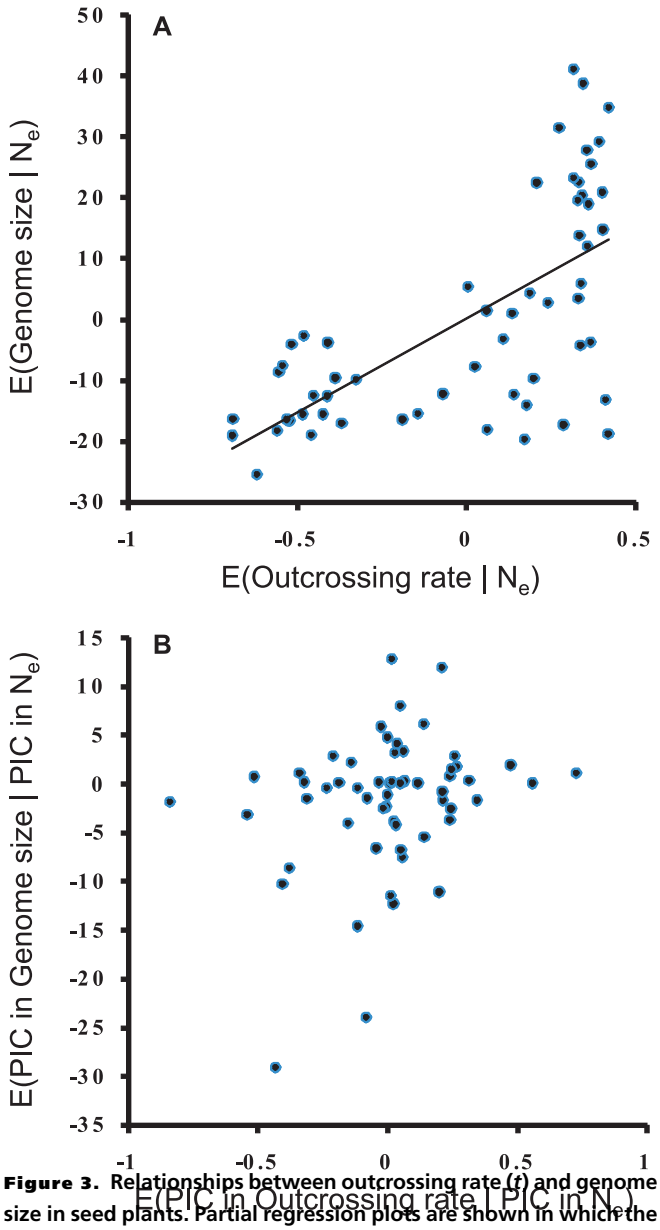

Figure 3. Relationships between outcrossing rate (*t*) and genome size in seed plants. Partial regression plots are shown in which the statistical effect of N_e has been removed. Both *t* and N_e were centered to mean = 0 prior to analysis. Lines are shown for significant relationships (see Table 1). (A) Raw *t* versus monoploid genome size (n = 58 species). (B) Phylogenetically independent contrasts, *t* versus monoploid genome size (n = 57 contrasts).

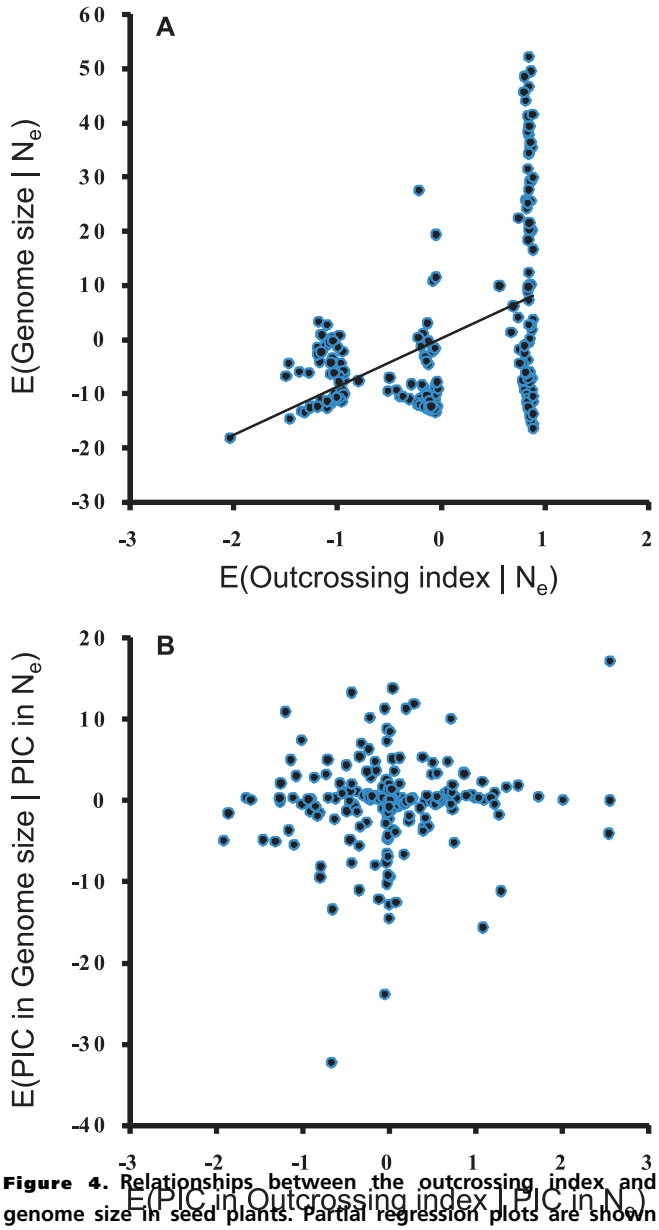

Figure 4. Relationships between the outcrossing index and genome size in seed plants. Partial regression plots are shown in which the statistical effect of N_e has been removed. Both the outcrossing index and N_e were centered to mean = 0 prior to analysis. Lines are shown for significant relationships (see Table 1). (A) Raw outcrossing index versus monoploid genome size (n = 205 species). (B) Phylogenetically independent contrasts, outcrossing index versus monoploid genome size (n = 198 contrasts).

evolutionary N_e—had smaller genomes. Kuo et al. interpret their results as evidence for an important role for nonadaptive processes in bacterial genome size evolution, but note that the direction of the relationship is opposite that predicted by Lynch and Conery (2003). Their explanation hinges on differences between bacteria and eukaryotes: deletions outweigh insertions in bacteria; furthermore gene duplications and proliferation of TEs are less prevalent in bacteria than in eukaryotes. Therefore, strong drift relative to selection should lead to reduced genomes over time in bacteria but not in eukaryotes (Kuo et al. 2009).

An important concern in these analyses is the reliability of estimates of N_e, given that use of alternative genetic markers can lead to different estimates. Such variation is expected from the varied fitness consequences of different mutations: for example,

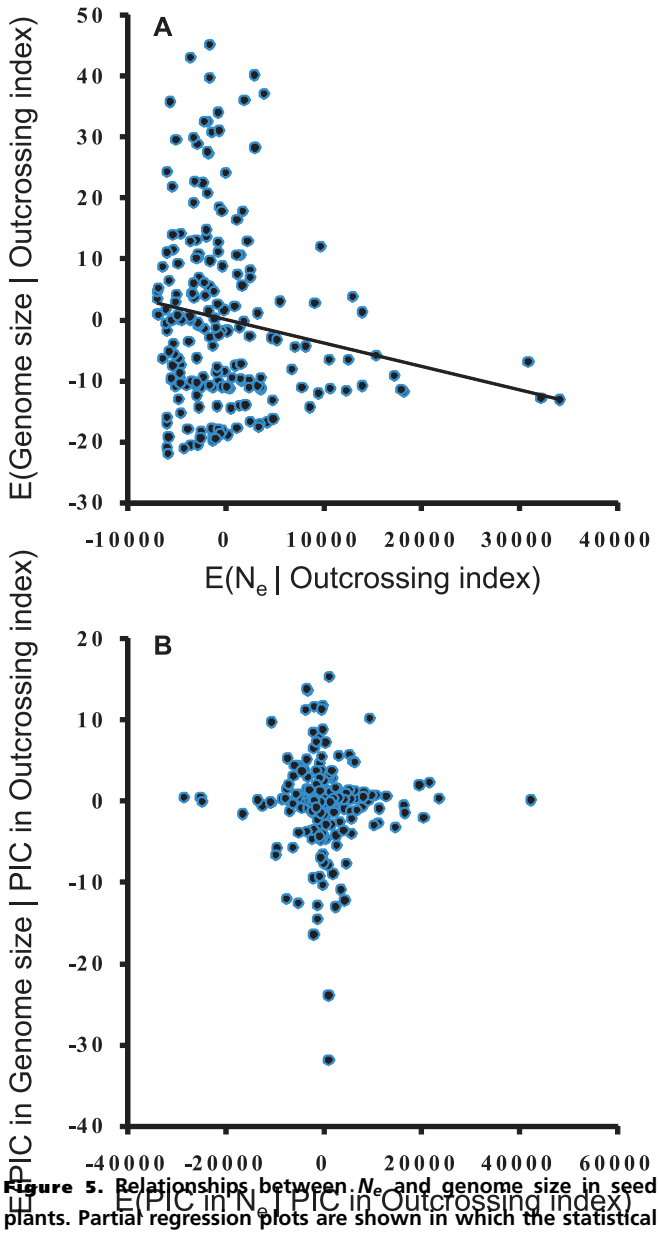

Figure 5. Relationships between N_e and genome size in seed plants. Partial regression plots are shown in which the statistical effect of the outcrossing index has been removed. Both the outcrossing index and N_e were centered to mean = 0 prior to analysis. Lines are shown for significant relationships (see Table 1). (A) Raw N_e versus monoploid genome size (n = 205 species). (B) Phylogenetically independent contrasts, N_e versus monoploid genome size (n = 198 contrasts).

replacement substitutions detectable by protein electrophoresis are far less likely to be neutral than silent substitutions, so that N_e estimates from allozymes are typically lower than estimates based on neutral sequence data (e.g., Strasburg and Rieseberg 2008). The diminished range of N_e derived from allozyme data may affect the power to detect a relationship between N_e and genome size. However, we do not expect the choice of data to introduce bias into analyses such as ours, in which all N_e estimates are derived from the same marker type. Furthermore, a significant positive correlation exists between allozyme-based N_e estimates and sequence-based N_e estimates ($P = 0.045$, $n = 14$ species), indicating that allozyme data are appropriate for estimating N_e. Most relevant here, for a 13-species subset of the Lynch and Conery (2003) dataset for which we have both allozyme and sequence data, allozyme-based N_e is as strongly related to genome size ($P = 0.052$, $r^2 = 0.301$) as is sequence-based N_e ($P = 0.061$, $r^2 = 0.284$; both correlations use phylogenetically uncorrected data).

MATING SYSTEM AND GENOME SIZE

Our PIC study found no relationship between mating system and genome size in one dataset and a weak, marginally significant relationship in the other (Table 1). It is possible that a biologically significant relationship exists in our data, but fails to achieve statistical significance because PIC analyses can inflate error variances relative to analyses of raw data (Ricklefs and Starck 1996). However, given the large difference between results from our own raw and PIC datasets, we suspect that most previously reported associations between mating system and genome size (e.g., Govindaraju and Cullis 1991) are simply driven by phylogenetic nonindependence of species and thus do not provide evidence of an evolutionary association between the variables. However, a recent study of the Veroniceae (Albach and Greilhuber 2004) used PICs and found a strong effect of mating system, wherein outcrossers have larger genome sizes. A possible explanation of the discrepancy between patterns in the Veroniceae and in our larger dataset may hinge on polyploidy. Some lineages of seed plants have undergone ancient genome duplications followed by reductions in genome size (Leitch and Bennett 2004), possibly in repeated cycles (Barker et al. 2008; Soltis et al. 2009). Although our analysis used monoploid genome sizes and thus accounted for recent polyploidy, it was not possible to account for older duplications and subsequent reductions in genome size. Studying the effects of mating system within a family, as in the Albach and Greilhuber (2004) analysis, may control for some of these ancient duplication events, making mating system effects more apparent. Similar studies in other families would help to determine the extent to which taxonomic scale is important, and would establish the reliability of the mating system effect. As genomic data from more taxa are studied for evidence of paleopolyploidy (Blanc and Wolfe 2004; Barker et al. 2008), it will become possible to construct analyses comparing families with and without paleopolyploidy, and thereby assess whether paleopolyploidy obscures the effects of mating system on genome size.

Mating system shifts may also obscure the effects of mating system on genome size. Transitions between outcrossing and self-pollination can occur rapidly in flowering plants (Barrett 2002).

Many studies suggest that outcrossing rate evolves rapidly in response to environmental shifts, including reduced population size at the range periphery (Busch 2005). Such lability would decrease any signal relating mating system to genome size.

It is also possible that transitions to selfing had no (or weak) net effects on genome size in our datasets because the resulting decreased TE transmission is coupled with decreased efficiency of selection. Findings that selfers have smaller genomes than outcrossers in other, smaller datasets (Albach and Greilhuber 2004; Wright et al. 2008) could indicate that avoidance of TEs is relatively more important than small N_e in determining genome size in these groups. Understanding the effects of mating system on TEs, however, requires molecular genetic characterization of TE copy number and population frequency in paired selfing and outcrossing taxa. Although such analyses have found evidence of decreased transmission of TEs in asexual lineages (Valizadeh and Crease 2008) and weaker selection against TEs in selfing lineages of some taxa (Wright et al. 2001; Dolgin et al. 2008; Hazzouri et al. 2008), other studies find little correlation between mating system and TE copy number (Tam et al. 2007). Predictions based on equilibrium expectations may not be easily applicable to the dynamic cycles of transmission, expansion, and quiescence likely important in many TEs (Le Rouzic et al. 2007.

CONCLUSIONS

Although comparative studies at the broadest scale (Lynch and Conery 2003) and at the intraspecific level (Lockton et al. 2008) detect correlations between effective population size and genome size, we found no evidence that effective population size shapes genome size in seed plants. Recent whole-genome duplications have clearly had a major effect on genome size in plants (Leitch and Bennett 2004), but other potential factors include earlier cycles of genome duplication and reduction, as well as mating system shifts. Although our analyses do not detect a strong effect of mating system on genome size, this could be due to opposing effects of selfing on TE transmission and the efficacy of natural selection. Overall, our results suggest that genome size within seed plants has likely been influenced by multiple factors and is not driven exclusively or even largely by effective population size.

ACKNOWLEDGMENTS

For discussions and helpful comments on the manuscript, many thanks to M. S. Barker, J. S. Johnston, M. Lynch, and J. A. Rudgers. Thanks also to M. F. Poelchau for her help with data compilation. Sincere thanks to P. F. Stevens for maintaining the Angiosperm Phylogeny Website and for advice on placement of genera.

LITERATURE CITED

Aiken, L. S., and S. G. West. 2001. Multiple regression: testing and interpreting interactions. Sage Publications, Thousand Oaks, CA.

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

Arkhipova, I., and M. Meselson. 2000. Transposable elements in sexual and ancient asexual taxa. Proc. Natl. Acad. Sci. USA97:14473–14477.

Barker, M. S., N. C. Kane, M. Matvienko, A. Kozik, W. Michelmore, S. J. Knapp, and L. H. Rieseberg. 2008. Multiple paleopolyploidizations during the evolution of the Compositae reveal parallel patterns of duplicate gene retention after millions of years. Mol. Biol. Evol. 25:2445–2455.

Barrett, S. C. H. 2002. The evolution of plant sexual diversity. Nat. Rev. Genet. 3:274–284.

Barringer, B. C. 2007. Polyploidy and self-fertilization in flowering plants. Am. J. Bot. 94:1527–1533.

Belsley, D. A., E. Kuh, and R. E. Welsch. 1980. Regression diagnostics: identifying influential data and sources of collinearity. John Wiley, New York.

Bennett, M. D., and I. J. Leitch. 2005a. Genome size evolution in plants. Pp. 89–162 *in* T. R. Gregory, ed. The evolution of the genome. Elsevier, Amsterdam.

———. 2005b. Plant DNA C-values database (release 4.0, Dec. 2005). Royal Botanic Gardens, Kew. Available at: http://data.kew.org/cvalues/.

Blanc, G., and K. H. Wolfe. 2004. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. The Plant Cell 16:1667–1678.

Busch, J. W. 2005. The evolution of self-compatibility in geographically peripheral populations of *Leavenworthia alabamica* (Brassicaceae). Am. J. Bot. 92:1503–1512.

Cassell, D. L. 2002. A randomization-test wrapper for SAS® PROCs. Pp. 251. Proceedings of the 27th Annual SAS User's Group International Conference. SAS Institute, Orlando, FL.

Charlesworth, B., and N. Barton. 2004. Genome size: does bigger mean worse? Curr. Biol. 14:R233–R235.

Charlesworth, D., and S. I. Wright. 2001. Breeding systems and genome evolution. Curr. Opin. Genet. Develop. 11:685–690.

Cruden, R. W. 1977. Pollen-ovule ratios: conservative indicator of breeding systems in flowering plants. Evolution 31:32–46.

Davies, T. J., T. G. Barraclough, M. W. Chase, P. S. Soltis, D. E. Soltis, and V. Savolainen. 2004. Darwin's abominable mystery: insights from a supertree of the angiosperms. Proc. Natl. Acad. Sci. USA 101:1904–1909.

Dolgin, E. S., B. Charlesworth, and A. D. Cutter. 2008. Population frequencies of transposable elements in selfing and outcrossing *Caenorhabditis* nematodes. Genet. Res. 90:317–329.

Drake, J. W., B. Charlesworth, D. Charlesworth, and J. F. Crow. 1998. Rates of spontaneous mutation. Genetics 148:1667–1686.

Felsenstein, J. 1985. Phylogenies and the comparative method. Am. Nat. 125:1–15.

Flowers, J. M., and M. D. Purugganan. 2008. The evolution of plant genomes—scaling up from a population perspective. Curr. Opin. Genet. Develop. 18:565–570.

Freckleton, R. P., P. H. Harvey, and M. Pagel. 2002. Phylogenetic analysis and comparative data: a test and review of evidence. Am. Nat. 160:712–726.

Fryxell, P. A. 1957. Mode of reproduction of higher plants. Bot. Rev. 23:135–233.

Garland, T., and R. Diaz-Uriarte. 1999. Polytomies and phylogenetically independent contrasts: examination of the bounded degrees of freedom approach. Syst. Biol. 48:547–558.

Garland, T., P. H. Harvey, and A. R. Ives. 1992. Procedures for the analysis of comparative data using phylogenetically independent contrasts. Syst. Biol. 41:18–32.

Garland, T., A. W. Dickerman, C. M. Janis, and J. A. Jones. 1993. Phylogenetic analysis of covariance by computer simulation. Syst. Biol. 42:265–292.

Goodwillie, C., S. Kalisz, and C. G. Eckert. 2005. The evolutionary enigma of mixed mating systems in plants: occurrence, theoretical explanations, and empirical evidence. Annu. Rev. Ecol. Evol. Syst. 36:47–79.

Govindaraju, D. R. 1988. Mating systems and the opportunity for group selection in plants. Evol. Trends Plants 2:99–106.

Govindaraju, D. R., and C. A. Cullis. 1991. Modulation of genome size in plants—the influence of breeding systems and neighborhood size. Evol. Trends Plants 5:43–51.

Gregory, T. R. 2005a. Animal Genome Size Database. Available at: http://www.genomesize.com.

———. ed. 2005b. The evolution of the genome. Elsevier, Amsterdam.

Gregory, T. R., and J. D. S. Witt. 2008. Population size and genome size in fishes: a closer look. Genome 51:309–313.

Greilhuber, J., J. Dolezel, M. A. Lysak, and M. D. Bennett. 2005. The origin, evolution and proposed stabilization of the terms 'genome size' and 'C-value' to describe nuclear DNA contents. Ann. Bot. 95:255–260.

Greilhuber, J., T. Borsch, K. Muller, A. Worberg, S. Porembski, and W. Barthlott. 2006. Smallest angiosperm genomes found in Lentibulariaceae, with chromosomes of bacterial size. Plant Biol. 8:770–777.

Grotkopp, E., M. Rejmanek, M. J. Sanderson, and T. L. Rost. 2004. Evolution of genome size in pines (*Pinus*) and its life-history correlates: supertree analyses. Evolution 58:1705–1729.

Hamrick, J. L., and M. J. W. Godt. 1989. Allozyme diversity in plant species. Pp. 43–63 *in* A. H. D. Brown, M. T. Clegg, A. L. Kahler, and B. S. Weir, eds. Plant population genetics, breeding and germplasm resources. Sinauer, Sunderland, MA.

———. 1996. Effects of life history traits on genetic diversity in plant species. Philos. Trans. R. Soc. Lond. B 351:1291–1298.

Hamrick, J. L., Y. B. Linhart, and J. B. Mitton. 1979. Relationships between life-history characteristics and electrophoretically detectable genetic variation in plants. Annu. Rev. Ecol. Syst. 10:173–200.

Hawkins, J. S., C. E. Grover, and J. F. Wendel. 2008a. Repeated big bangs and the expanding universe: directionality in plant genome size evolution. Plant Sci. 174:557–562.

Hawkins, J. S., G. J. Hu, R. A. Rapp, J. L. Grafenberg, and J. F. Wendel. 2008b. Phylogenetic determination of the pace of transposable element proliferation in plants: copia and LINE-like elements in *Gossypium*. Genome 51:11–18.

Hazzouri, K. M., A. Mohajer, S. I. Dejak, S. P. Otto, and S. I. Wright. 2008. Contrasting patterns of transposable-element insertion polymorphism and nucleotide diversity in autotetraploid and allotetraploid *Arabidopsis* species. Genetics 179:581–592.

Kullman, B., H. Tamm, and K. Kullman. 2005. Fungal genome size database. Available at: http://www.zbi.ee/fungal-genomesize.

Kuo C. H., N. A. Moran, and H. Ochman. 2009. The consequences of genetic drift for bacterial genome complexity. Genome Res. 19:1450–1454.

Leitch, I. J., and M. D. Bennett. 2004. Genome downsizing in polyploid plants. Biol. J. Linn. Soc. 82:651–663.

Le Rouzic, A., T. S. Boutin, and P. Capy. 2007. Long-term evolution of transposable elements. Proc. Natl. Acad. Sci. USA 104:19375–19380.

Lockton, S., J. Ross-Ibarra, and B. S. Gaut. 2008. Demography and weak selection drive patterns of transposable element diversity in natural populations of *Arabidopsis lyrata*. Proc. Natl. Acad. Sci. USA 105:13965–13970.

Lynch, M. 2007. The origins of genome architecture. Sinauer Associates, Sunderland, MA.

Lynch, M., and J. S. Conery. 2003. The origins of genome complexity. Science 302:1401–1404.

Maddison, W. P., and D. R. Maddison. 2009. Mesquite: a modular system for evolutionary analysis, v. 2.71. Available at: http://mesquiteproject.org.

Midford, P. E., T. Garland Jr, and W. Maddison. 2002. PDAP:PDTREE package for Mesquite, version 1.00. Available at: http://mesquiteproject.org/pdap_mesquite/.

Morgan, M. T. 2001. Transposable element number in mixed mating populations. Genet. Res. 77:261–275.

Neter, J., M. H. Kutner, C. J. Nachsteim, and W. Wasserman. 1996. Applied linear statistical models. Irwin, Chicago.

Nordborg, M. 2000. Linkage disequilibrium, gene trees and selfing: an ancestral recombination graph with partial self-fertilization. Genetics 154:923–929.

Ohta, T., and M. Kimura. 1973. Model of mutation appropriate to estimate number of electrophoretically detectable alleles in a finite population. Genet. Res. 22:201–204.

Oliver, M. J., D. Petrov, D. Ackerly, P. Falkowski, and O. M. Schofield. 2007. The mode and tempo of genome size evolution in eukaryotes. Genome Res. 17:594–601.

Pagel, M. 1999. Inferring the historical patterns of biological evolution. Nature 401:877–884.

Pagel, M., and A. Meade. 2009. BayesTraits. Univ. of Reading, UK. Available at: http://www.evolution.rdg.ac.uk/BayesTraits.html.

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

———. 2002. Mutational equilibrium model of genome size evolution. Theor. Popul. Biol. 61:531–544.

Price, H. J. 1976. Evolution of DNA content in higher plants. Bot. Rev. 42:27–52.

Ricklefs, R. E., and J. M. Starck. 1996. Applications of phylogenetically independent contrasts: a mixed progress report. Oikos 77:167–172.

SAS Institute. 2003. The SAS system for Windows, release 9.1. SAS Institute, Cary, NC.

Schoen, D. J., and A. H. D. Brown. 1991. Intraspecific variation in population gene diversity and effective population-size correlates with the mating system in plants. Proc. Natl. Acad. Sci. USA 88:4494–4497.

Soltis, D. E., V. A. Albert, J. Leebens-Mack, C. D. Bell, A. H. Paterson, C. F. Zheng, D. Sankoff, C. W. dePamphilis, P. K. Wall, and P. S. Soltis. 2009. Polyploidy and angiosperm diversification. Am. J. Bot. 96:336–348.

Strasburg, J. L., and L. H. Rieseberg. 2008. Molecular demographic history of the annual sunflowers *Helianthus annuus* and *H. petiolaris*—Large effective population sizes and rates of long-term gene flow. Evolution 62:1936–1950.

Tam, S. M., M. Causse, C. Garchery, H. Burck, C. Mhiri, and M. A. Grandbastien. 2007. The distribution of copia-type retrotransposons and the evolutionary history of tomato and related wild species. J. Evol. Biol. 20:1056–1072.

Valizadeh, P., and T. J. Crease. 2008. The association between breeding system and transposable element dynamics in *Daphnia pulex*. J. Mol. Evol. 66:643–654.

Vinogradov, A. E. 2003. Selfish DNA is maladaptive: evidence from the plant Red List. Trends Genetics 19:609–614.

Webb, C. O., and M. J. Donoghue. 2005. Phylomatic: tree assembly for applied phylogenetics. Mol. Ecol. Notes 5:181–183.

Wright, S. I., and D. J. Schoen. 1999. Transposon dynamics and the breeding system. Genetica 107:139–148.

Wright, S. I., Q. H. Le, D. J. Schoen, and T. E. Bureau. 2001. Population dynamics of an Ac-like transposable element in self- and cross-pollinating *Arabidopsis*. Genetics 158:1279–1288.

Wright, S. I., R. W. Ness, J. P. Foxe, and S. C. H. Barrett. 2008. Genomic consequences of outcrossing and selfing in plants. Int. J. Plant Sci. 169:105–118.

Yi, S., and J. T. Streelman. 2005. Genome size is negatively correlated with effective population size in ray-finned fish. Trends Genet. 21:643–646.
Zuccolo, A., A. Sebastian, J. Talag, Y. Yu, H. Kim, K. Collura, D. Kudrna, and R. A. Wing. 2007. Transposable element distribution, abundance and role in genome size variation in the genus *Oryza*. BMC Evol. Biol. 7:152.

Associate Editor: J. Vamosi

Supporting Information

The following supporting information is available for this article:

Appendix S1. Allozyme-based estimates of effective population sizes (N_e) for the 205 species in the dataset.
Appendix S2. Sources for within-family and within-genus phylogenetic relationships used in constructing the supertrees.

Supporting Information may be found in the online version of this article.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.