vol. 185, no. 4 the american naturalist april 2015

Exploring the Relationships between Mutation Rates, Life History, Genome Size, Environment, and Species Richness in Flowering Plants

Lindell Bromham,[1,*] Xia Hua,[1] Robert Lanfear,[1,2] and Peter F. Cowman[1,3]

1. Centre for Macroevolution and Macroecology, Division of Evolution, Ecology and Genetics, Research School of Biology, Australian National University, Canberra, Australian Capital Territory 0200, Australia; 2. Department of Biological Sciences, Macquarie University, North Ryde, New South Wales 2109, Australia; 3. Department of Ecology and Evolutionary Biology, Yale University, New Haven, Connecticut 06511

Submitted September 23, 2014; Accepted November 13, 2014; Electronically published February 10, 2015

Online enhancements: appendix, zip file. *Dryad data:* http://dx.doi.org/10.5061/dryad.31614.

ABSTRACT: A new view is emerging of the interplay between mutation at the genomic level, substitution at the population level, and diversification at the lineage level. Many studies have suggested that rate of molecular evolution is linked to rate of diversification, but few have evaluated competing hypotheses. By analyzing sequences from 130 families of angiosperms, we show that variation in the synonymous substitution rate is correlated among genes from the mitochondrial, chloroplast, and nuclear genomes and linked to differences in traits among families (average height and genome size). Within each genome, synonymous rates are correlated to nonsynonymous substitution rates, suggesting that increasing the mutation rate results in a faster rate of genome evolution. Substitution rates are correlated with species richness in protein-coding sequences from the chloroplast and nuclear genomes. These data suggest that species traits contribute to lineage-specific differences in the mutation rate that drive both synonymous and nonsynonymous rates of change across all three genomes, which in turn contribute to greater rates of divergence between populations, generating higher rates of diversification. These observations link mutation in individuals to population-level processes and to patterns of lineage divergence.

Keywords: comparative, phylogeny, substitution rates, angiosperm, diversification rate.

Introduction

Diversification is the process of changes in diversity by the addition of new species through speciation and the loss of species by extinction. Phylogenies allow a way of comparing the rate of diversification between lineages, for example, by comparing the number of extant species in clades relative to their age (the net diversification rate). In addition to allowing widespread comparison of diversification rates, molecular phylogenies have been used to compare the net rate of diversification to the rate of molecular evolution, estimated from phylogenetic branch lengths. A growing number of studies have reported significant links between rates of molecular evolution and net diversification for a range of different genes and lineages. Some of these studies compared rates of molecular evolution in sister pairs of lineages that differ in the number of extant species, showing that the lineages with a greater net rate of diversification tend to have higher substitution rates (Barraclough and Savolainen 2001; Davies et al. 2004*b*; Eo and DeWoody 2010; Lancaster 2010; Lanfear et al. 2010*a*; Duchene and Bromham 2013). Others have shown that estimates of molecular change along paths through phylogenies are correlated to the number of inferred speciation events (Webster et al. 2003; Pagel et al. 2006; Ezard et al. 2013).

The link between rates of genome change (as measured from molecular phylogenetic branch lengths) and net diversification (as measured by extant species richness) provides a fascinating insight into evolutionary processes. The role of genetic change in speciation has been intensively studied for more than a century (Bateson 1894), but most research has focused on specific loci or mechanisms that cause populations to become genetically isolated from one another (e.g., Nosil and Schluter 2011; Strasburg et al. 2012; Abbott et al. 2013). Molecular phylogenetic studies provide a very different view of the relationship between genetic change and diversification, because they typically do not include genes likely to be involved with mate recognition or local adaptation. Instead, they are usually based on "housekeeping genes" with general metabolic and biochemical functions. Since phylogenetic markers come from a range of loci across nuclear and organellar genomes, any general relationship between phylogenetic branch lengths and spe-

* Corresponding author; e-mail: lindell.bromham@anu.edu.au.

Am. Nat. 2015. Vol. 185, pp. 507–524.
DOI: 10.1086/680052

cies richness suggests that genome-wide rates of molecular change are correlated with diversification rate.

However, while there is growing evidence for a widespread relationship between rates of genome evolution and net diversification (though not universal; see Pagel et al. 2006; Goldie et al. 2011), there has been relatively little progress in explaining this pattern. Broadly speaking, there are three possible explanations for the link between the diversification and molecular evolution rates. First, an increased rate of molecular evolution could promote a higher rate of diversification. A higher rate of genomic change could speed the accumulation of substitutions that cause genomic incompatibility between hybrids, thus reinforcing reproductive isolation between incipient species. Increasing the supply of variation might also provide more genetic variation for selection for local adaptations or isolating mechanisms, or reduce the likelihood of extinction by increasing standing genetic variation.

Second, a link between the rates of diversification and molecular evolution might be mediated by an association between the process of speciation increasing the rate of molecular evolution, for example, due to adaptation or through the influences of population subdivision on substitution. If speciation is typically accompanied by a reduction in effective population size, for example, through isolated founder populations, then genetic drift may have a large effect on substitution rates immediately after the formation of a new, isolated population (Venditti and Pagel 2009). This could lead to a transient rise in the fixation of nearly neutral substitutions. Conversely, speciation may be associated with a burst of substitutions in loci under positive selection corresponding to adaptations to new niches or reproductive isolating mechanisms.

A third possibility is that there is some other factor that is correlated with both diversification rate and rate of molecular evolution, such as life-history or environmental factors, that causes an indirect correlation between the two. For example, rapid generational turnover could increase both the rate of molecular evolution (through the generation time effect on DNA mutation rates; Bromham 2009) and the rate of diversification (if higher intrinsic rates of population growth reduce extinction risk; Davies et al. 2004*a*). It has also been suggested that the relationship between species richness and substitution rate in plants could be an artefact of both variables being correlated with environmental energy (Davies et al. 2004*b*).

Our aim in this study was to weigh up evidence for each of these alternative (nonexclusive) hypotheses concerning the link between rates of molecular evolution and diversification by building on the analysis of an earlier study that used a broad, family-level database to study rates of molecular evolution in flowering plants. While the focus of that paper was on the relationship between plant height and rate of molecular evolution, it also reported a significant positive correlation between chloroplast substitution rates and species richness and a similar, but nonsignificant, positive correlation with nuclear ribosomal RNA (rRNA; Lanfear et al. 2013). Here we wish to investigate this reported correlation, in the hope that additional data and more detailed analysis may help to uncover the underlying causes of the link between species richness and rate of molecular evolution. To this end, we expand the database to include genes from all three genomes, add data on an additional explanatory variable (genome size), and perform more detailed statistical analyses in order to dissect the relationship between life history, environment, species richness, and rates of molecular evolution in plants. Our aims in this study were to (*a*) test for a link between rate of molecular evolution and net diversification rate in each of the three plant genomes (nuclear, mitochondrial, chloroplast); (*b*) ask whether patterns of substitution can distinguish between proposed explanations for the relationship between rates of net diversification and molecular evolution; (*c*) evaluate the possible causes of this correlation by investigating the relationship between substitution rate, family traits, and environment; and (*d*) discern any links between the substitution rates of the chloroplast, mitochondrial, and nuclear genomes.

Material and Methods

Traits

For the sake of brevity, we will use the term traits to refer to the explanatory variables that describe the average characteristics (height, genome size) and environmental conditions (latitude, temperature, ultraviolet [UV] radiation) for a taxon. Our starting point for this analysis is a data set consisting of phylogenetically independent comparisons between sister families of flowering plants (Lanfear et al. 2013) because it has previously been shown to support a general, broadscale correlation between species richness and rates of molecular evolution for chloroplast genes. We build on this database in order to undertake a more comprehensive investigation of rates of molecular evolution in plant families by extending the analysis to protein-coding genes from all three genomes (chloroplast, nuclear, mitochondrial). We also add genome size, which is potentially an important correlate of life history, species richness, ecology, and rate of molecular evolution in some plant lineages (e.g., Bennett 1972; Beaulieu et al. 2007, 2008, 2010; Knight and Beaulieu 2008; Hodgson et al. 2010; Kraaijeveld 2010; Lavergne et al. 2010; Herben et al. 2012; Kang et al. 2014).

Family-level estimates of average height were derived from a large database of maximum recorded plant height (Moles et al. 2009) by first calculating the mean log-

transformed height for all species in each genus and then calculating the mean of all genera in each family (table S1; tables S1–S9 available online in a zip file). We included three environmental variables for each family, each of which is calculated for the entire range of the family (i.e., the polygon that contains the range of all species in that family): the mean level of current UV irradiance experienced by each family, the mean current temperature experienced by each family, and centroid of each family's absolute latitude (Davies et al. 2004*a*; Lanfear et al. 2013). Clearly many families will contain species with a great diversity of heights, and the mean environmental variables for widespread families will not reflect the conditions encountered by all of the species within that family. But these clade averages have been shown to be correlated with rates of molecular evolution, and the correlations they reveal are broadly consistent with finer-scale investigations of the effect of body size and life-history characters on rates of molecular evolution (e.g., Smith and Donoghue 2008). This implies that, for the purposes of investigating general patterns in rates of molecular evolution, these family-level averages provide an adequate approximation of the difference in average body size between sister lineages since they last shared a common ancestor (given that much of this history will be shared by all species in a family). In this study, we are specifically concerned with explaining the intriguing relationship already detected in family-level estimates of species richness and substitution rates, rather than a more general test of the influence of environmental and life-history traits on diversification rates.

We used clade size (extant species richness for each family) to represent net diversification rate. For each phylogenetically independent pair of sister clades, any difference in the number of species between sister families must have arisen since they last shared a common ancestor through a difference in the speciation rate or extinction rate or both (Lanfear et al. 2011). The number of recognized species per family was derived from the Families of Flowering Plants database (http://delta-intkey.com). Although the absolute numbers of species per family may vary with taxonomic treatment, the relative differences in species between sister families in this database have been shown to be associated with rates of molecular evolution (Lanfear et al. 2013).

To represent genome size, we collected 1C-values from the Royal Botanic Gardens, Kew, Plant DNA *C*-values database (Bennett and Leitch 2012) for plant families in our comparisons. The C-value, representing the amount of DNA in a haploid nucleus in picograms, is the most widely used measure of genome size and so is available for a large number of species. Variation in C-value is influenced by many different aspects of genome content and organization, including number of genes, degree of duplication within the genome, repetitive elements, transposable elements, and endogenous viruses, but separate estimates of these different components of genome size are not available for a large enough sample of species to allow comparison between the sister-family comparisons included in this study. The 1C measure of genome size has been shown to correlate with cell size and cell generation time (Bennett and Leitch 2005), making it an appropriate measure of genome size for this analysis.

To estimate the average genome size for each family, we first averaged all available species estimates in each genus and then averaged the genus-level means to get the genome size estimates for the family. Given that comprehensive phylogenies are not available for the majority of families included in this study, this taxonomic averaging approach approximates a phylogenetic average and helps to make life-history and molecular rate estimates more directly comparable (Welch and Waxman 2008). Since C-value estimates are not available for all families, we could not include an average genome size for all of our comparisons. In total, there were 44 comparisons for which we had genome size estimates for both sister families (table S1).

Rates

For convenience, we will use the term rates to describe all estimates of substitution rate made from analyses of DNA sequences. We estimate the rate of all substitutions (total), synonymous substitution rate (dS), nonsynonymous substitution rate (dN), and the ratio of nonsynonymous to synonymous substitutions (dN/dS). Synonymous changes typically do not affect the phenotype so are expected to be neutral with respect to fitness (or very nearly so). As a result, the synonymous substitution rate is largely determined by the rate at which synonymous changes are generated by mutation (Kimura 1983). Nonsynonymous substitutions are expected to show a range of fitness effects, from deleterious to neutral to advantageous, so nonsynonymous substitution rates may therefore be influenced by the mutation rate (which determines the rate of generation of variants), by selection (which affects the rate of substitution of advantageous mutations or the removal of deleterious changes), and by population size (which influences the rate of substitution of nearly neutral mutations that have small selective effects). Changes in the action of selection or variation in population size may be reflected in the ratio of nonsynonymous to synonymous changes.

To compare synonymous and nonsynonymous mutation rates from all three genomes between families, we collected available sequence data for sister pairs of plant families from protein-coding genes from the nuclear, chloroplast, and mitochondrial genomes. For a gene to be suitable for this study, it needed to be available for a large enough number of plant

families that we could make sufficient sister-family comparisons for this statistical analysis. The genes needed to be sufficiently conserved to allow alignment across diverse plant families yet variable enough to allow estimation of both synonymous and nonsynonymous substitution rates. We used four mitochondrial genes (*atp1*, *matR*, *nad5*, and *rps3*) taken from the data set of Soltis et al. (2011), resulting in 48 comparisons for the mitochondrial alignment (192 taxa, 5,211 bp; table S2). For the chloroplast genome, we used the substitution rate differences previously estimated for the 65 sister-family comparisons (Lanfear et al. 2013) based on the protein-coding genes *atpB* and *rbcL* (196 taxa, 2,897 bp). For the nuclear genome, we could find only one suitable protein-coding gene that was available for enough of the sister-family comparisons: we assembled an alignment of sequences from the protein-coding gene *xanthine dehydrogenase* (*Xdh*; Gorniak et al. 2010; Morton 2011), representing 23 sister-family comparisons (62 taxa, 1,113 bp; table S3). We also included previous estimates of substitution rates for these sister families from the nuclear rRNA genes *18s* and *26s* (196 taxa, 5,275 bp; Burleigh et al. 2009; Lanfear et al. 2013).

For each comparison, we used the same number of taxa to represent each sister clade. Balancing the number of tips in each sister clade helps to reduce the impact of the node-density effect, which is the propensity for molecular rates to be underestimated on long branches. Substitution rate estimates will tend to increase as more taxa are added to a clade, because adding more lineages will tend to break up the internal edges of the phylogeny, allowing better inference of multiple hits (substitutions overwritten by subsequent changes). While node density is most clearly a problem for parsimony reconstructions (Sanderson 1990), it may also impact maximum likelihood estimates of branch length (Hugall and Lee 2007). Node-density effect is a particular concern for studies examining the relationship between diversification rate and rate of molecular evolution, because it can generate spurious correlations between the substitution rate estimates and clade size by increasing substitution rate estimates in more species-rich clades (those containing more nodes relative to age). While tests of the node-density effect have been developed, these are designed for whole-tree analyses, comparing the root-to-tip path lengths to the number of descendant lineages (Venditti et al. 2008). The best strategy for avoiding node-density effect for sister-clade analyses is careful data selection (Lanfear et al. 2010*b*).

Where we had a choice of several taxa, we preferentially selected taxa with the greatest coverage of sequence data, but all else being equal we chose taxa at random. In most of the comparisons, each sister lineage was represented by sequences from a single taxon; but in seven comparisons, each family is represented by two taxa (tables S2, S3). While a single sequence will not capture species-specific variation in rate of molecular evolution within families, comparisons between single sequences will represent much of the shared history of each family, and this approach has previously demonstrated an association between clade size and rate of molecular evolution (Davies et al. 2004*b*; Lanfear et al. 2010*a*; Duchene and Bromham 2013).

Initial alignments of mitochondrial and nuclear protein-coding genes were performed using the MAFFT translation alignment plug-in for Geneious 6.0 (Katoh et al. 2002; Drummond et al. 2011). We manually removed any incomplete codons, for example, due to gaps in the original sequence or truncated codons at the ends of the sequences. We then created separate mitochondrial and nuclear alignments for each sister-clade comparison, including representative sequences for each of the sister families, plus a closely related family as an outgroup (see tables S2, S3). For each of these sister-clade alignments, we included only sequences that had coverage for both families in the pair. For the mitochondrial data, 32 of the 48 comparisons had complete coverage for all four mitochondrial genes, 14 comparisons had two or three genes, and two comparisons had only one gene (table S2). The topology for these sister pairs plus outgroups were taken from the maximum likelihood phylogeny (Lanfear et al. 2013) by extracting subtrees using the APE package in R (Paradis et al. 2004).

For each comparison, we estimated synonymous (dS) and nonsynonymous (dN) branch lengths and dN/dS ratios using the GY94 codon substitution model (Goldman and Yang 1994) in the codeml program of the PAML package (ver. 4.4b; Yang 2007), with dN/dS values free to vary across the tree. For comparisons with more than one taxon per family, we calculated clade averages for dS and dN and a single dN/dS ratio for the family (Lanfear et al. 2013). For the *Xdh* gene, we also estimated total branch length using the baseml program in PAML. This was to facilitate comparison with the nuclear rRNA genes, since dS and dN cannot be estimated for RNA-coding genes.

All the variables were calculated as the differences in family-average estimates between the two sister families in each comparison. We performed diagnostic tests on the data, as described in Lanfear et al. (2010*b*), to evaluate the three basic assumptions of generalized linear models for data generated by phylogenetic comparative methods. The first criterion is whether variance of trait differences is still related to their absolute values after appropriate data transformation. The second criterion is whether variance of trait and rate differences still increases linearly with evolutionary time after standardizing the sister pairs by branch lengths: differences in shallow pairs tend to decrease with evolutionary time due to stochastic fluctuations in substitution numbers (Welch and Waxman 2008); therefore, these

shallow pairs are problematic data points. The third criterion is whether there still exists a negative relationship between contrasts and evolutionary time after excluding shallow pairs. These tests indicated that log transformations were appropriate for all the contrasts except for the environmental variables, for which temperature and UV were squared and latitude was not transformed.

Based on the diagnostic tests, only substitution rate estimates were standardized by branch length. Standardization of rate estimates by the relative depth of the comparison is necessary to ensure that the data meet the assumptions of homogeneity of variance by accounting for the fact that both trait and rate differences are expected to increase with evolutionary time. Ideally, we would use the time since the last common ancestor for each sister pair to standardize the comparison, but unfortunately, dates of divergence (independent of molecular data) are not available for the majority of the comparisons in our study. Therefore, we use the best estimator for evolutionary time that we could obtain, which is the square root of the total number of substitutions in the chloroplast genes and nuclear rRNA (as these two sequences are available for all the family comparisons). While using branch length to standardize the comparisons is not perfect, it is preferable to not using any standardization for comparison depth.

We applied the test of Welch and Waxman (2008) to detect and remove sister pairs for which we were unable to make reliable estimates of differences in molecular branch lengths. This test removes shallow sister pairs until there no longer exists a negative relationship between contrasts and evolutionary time, and is not based on the absolute value of the difference between the sister lineages. After removing these problematic pairs, our sample sizes for analyses were as follows: 65 sister-family comparisons for both the chloroplast genes and nuclear rRNA; 48 sister-family comparisons for mitochondrial genes; and 22 sister-family comparisons for the nuclear protein-coding gene, *Xdh*. Because genome size estimates were not available for all families in the data set, the set of comparisons that includes genome size for both sister families was reduced to 42 comparisons for chloroplast genes, 40 for nuclear rRNA, 27 for mitochondrial genes, and 14 for the nuclear *Xdh* gene. Because the reduction in the number of comparisons might result in lower power, we repeated key analyses on both the full set of comparisons without genome size and on the reduced data set including genome size.

Analysis

Our aims were to describe the relationships between traits and the rates of molecular evolution and net diversification and to ask whether any associations are due to direct or indirect relationships. Therefore, we employed a number of statistical methods that control for colinearity of variables and that accounted for relationship structures among variables based on a priori hypotheses.

Are Traits Correlated with Rates? We began by asking how family traits are associated with substitution rates and species richness, respectively. The environmental variables (latitude, temperature, and UV) are highly correlated with one another ($P < 10^{-5}$; table S4), so in order to minimize the effect of multicollinearity, we performed ridge regression and partial least squares regression (Izenman 2008). Ridge regression reduces the mean squared errors of regression coefficients that are inflated by multicollinearity. Partial least squares regression (PLSR) finds the components on which explanatory and response variables have the largest covariance.

We performed ridge regression and PLSR using the environmental variables, plant height, and genome size as the explanatory variables and the substitution rates and species richness as the response variables. We also performed ridge regressions using the substitution rates from the chloroplast, mitochondrial, and nuclear genes as the explanatory variables and the species richness as the response variable. The root mean square error of prediction, calculated by jackknifing cross validation, was used for choosing the optimal number of components. Jackknifing was used to estimate the variances of PLSR coefficients, although the resulting *P* value is not as reliable as ridge regression (Mevik et al. 2011). To further reduce the impact of multicollinearity and clarify the effects of environmental variables on substitution rates and species richness, we performed principal component analysis on latitude, temperature, and UV and then performed ridge regression using all three principal components as the explanatory variables.

Are Rates Correlated across Genomes? Variation in rates of molecular evolution may be correlated across the three cellular genomes, either directly, through selection for compensating mutations that allow the genomes to continue to work together in the face of genomic change, or indirectly, through common effects such as cell generation time and copy frequency. Therefore, any attempt to understand the driving forces of rate variation across flowering plants should ideally take into account correlation between genomes. We looked for correlations between substitution rate estimates from the chloroplast, mitochondrial, and nuclear alignments with the aim of detecting any interactions between rates of molecular evolution in different genomes. We used Kendall's rank correlation analysis because it does not rely on the assumption that missing data is random (Sokal and Rohlf 1995).

Are Rates Correlated with Species Richness? We performed path analysis to explore the nature of the link between family-average traits (height, genome size, and environmental variables), substitution rates, and species richness. Path analysis (and structural equation modeling in general) calculates the fit of a predefined causal structure among variables by comparing the actual covariance matrices among variables and the estimated covariance matrices of the fitted structure. Dependence between variables can be represented by lines in a diagram, with only those paths that represent a plausible interpretation of the data included in the model (Wright 1934). This provides a way to assess the relationships between variables, including through the effect on other variables. While path analysis cannot be used to establish causal mechanisms, correlations in the data do imply some causal connection (whether direct or indirect) between the variables (Shipley 2002). The role of path analysis is to evaluate the relative strengths of alternative models by asking how well the pattern of correlations matches predefined links between variables. Because path analysis provides a way of assessing the support in the data for predefined causal models, researchers must explicitly describe reasonable hypotheses for the relationships between variables (Lleras 2005). We based our models on a priori hypotheses discussed in the literature (see "Introduction" and "Discussion") but also use the results of the regression analyses to inform these models, as we do not include links between variables if they are not correlated with each other.

We test the data against four models that describe the causal relationship between traits, rates, and species richness (fig. A1, available online). Models 1 and 2 differ in the direction of the causal relationship between substitution rates and species richness. Model 1 represents an influence of rates of molecular evolution on rates of diversification, so substitution rates may influence species richness, and traits (environment and life history) can influence both rates and species richness. Model 2 represents an effect of rates of diversification on rates of molecular evolution, so in this model species richness may influence rates, while traits (height, genome size, environment) can influence both rates and species richness. Model 3 represents a direct connection between molecular evolution and diversification rates, so traits can influence both species richness and rates, and species richness has residual covariance with substitution rates. Model 4 represents an indirect connection between species richness and substitution, so traits can influence species richness and rates, but species richness has no residual covariance with substitution rates.

Path analysis does not provide proof of causal mechanisms; instead, it provides a way of rejecting some causal hypotheses as being a less adequate description of the data if it does not fit the pattern of covariance as well as an alternative model. For example, in model 1, we assume substitution rates have a direct effect on diversification rates, so the estimated covariance matrix involves the variance of diversification rates conditional on the substitution rates; in model 2, we assume diversification rates have a direct effect on substitution rates, so the estimated covariance matrix involves the variance of substitution rates conditional on the diversification rates. As a result, different predefined causal structures lead to different estimated covariance matrices among variables, and the best structure is the one whose estimated covariance matrices are most similar to the actual covariance matrices.

We compare the relative fit of different models using the Aikake information criterion (AIC), which compares the goodness of fit of different models with respect to the number of parameters of the model. In this way, the AIC compares models with respect to both their explanatory power and complexity. We use a χ^2 test to reject models, which compares the difference between the observed covariance matrix among variables and the expected covariance matrix given the predefined casual structure against a χ^2 null distribution.

All the statistical analyses were performed in R (R Development Core Team 2013), for ridge regression using the ridge package (Cule 2012), for partial least squares regression using the pls package (Mevik et al. 2011), and for path analysis using the lavaan package (Rosseel 2012). All data used have been deposited in the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.31614 (Bromham et al. 2015).

Results

Are Traits Correlated with Rates? Results of ridge regression and partial least squares regression confirm a consistent negative association between average height in flowering plant families and substitution rates, for synonymous substitutions in the mitochondrial and chloroplast genes, and for total branch length in the nuclear rRNA (table 1). Families with lower average height also tend to have a lower ratio of nonsynonymous to synonymous substitution rates (dN/dS) in the chloroplast genes (table 1), a pattern also seen in the mitochondrial genes (tables S5, S6, S8) though not significant for the reduced data set (which has fewer comparisons due to the inclusion of genome size; tables 1, S7, S9). Plant families with smaller average genome sizes tend to have faster synonymous substitution rates, total substitution rates, and lower dN/dS in the nuclear protein-coding gene *Xdh* (table 1).

In the principal component analysis, temperature (loading $= -0.58$), latitude (loading $= 0.58$), and UV (loading $= -0.58$) have the same absolute loadings on the

Table 1: Results of ridge regression for substitution rates and species richness against plant height, genome size, and principal components 1–3 (PC1–PC3) of environmental variables

	Chloroplast			Mitochondrial			Nuclear					
	dN	dS	dN/dS	dN	dS	dN/dS	dN	dS	dN/dS	*Xdh* total	rRNA total	Species richness
Height:												
r	−.00	**−1.53**	**.82**	−.39	**−2.06**	.58	−.00	−.17	.08	−.08	**−3.25**	.32
P	.66	**.01**	**.05**	.54	**.02**	.13	.61	.69	.85	.75	**.00**	.81
Genome size:												
r	.00	−.15	.31	.61	.57	.13	−.00	**−1.21**	**1.61**	**−.55**	.37	1.27
P	.57	.80	.47	.33	.53	.71	.54	**.00**	**.00**	**.03**	.66	.35
PC1:												
r	**−.02**	−.62	−.33	.54	1.29	−.28	−.00	−.46	.45	−.26	.67	.14
P	**.05**	.28	.44	.40	.15	.45	.82	.28	.25	.33	.43	.92
PC2:												
r	.00	.08	−.01	−.33	.35	−.42	−.00	−.15	−.06	−.07	.16	**3.81**
P	.91	.88	.97	.60	.70	.25	.99	.72	.88	.78	.85	**.00**
PC3:												
r	−.00	−.16	.19	−1.21	−1.45	.10	−.00	−.21	.03	−.19	.39	.71
P	.99	.78	.47	.06	.11	.79	.99	.59	.95	.44	.65	.60
R^2	.01	.15	.11	.14	.15	.11	.00	.36	.55	.22	.30	.18
Species richness:												
r	4.50	4.38	−.06	2.47	1.13	.24	.84	.51	−.35	–	.35	–
P	.06	.07	.90	.15	.51	.69	.40	.61	.72	–	.33	–
R^2	–	.12	.00	–	.05	.00	–	.03	.01	–	.02	–

Note: PC1 accounts for the general trend of lower temperature and lower UV in higher latitude. PC2 accounts for temperature variation that cannot be explained by latitude. PC3 accounts for UV variation that is independent to temperature and latitude. For protein-coding genes, synonymous (dS) and nonsynonymous (dN) substitution rates were estimated, as well as dN/dS. In addition, for the nuclear ribosomal RNA (rRNA) genes and *Xdh*, total substitution rate was estimated to facilitate comparison between the rRNA gene and the protein-coding gene. In the first section of the table ("Height" through "PC3" rows), each column shows the regression coefficients (*r*) and *P* values of different exploratory variables on a response variable, as well as the total effect size (R^2) of all the exploratory variables. The second section ("Species richness" rows) shows the regression coefficients (*r*), *P* values, and the total effect size (R^2) of nonsynonymous and synonymous substitution rates on species richness, as well as those values of dN/dS on species richness. Significant coefficients and their *P* values are indicated in boldface. The numbers of comparisons for each data set are as follows: chloroplast, 42; mitochondrial, 27; nuclear *Xdh*, 14; nuclear rRNA, 40. A dash indicates an analysis was not performed.

first principal component (PC1), indicating that variation along PC1 shows the general trend of lower temperature and lower UV at higher latitudes. On the second principal component (PC2), the loading of temperature (−0.82) is in the opposite direction to that expected for the latitudinal gradient, and temperature has higher absolute loadings than latitude (−0.37) and UV (0.45). On the third principal component (PC3), only latitude (0.73) and UV (0.68) have loadings, and the loading of UV is in the opposite direction to the latitudinal trend. Thus, PC2 and PC3 together account for environmental variation that is independent of latitudinal gradient, with PC2 primarily accounting for temperature (e.g., decrease in temperature with altitude) and PC3 primarily accounting for UV (e.g., increasing UV with altitude).

Results of ridge regression using the three PCs show that plant families at lower latitudes tend to have higher nonsynonymous substitution rates in their chloroplast genomes (tables 1, S5). Ridge regression on the three PCs also suggests that plant families that are exposed to higher average levels of UV tend to have lower substitution rates and higher dN/dS in chloroplast and mitochondrial genomes (table S5; this is also suggested by other regression analyses: see tables S6–S8), but this result is not consistent or significant in the reduced data set that includes genome size (tables 1, S9).

Are Rates Correlated across Genomes? We identified significant pairwise correlations between substitution rates across the three genomes (table 2). Within each of the genomes, synonymous and nonsynonymous substitution rates are correlated with each other (fig. 1). Synonymous substitution rates in the chloroplast genome are correlated with synonymous and nonsynonymous rates estimated from mitochondrial and nuclear genes and nuclear rRNA total substitution rates. Synonymous substitution rates in the mitochondrial genome are correlated with nonsynonymous substitution rates and total rRNA rates in the nuclear gene. Nonsynonymous substitution rates are not correlated across genomes.

Are Rates Correlated with Species Richness? For the path analysis, we tested four models (see appendix, fig. A1) but included only variables that show significant correlations in any of the regression analyses. So we include height in analyses of rates in mitochondrial, chloroplast, and nuclear rRNA genes and genome size only in the analysis of nuclear protein-coding gene. Because the environmental variables are highly intercorrelated, we use principal components in the path analysis.

For chloroplast genes, the best-fit model is model 3, which includes a direct connection between species richness and substitution rates, in addition to the connection caused by the species traits examined in the study (table 3). However, model 1 (substitution rate influences species richness) fits the data nearly as well, given that its AIC value does not differ much from model 3 ($\Delta AIC < 2$). We can reject model 4 (no direct association between substitution rates

Table 2: Pairwise correlation between substitution rates among three genomes: chloroplast (Cp), mitochondrial (Mt), and nuclear (*Xdh* and ribosomal RNA [rRNA]) genes

	Cp dN	Mt dS	Mt dN	*Xdh* dS	*Xdh* dN	rRNA
Cp dS:						
r	**.26**	**.38**	**.22**	**.65**	**.33**	**.31**
P	**.00**	**.00**	**.04**	**.00**	**.05**	**.00**
Cp dN:						
r	. . .	.06	−.02	.01	.00	.00
P	. . .	.56	.90	.97	1.00	.99
Mt dS:						
r	. . .	. . .	**.51**	.12	**.40**	**.28**
P	. . .	. . .	**.00**	.54	**.03**	**.01**
Mt dN:						
r	. . .	. . .	. . .	.07	.21	.17
P	. . .	. . .	. . .	.72	.27	.11
Xdh dS:						
r	. . .	. . .	. . .	. . .	**.34**	.27
P	. . .	. . .	. . .	. . .	**.04**	.10
Xdh dN:						
r	. . .	. . .	. . .	. . .	. . .	.12
P	. . .	. . .	. . .	. . .	. . .	.50

Note: Significant coefficients (*r*) are indicated in boldface, with *P* values given below. dN = nonsynonymous substitution rate; dS = synonymous substitution rate.

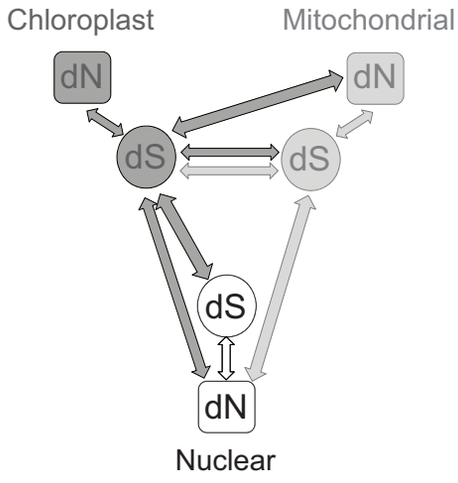


Figure 1: Schematic diagram of the significant correlations between the estimated substitution rates for chloroplast, nuclear, and mitochondrial genomes. For correlation coefficients and *P* values, see table 2. dN = nonsynonymous substitution rate; dS = synonymous substitution rate.

and species richness) for the chloroplast sequences. For the nuclear *Xdh* gene, the best-fitting model is model 1 (substitution rates influence species richness), and it fits the data significantly better than other models (all $\Delta AIC > 2$). For mitochondrial and nuclear rRNA genes, the best-fit model is model 4 (no direct association between rates and species richness), but none of the models can be rejected in favor of any of the others.

Figure 2 shows the results of significance tests on path coefficients. Synonymous substitution rates in the chloroplast and nuclear protein-coding genes are significantly correlated with species richness, and the analyses suggest that these correlations are not explained by an indirect covariation between species richness and rates with the family-average traits. Chloroplast substitution rates are also positively correlated with species richness in the ridge regression analysis, though this is not significant (dS, $P = .07$; dN, $P = .06$; table 1). These analyses do not provide any evidence of a link between species richness and substitution rates in the mitochondrial genes or the nuclear rRNA sequences (fig. 2).

This study focuses on the relationship between diversification rates and substitution rates and was not designed to test the influence of environmental or species traits on species richness. However, we note that temperature is also identified as a significant correlate of species richness identified in the regression analysis (tables 1, S5, S6, S8, S9). As suggested by PC2, this negative association between temperature and net diversification rate is independent of any latitudinal trend in temperature or diversification rates and is likely caused by other environmental gradients. This pattern may be an artifact of taking average measurements over the family geographic range: a family with more species might cover a larger area, which might contain a wider range of environmental conditions (such as variation in elevation), and thus family mean values calculated over the range might be lower than the mean values of locations in which members of the family are found. For example, a larger family range might include more high-elevation regions even if the species in the family are not distributed at high altitude, so the average temperature of the range may be lower than the actual temperature niche of the constituent species. Alternatively, if environmental heterogeneity drives diversification (e.g., montane speciation), then more heterogeneous environments might contain larger families, potentially making the mean temperature of more species-rich families lower than that of smaller and less widely distributed families.

Discussion

We show that plant traits can have a consistent effect on rates of molecular evolution across all three genomes, with patterns of substitutions suggesting that differences in mutation rate generate variation in both synonymous and

Table 3: Results of path analysis on the associations between substitution rates, traits, and species richness

	Chloroplast		Mitochondrial		*Xdh*		rRNA	
Model	χ^2 (*P*)	AIC (df)	χ^2 (*P*)	AIC (df)	χ^2 (*P*)	AIC (df)	χ^2 (*P*)	AIC (df)
1	7.01 (.54)	1,271 (8)	4.55 (.80)	807.1 (8)	**4.44 (.82)**	**262.1 (8)**	.73 (.98)	1,104 (5)
2	7.51 (.48)	1,272 (8)	5.47 (.71)	808.0 (8)	9.16 (.33)	266.8 (8)	.39 (1.00)	1,104 (5)
3	**4.67 (.79)**	**1,269 (8)**	4.28 (.83)	806.8 (8)	8.65 (.37)	266.3 (8)	.77 (.98)	1,104 (5)
4	18.70 (.04)	1,279 (10)	**7.78 (.65)**	**805.7 (10)**	15.77 (.11)	269.4 (10)	**1.45 (.96)**	**1,103 (6)**

Note: Four variants of the model are fitted to each data set. In model 1, substitution rates influence species richness. In model 2, species richness influences substitution rates. In model 3, species richness and substitutions are correlated. In model 4, species richness and substitutions are not correlated. Model fitness is evaluated by χ^2 (with *P* value in parentheses) and Aikake information criterion (AIC) value (with degrees of freedom in parentheses). Significant results are indicated in boldface. Figure 1 shows the detailed results of model 3.

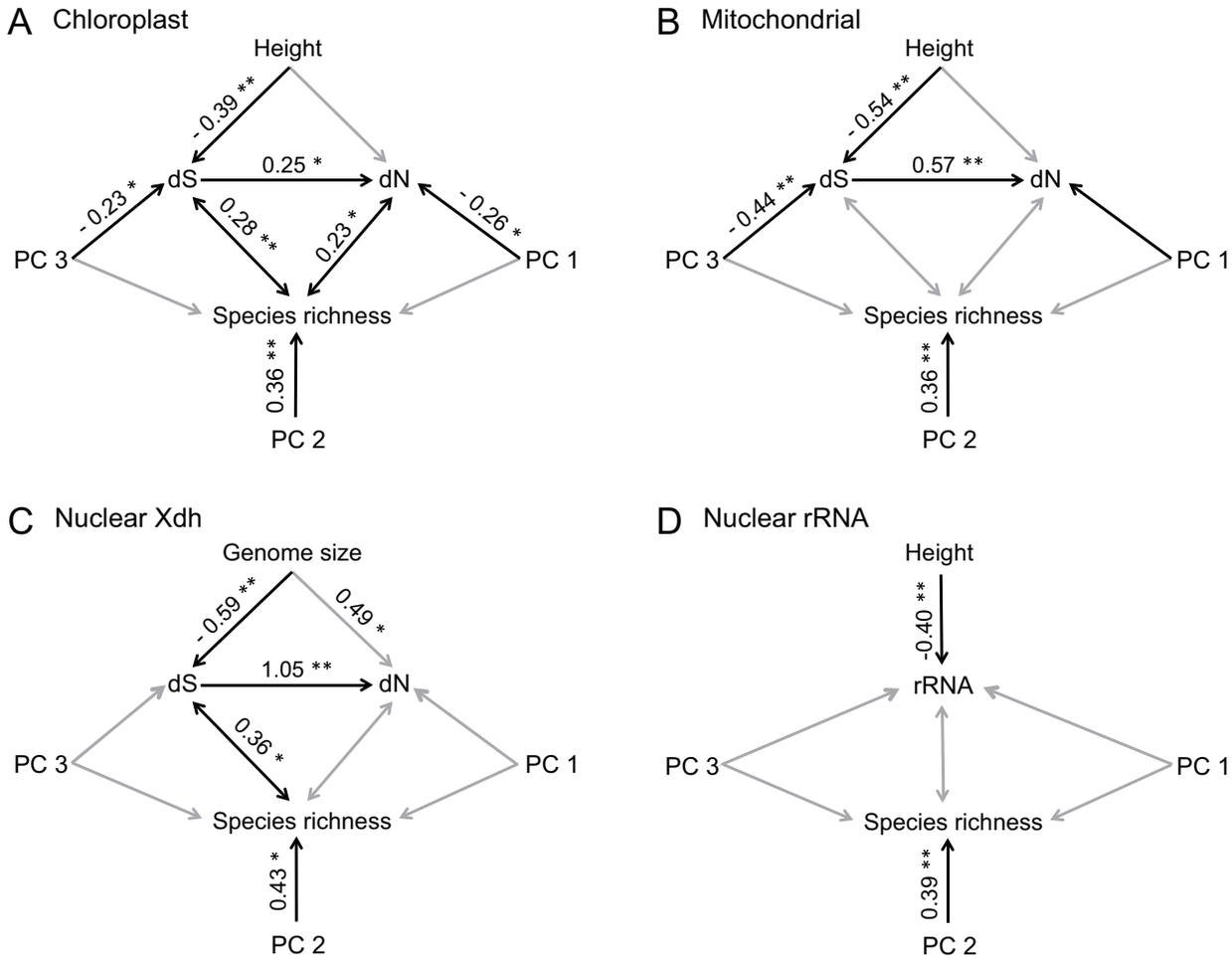


Figure 2: Results of path analysis. Single arrows indicate causal relationship, and double arrows indicate correlation. Arrows for correlations among explanatory variables are not shown, for the sake of clarity. Values along each path are path coefficients with significant level (one asterisk, $P < .05$; two asterisks, $P < .01$). Nonsignificant coefficients are not shown. Gray indicates a path with no significant coefficients in any data set. PC1–PC3 are the principal components of environmental variables (see table 1).

nonsynonymous rates in all three genomes. We do not find any significant associations between nonsynonymous rates estimated from the three different genomes, so these data provide no evidence that the patterns observed are largely driven by compensatory substitutions to maintain genome compatibility. For chloroplast and mitochondrial genes and nuclear rRNA, our analyses confirm that plant families with shorter average height have faster rates of molecular evolution. Analysis of the nuclear protein-coding gene, *Xdh*, suggests that greater family-average genome size is associated with lower mutation rate. There is some evidence of environmental influence on rates of molecular evolution, with the signal of increased substitution rates at lower latitudes in chloroplast and mitochondrial genomes. The path analyses support a link between species richness and substitution rates for the chloroplast and mitochondrial genomes. While the strongest support in the data is for either a direct influence of rates of molecular evolution on diversification rate or a direct link between them after accounting for the influence of family traits, we cannot reject the alternative explanation that differences in diversification rates drive differences in substitution rates. However, given that the patterns of rate variation are most strongly reflected in the synonymous substitution rate, which reflects the mutation rate, we think the most likely explanation is that differences in mutation rates between lineages, at least partly driven by differences in family-average traits, influence both synonymous and nonsynonymous rates of change across all three genomes, which in turn drive greater rates of divergence between populations, generating higher

rates of diversification (fig. 3). We will discuss each of these links in more detail.

Plant Height and Mutation Rate

The clearest and most consistent pattern that emerges from our analyses is that plant families with a shorter average height have faster rates of molecular evolution. This is consistent with previous analysis of these nuclear rRNA and chloroplast sequences (Lanfear et al. 2013), but our study extends this result to all three plant genomes. Height is related to synonymous substitution rate and dN/dS in the chloroplast and mitochondrial genomes and to substitution rate in the nuclear rRNA sequences. Since synonymous substitution rate (dS) is primarily influenced by the mutation rate, this suggests that taller plants have lower rates of mutation per year.

Size is an important determinant of rate of molecular evolution for many taxa, a pattern often attributed to an assumed difference in the number of genome replications per unit time (Sarich and Wilson 1973; Ohta 1993; Bromham et al. 1996; Thomas et al. 2010). While the relationship between body size, cell divisions, and generation time is complicated in plants, we can make predictions about the relationship between plant height and genome turnover. Larger plants often have lower absolute growth rates (Petit and Hampe 2006) and thus undergo fewer cell divisions per unit time than shorter plants (Lanfear et al. 2013). Fewer cell divisions means fewer opportunities for replication errors to occur, so for the same per-replication mutation rate, we would expect a taller plant to have fewer DNA replication errors per unit time than a shorter plant. The effect of number of cell divisions on the per-unit-time mutation rate can be demonstrated through male-driven evolution in paternally inherited organelle genomes: pollen production requires more cell divisions than ovule production, so genes passed through the male line accumulate more copy errors per unit time (Whittle and Johnston 2002).

Reduction in the per-unit-time mutation rate is expected for taller plants on purely mechanistic grounds (fewer opportunities for replication errors), but it may be enhanced by the influence of selection on mutation rates. In plants, germ lines are not sequestered early in development but arise from the apical meristem (growing tip) when it becomes a reproductive apex. So taller plants will tend to have more cell generations per reproductive generation (Schultz and Scofield 2009). The taller the plant, the more cell divisions occur between the seed and the apical meristem, so the more opportunity for mutations to occur per generation (Petit and Hampe 2006; Bobiwash et al. 2013). Therefore, a taller plant is at greater risk of accumulating deleterious mutations during its lifetime. These deleterious mutations may or may not be selected against in somatic tissues (Klekowski and Godfrey 1989; Marcotrigiano 2000; Morgan 2001), but if they accumulate in a cell line that gives rise to reproductive cells, then offspring fitness may be reduced. Populations formed by vegetative propagation can accumulate somatic mutations over time (Warren 2009), even if these mutations are deleterious (Gross et al. 2012). One way for taller plants to avoid the mutation costs of more cell divisions per generation is to reduce the error rate per replication. Selection pressure to reduce the per-genome-replication mutation rate could result in a lower per-site, per-replication mutation rate, which should be reflected in the rate of synonymous substitutions per site (Sung et al. 2012). Our results are consistent with this hypothesis, because we see a consistent negative relationship between synonymous substitution rates and height in the mitochondrial and chloroplast genes. We do not detect this relationship for the single protein-coding nuclear gene tested, although the nuclear

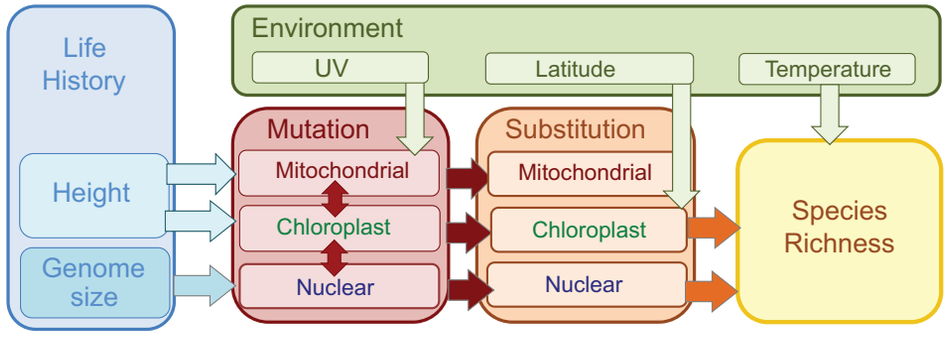


Figure 3: Schematic diagram of the inferred links between life history, environmental variables, mutation rates, substitution rates, and species richness suggested by the results of our analysis. Each arrow is based on a significant correlation in one of the analyses, but not all significant correlations are shown.

rRNA genes show a negative association between substitution rate and height.

Genome Size and Molecular Evolution

Families with larger average genomes have lower synonymous substitution rates and higher dN/dS in the nuclear protein-coding gene. One possible explanation for the link between genome size and synonymous substitution rate is that, if increased genome size correlates with larger cells and reduction in growth rates, then plants with larger genomes might have fewer genome replications per unit time and, therefore, less opportunity to accumulate mutations. But cell division rates should affect all three genomes, and we see no evidence of a link between genome size and organelle mutation rates. Alternatively, increased genome size may result in selection for lower mutation rates to reduce the per-genome number of mutations per generation. This might explain why we see a link between nuclear genome size and substitution rates in the nuclear protein-coding gene although this relationship is not reflected in organelle substitution rates, since organelle genomes vary less in size between families and are copied using, at least in part, their own DNA replication machinery.

While some variation in total genome size is related to genome duplications, much of the variation is accounted for by differences in the amount of repetitive DNA, including transposable elements (Bennetzen et al. 2005; Grover and Wendel 2010; Tenaillon et al. 2010). Given that mutations in transposon sequences should typically carry no fitness cost to the host, it has been suggested that the relevant parameter is the mutation rate per replication per base of the effective genome, which is that part of the genome where mutations can produce deleterious effects on which selection can act (Drake et al. 1998). An exception is the fitness cost of gain-of-function mutations in noncoding DNA: the more DNA there is in the genome, the more chance of a mutation that accidentally changes the regulation and maintenance of the working parts of the genome (Lynch 2007). Unfortunately, the coding fraction of the genome and the relative proportion of transposable elements is not available for sufficient species to allow us to make the distinction between total genome size and effective genome size in this study.

Genome size might be indirectly linked to rates of molecular evolution through effective population size. If the selective cost of increase in transposon copy number is typically only slightly deleterious, then we can expect more effective reduction in transposon copies in larger populations where selection against slightly deleterious alleles is most effective (Grover and Wendel 2010). If this is the case, then lineages characterized by smaller populations should accumulate more transposon copies and more nearly neutral substitutions, leading to an increase in both genome size and dN/dS. This is consistent with our observation of higher dN/dS in families with larger average genome size (table 1). However, this effect does not provide an easy explanation for the observation that plant families with larger average genome sizes also have lower mutation rates. Furthermore, the evidence for a link between genome size and effective population size in plants has been mixed (Lockton et al. 2008; Whitney et al. 2010; Ai et al. 2012).

Mutation Rate and Diversification

Our finding of a link between synonymous substitution rate and species richness in this analysis is consistent with previous findings in plants (Barraclough and Savolainen 2001; Duchene and Bromham 2013; Lanfear et al. 2013), fish (Venditti and Pagel 2009), and birds and reptiles (Eo and DeWoody 2010; Lanfear et al. 2010*a*). Mutation rate might affect the net rate of diversification by contributing to hybrid incompatibility, which can arise from any noncompatible substitutions, even of alleles that have no fitness cost or benefit in their own population. The more substitutions acquired by either population, the less chance that genomes drawn from the different populations could be combined in one individual to produce a viable hybrid (Orr 1995). These Dobzhansky-Muller incompatibilities are typically modeled with a relatively small number of loci (Rieseberg and Willis 2007) but could occur through the continuous accumulation of substitutions of small effect across a large number of loci (Hua and Wiens 2013). Even if the initial formation of reproductive isolation is due to particular changes in few key genes, the accumulation of substitutions in each of the separated lineages thereafter codifies the reproductive isolation (Coyne and Orr 2004), so the number of loci contributing to isolation might increase with the square of time or even more rapidly (Matute et al. 2010). If many loci across the genome can contribute to hybrid incompatibility, then the genome-wide substitution rate could influence the speed of evolution of barriers to reproduction between populations. Because the accumulation of incompatible alleles does not need to be symmetrical (Welch 2004), a faster rate of molecular evolution in one lineage should speed the evolution of reproductive isolation between the two.

Faster substitution rates in all three plant genomes might contribute to cytonuclear conflict, which plays an important role in some cases of hybrid incompatibility (Levin 2003; Greiner et al. 2011; Greiner and Bock 2013), although its general contribution to patterns of speciation is not yet clear (Crespi and Nosil 2012). For example, changes to mitochondrial genes that alter their interaction with nuclear-produced proteins can induce cytoplasmic male sterility (Ma 2013), but male function can be restored by compensa-

tory changes to nuclear genes (Luo et al. 2013). Organelle-nucleus interactions could potentially involve many different genes: for example, scores of nuclear genes must work with the dozen or so mitochondrial genes on the oxidative phosphorylation pathway (Burton and Barreto 2012). Critically, these are housekeeping genes typical of those included in phylogenetic studies. For example, substitutions in the mitochondrial *nad* and *atp1* genes, included in this study, have been shown to drive cytoplasmic male sterility (Greiner and Bock 2013; Yoshimi et al. 2013). However, we did not find any direct support for cytonuclear incompatibility in the sequences analyzed for this study, as we see no evidence of a link between nonsynonymous substitution rates across genomes.

Is it possible that the causal arrow runs the other way, so that higher diversification rates drive increased synonymous substitution rates? It is difficult to imagine how diversification rate could directly impact mutation rate, though it may do so indirectly through reduction in effective population size. Population size could be negatively associated with diversification rate if speciation tends to divide populations. Selection on mutations of small selective effect is less efficient in small populations, so slightly beneficial changes to DNA repair may fail to go to fixation, but slightly deleterious decreases in replication fidelity may be fixed by drift (Lynch 2007, 2010). If this were a common phenomenon, then we would expect increased mutation rates to be associated with smaller population sizes. Our results are in the opposite direction, with increase in dS associated with decrease in dN/dS (table 1). This suggests that higher mutation rates are not associated with relaxed selection in small populations, unless the increase in dN across the genome due to fixation of nearly neutral alleles by drift is much greater in magnitude than the increase in dS due to erosion of DNA repair systems.

Environmental Effects on Molecular Evolution

Our results are compatible with a latitudinal gradient in organelle substitution rates (Wright et al. 2006, 2010, 2011; Gillman et al. 2009; Lourenço et al. 2012; Lanfear et al. 2013), though these associations are inconsistently identified in our analyses and not always significant. It is important to note that our comparisons were not chosen specifically to test the latitudinal gradient: sister families may have overlapping latitudinal ranges, limiting the power to detect latitudinal patterns (Cardillo 1999). However, we can ask whether the patterns we have detected in this data are consistent with different hypothesized links between environmental conditions and rate of molecular evolution.

Environmental variables have been proposed as both direct and indirect determinants of rates of molecular evolution in plants. The indirect link between environment and molecular evolution is primarily discussed in terms of the latitudinal gradient in plant productivity and diversity (Brown 2014). Energy availability, determined by temperature, light, and water, has been proposed as the primary driver of patterns of species richness through its effects on plant growth and, therefore, on primary productivity (Hawkins et al. 2003). Growth rates could influence the rate of molecular evolution by affecting the number of cell divisions per unit time and, therefore, rates of genome turnover. Consistent with this hypothesis, temperature has been identified as a correlate of rate of molecular evolution in plants (Davies et al. 2004*b*), as has water availability (Goldie et al. 2010). Average temperature and length of growing season tend to decrease with increasing latitude (De Frenne et al. 2013), so it is reasonable to expect that plants at low latitude can undergo more cell divisions per year than plants at higher latitudes and thus accumulate more copy errors in their genomes (Gillman et al. 2009; Gillman and Wright 2014). If the latitudinal effect on rates of molecular evolution was primarily driven by higher growth rates causing more replication errors per unit time, then we would expect to see this reflected in both the synonymous and nonsynonymous substitution rates. But we detect a latitudinal gradient only in the rate of nonsynonymous substitutions in the chloroplast genes.

A direct link between environmental factors and molecular evolution has been suggested by proposing that aspects of the environment, particularly temperature and UV, could have a directly mutagenic effect on the genome (Rohde 1992; Willis et al. 2009; Flenley 2011). UV-B can damage cellular structures and induce mutations and genome rearrangements. Since UV exposure increases toward the equator, plants living at low latitudes might experience more UV-induced DNA damage. On this basis, UV-induced mutagenesis has been cited as a potential driver of species richness in plants (Willis et al. 2009). However, our data provide no support for this hypothesis and indeed suggest the opposite pattern: average UV exposure is negatively correlated with organelle mutation rate. The role of UV as a determinant of substitution rate is also undermined by patterns of variation in rate of molecular evolution with altitude (Dowle et al. 2013). UV exposure increases with altitude, yet several studies have shown that rates of molecular evolution decrease with altitude (Gillman et al. 2009; Wright et al. 2010).

The negative relationship between UV exposure and mutation rate might be explained by the adjustment of DNA repair in response to mutagen levels. There are many DNA repair pathways that ameliorate the effect of UV, and the efficiency of these pathways can vary between species, impacting on the mutation rate (Lucas-Lledo and Lynch 2009). Lineages may adapt to different environments through changes to the activity or efficiency of DNA repair

mechanism; for example, species in low UV environments may lose photolyase enzymes (Eisen and Hanawalt 1999). Plants living at high altitude may be adapted to resist UV-B damage (Sullivan et al. 1992), for example, through changes to DNA repair systems (Albarracín et al. 2012). Moreover, UV acclimation responses can be inducible, employed at a level appropriate to conditions (Jansen et al. 1998), including through moderation of DNA repair (Ries et al. 2000). For example, photoreactivation, which repairs thymine dimers caused by UV light, is driven by energy from UV-A radiation, so the capacity for repairing UV damage may increase with the level of insolation (Jansen et al. 1998). Although plant families distributed at lower average latitudes may be exposed to greater levels of UV-B, the effect of UV-B on mutation rates may be ameliorated by investment in cellular mechanisms that prevent or repair UV-induced DNA damage. So while we do find some support for a link between environment conditions and rate molecular evolution, we do not see any clear pattern that environment directly influences the mutation rate (e.g., through increased mutagenic burden in warmer, brighter environments) nor indirectly (e.g., increasing mutation rates in environmental areas likely to stimulate higher growth rates).

Population Size, Selection, and Diversification

It has been suggested that the link between diversification and rates of molecular evolution is mediated by changes in population size accompanying speciation, for example, due to the founder effect as a new habitat is colonized or the reproductive isolation of small peripheral populations (Webster et al. 2003; Pagel et al. 2006; Venditti and Pagel 2009). Following population subdivision, neutral alleles carried over from the larger parent population will be either lost or fixed, potentially providing a burst of substitutions (Venditti and Pagel 2009). In smaller populations, negative selection will be less efficient at removing slightly deleterious mutations, so these alleles will have an increased chance of going to fixation by drift. If slightly deleterious mutations make up a nontrivial proportion of all mutations (Eyre-Walker and Keightley 2007), then small populations will tend to have higher rates of nonsynonymous substitution relative to the rate of synonymous substitutions (Charlesworth 2009; Woolfit 2009).

Comparative studies have demonstrated that lineages with reduced population size can have consistently higher dN/dS (e.g., Woolfit and Bromham 2003, 2005). But, thus far, there has been little direct evidence that population size changes are driving the link between substitution rates and net diversification rate: studies have either not directly looked for the signature of population size change (Barraclough et al. 1996; Webster et al. 2003; Pagel et al. 2006; Lancaster 2010) or, if they have looked for such evidence, have failed to find it (Barraclough and Savolainen 2001; Venditti and Pagel 2009; Lanfear et al. 2010*a*; Duchene and Bromham 2013; Lanfear et al. 2013). While we find a consistent association between family-average traits and dN/dS in these data (height for the organelle genes, genome size for the nuclear *Xdh* gene), we do not see a positive association between dN/dS and species richness. This may be because there is no consistent pattern of population size reduction associated with speciation or that any such effects are too transient to have a significant impact on family-average substitution rates. Or it may be that other factors that influence substitution rates override or dissemble any effect of population size associated with speciation.

An alternative explanation for the link between substitution rate and net diversification rate observed in this study is that selection on these genes is either a cause of, or a response to, speciation. While speciation may be accompanied by selection to adapt to a new habitat or reinforce reproductive isolation, it seems unlikely that the genes typically included in phylogenetic studies would all be subject to strong directional selection in a newly isolated population. If widespread positive selection across the genome were to account for the link between net diversification rate and rate of molecular evolution, then we would expect to see this reflected in raised dN/dS in species-rich lineages. Instead, phylogenetic studies of the relationship between molecular rates and net diversification rates can be more easily interpreted as reflecting genome-wide processes such as mutation rate and drift, rather than locus-specific selection for adaptation or isolating mechanisms.

Conclusion

While details of the processes of diversification are lost when taking the broad, comparative view, phylogenetic studies do permit the search for general patterns that may highlight some common mechanisms. This study of rates of molecular evolution in flowering plant families provides evidence that families of taller plants are characterized by lower mutation rates per unit of time than families of shorter plants in genes from the nuclear, chloroplast, and mitochondrial genomes. Substitution rates estimated from protein-coding sequences from the chloroplast and nuclear genomes are associated with differential rates of diversification in flowering plant families. Environmental factors are also associated with variation in rates of molecular evolution, but we find no support in these data for the hypothesis that this is driven by higher rates of mutation at lower latitudes, whether due to a direct association (impact of temperature or UV on mutation generation) or indirect

connection (raised growth rates increasing the accumulation of replication errors). Our findings are consistent with a role for the accumulation of substitutions across all three genomes in driving diversification through their cumulative action on genetic incompatibility between sister populations and thus the pace of formation of species. This finding links change at the genomic level to species characteristics and biodiversity generation, highlighting the continuity of processes of mutation (generation of variation), microevolution (population divergence), and macroevolution (lineage diversification).

Literature Cited

Abbott, R., D. Albach, S. Ansell, J. W. Arntzen, S. J. E. Baird, N. Bierne, J. Boughman, et al. 2013. Hybridization and speciation. Journal of Evolutionary Biology 26:229–246.

Ai, B., Z.-S. Wang, and S. Ge. 2012. Genome size is not correlated with effective population size in the *Oryza* species. Evolution 66:3302–3310.

Albarracín, V. H., G. P. Pathak, T. Douki, J. Cadet, C. D. Borsarelli, W. Gärtner, and M. E. Farias. 2012. Extremophilic *Acinetobacter* strains from high-altitude lakes in Argentinean Puna: remarkable UV-B resistance and efficient DNA damage repair. Origins of Life and Evolution of Biospheres 42:201–221.

Barraclough, T. G., P. H. Harvey, and S. Nee. 1996. Rate of *rbcL* gene sequence evolution and species diversification in flowering plants (angiosperms). Proceedings of the Royal Society B: Biological Sciences 263:589–591.

Barraclough, T. G., and V. Savolainen. 2001. Evolutionary rates and species diversity in flowering plants. Evolution 55:677–683.

Bateson, W. 1894. Materials for the study of variation treated with especial regard to discontinuity in the origin of species. Macmillan, London.

Beaulieu, J. M., I. J. Leitch, S. Patel, A. Pendharkar, and C. A. Knight. 2008. Genome size is a strong predictor of cell size and stomatal density in angiosperms. New Phytologist 179:975–986.

Beaulieu, J. M., A. T. Moles, I. J. Leitch, M. D. Bennett, J. B. Dickie, and C. A. Knight. 2007. Correlated evolution of genome size and seed mass. New Phytologist 173:422–437.

Beaulieu, J. M., S. A. Smith, and I. J. Leitch. 2010. On the tempo of genome size evolution in angiosperms. Journal of Botany, http://dx.doi.org/10.1155/2010/989152.

Bennett, M. D. 1972. Nuclear DNA content and minimum generation time in herbaceous plants. Proceedings of the Royal Society B: Biological Sciences 181:109–135.

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

———. 2012. Plant DNA C-values database. Release 6.0. December 2012. http://data.kew.org/cvalues/.

Bennetzen, J. L., J. Ma, and K. M. Devos. 2005. Mechanisms of recent genome size variation in flowering plants. Annals of Botany 95: 127–132.

Bobiwash, K., S. Schultz, and D. Schoen. 2013. Somatic deleterious mutation rate in a woody plant: estimation from phenotypic data. Heredity 111:338–344.

Bromham, L. 2009. Why do species vary in their rate of molecular evolution? Biology Letters 5:401–404.

Bromham, L., X Hua, R. Lanfear, and P. F. Cowman. 2015. Exploring the relationships between mutation rates, life history, genome size, environment, and species richness in flowering plants. American Naturalist, Dryad Digital Repository, http://dx.doi.org/10.5061/dryad.31614.

Bromham, L., A. Rambaut, and P. H. Harvey. 1996. Determinants of rate variation in mammalian DNA sequence evolution. Journal of Molecular Evolution 43:610–621.

Brown, J. H. 2014. Why are there so many species in the tropics? Journal of Biogeography 41:8–22.

Burleigh, J. G., K. Hilu, and D. Soltis. 2009. Inferring phylogenies with incomplete data sets: a 5-gene, 567-taxon analysis of angiosperms. BMC Evolutionary Biology 9:61.

Burton, R. S., and F. S. Barreto. 2012. A disproportionate role for mtDNA in Dobzhansky-Muller incompatibilities? Molecular Ecology 21:4942–4957.

Cardillo, M. 1999. Latitude and rates of diversification in birds and butterflies. Proceedings of the Royal Society B: Biological Sciences 266:1221–1225.

Charlesworth, B. 2009. Effective population size and patterns of molecular evolution and variation. Nature Reviews Genetics 10:195–205.

Coyne, J. A., and H. A. Orr. 2004. Speciation, Sinauer, Sunderland, MA.

Crespi, B., and P. Nosil. 2012. Conflictual speciation: species formation via genomic conflict. Trends in Ecology and Evolution 28:48–57.

Cule, E. 2012. ridge: Ridge Regression with automatic selection of the penalty parameter. Version 2.1. R package.

Davies, T. J., T. G. Barraclough, V. Savolainen, and M. W. Chase. 2004*a*. Environmental causes for plant biodiversity gradients. Philosophical Transactions of the Royal Society B: Biological Sciences 359:1645–1656.

Davies, T. J., V. Savolainen, M. W. Chase, J. Moat, and T. G. Barraclough. 2004*b*. Environmental energy and evolutionary rates in flowering plants. Proceedings of the Royal Society B: Biological Sciences 271:2195–2220.

De Frenne, P., B. J. Graae, F. Rodríguez-Sánchez, A. Kolb, O. Chabrerie, G. Decocq, H. De Kort, et al. 2013. Latitudinal gradients as natural laboratories to infer species' responses to temperature. Journal of Ecology 101:784–795.

Dowle, E., M. Morgan-Richards, and S. Trewick. 2013. Molecular evolution and the latitudinal biodiversity gradient. Heredity 110: 501–510.

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

Drummond, A. J., B. Ashton, S. Buxton, M. Cheung, A. Cooper, C. Duran, M. Field, et al. 2011. Geneious. http://www.geneious.com.

Duchene, D., and L. Bromham. 2013. Rates of molecular evolution and diversification in plants: chloroplast substitution rates correlate with species richness in the Proteaceae. BMC Evolutionary Biology 13:65.

Eisen, J. A., and P. C. Hanawalt. 1999. A phylogenomic study of DNA repair genes, proteins, and processes. Mutation Research/DNA Repair 435:171–213.

Eo, S. H., and J. A. DeWoody. 2010 Evolutionary rates of mitochondrial genomes correspond to diversification rates and to contemporary species richness in birds and reptiles. Proceedings of the Royal Society B: Biological Sciences 277:3587–3592.

Eyre-Walker, A., and P. D. Keightley. 2007. The distribution of fitness effects of new mutations. Nature Reviews Genetics 8:610–618.

Ezard, T. H. G., G. H. Thomas, and A. Purvis. 2013. Inclusion of a near-complete fossil record reveals speciation-related molecular evolution. Methods in Ecology and Evolution 4:745–753.

Flenley, J. R. 2011. Why is pollen yellow? and why are there so many species in the tropical rain forest? Journal of Biogeography 38: 809–816.

Gillman, L. N., D. J. Keeling, H. A. Ross, and S. D. Wright. 2009. Latitude, elevation and the tempo of molecular evolution in mammals. Proceedings of the Royal Society B: Biological Sciences 276: 3353–3359.

Gillman, L. N., and S. D. Wright. 2014. Species richness and evolutionary speed: the influence of temperature, water and area. Journal of Biogeography 41:39–51.

Goldie, X., L. Gillman, M. Crisp, and S. Wright. 2010. Evolutionary speed limited by water in arid Australia. Proceedings of the Royal Society B: Biological Sciences 277:2645–2653.

Goldie, X., R. Lanfear, and L. Bromham 2011. Diversification and the rate of molecular evolution: no evidence of a link in mammals. BMC Evolutionary Biology 11:286.

Goldman, N., and Z. Yang. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. Molecular Biology and Evolution 11:725–736.

Gorniak, M., O. Paun, and M. W. Chase. 2010. Phylogenetic relationships within Orchidaceae based on a low-copy nuclear coding gene, *Xdh*: congruence with organellar and nuclear ribosomal DNA results. Molecular Phylogenetics and Evolution 56:784–795.

Greiner, S., and R. Bock. 2013. Tuning a ménage à trois: co-evolution and co-adaptation of nuclear and organellar genomes in plants. BioEssays 35:354–365.

Greiner, S., U. W. E. Rauwolf, J. Meurer, and R. G. Herrmann. 2011. The role of plastids in plant speciation. Molecular Ecology 20: 671–691.

Gross, C. L., P. A. Nelson, A. Haddadchi, and M. Fatemi. 2012. Somatic mutations contribute to genotypic diversity in sterile and fertile populations of the threatened shrub, *Grevillea rhizomatosa* (Proteaceae). Annals of Botany 109:331–342.

Grover, C. E., and J. F. Wendel. 2010. Recent insights into mechanisms of genome size change in plants. Journal of Botany 2010: 382732.

Hawkins, B. A., R. Field, H. V. Cornell, D. J. Currie, J.-F. Guégan, D. M. Kaufman, J. T. Kerr, et al. 2003. Energy, water, and broad-scale geographic patterns of species richness. Ecology 84:3105–3117.

Herben, T., J. Suda, J. Klimešová, S. Mihulka, P. Říha, and I. Šímová. 2012. Ecological effects of cell-level processes: genome size, functional traits and regional abundance of herbaceous plant species. Annals of Botany 110:1357–1367.

Hodgson, J. G., M. Sharafi, A. Jalili, S. Diaz, G. Montserrat-Marti, C. Palmer, B. Cerabolini, et al. 2010. Stomatal vs. genome size in angiosperms: the somatic tail wagging the genomic dog? Annals of Botany 105:573–584.

Hua, X., and J. J. Wiens. 2013. How does climate influence speciation. American Naturalist 182:1–12.

Hugall, A. F., and M. S. Lee. 2007. The likelihood node density effect and consequences for evolutionary studies of molecular rates. Evolution 61:2293–2307.

Izenman, A. J. 2008. Modern multivariate statistical techniques: regression, classification, and manifold learning. Springer, New York.

Jansen, M. A., V. Gaba, and B. M. Greenberg. 1998. Higher plants and UV-B radiation: balancing damage, repair and acclimation. Trends in Plant Science 3:131–135.

Kang, M., J. Tao, J. Wang, C. Ren, Q. Qi, Q. Ä. Xiang, and H. Huang. 2014. Adaptive and nonadaptive genome size evolution in Karst endemic flora of China. New Phytologist 202:1371–1381.

Katoh, K., K. Misawa, K. Ä. Kuma, and T. Miyata. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research 30:3059–3066.

Kimura, M. 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge.

Klekowski, E. J., and P. J. Godfrey. 1989. Ageing and mutation in plants. Nature 340:389–391.

Knight, C. A., and J. M. Beaulieu. 2008. Genome size scaling through phenotype space. Annals of Botany 101:759–766.

Kraaijeveld, K. 2010. Genome size and species diversification. Evolutionary Biology 37:227–233.

Lancaster, L. T. 2010. Molecular evolutionary rates predict both extinction and speciation in temperate angiosperm lineages. BMC Evolutionary Biology 10:162.

Lanfear, R., L. Bromham, and S. Y. Ho. 2011. Molecular evolution and diversification: counting species is better than counting nodes when the phylogeny is unknown. Proceedings of the National Academy of Sciences of the USA 108:E85–E86.

Lanfear, R., S. Y. W. Ho, T. J. Davies, A. T. A. Moles, L. N. G. Swenson, L. Warman, A. E. Zanne, et al. 2013. Taller plants have lower rates of molecular evolution: the rate of mitosis hypothesis. Nature Communications 4:1879.

Lanfear, R., S. Y. W. Ho, D. Love, and L. Bromham. 2010*a*. Mutation rate influences diversification rate in birds. Proceedings of the National Academy of Sciences of the USA 107:20423–20428.

Lanfear, R., J. J. Welch, and L. Bromham. 2010*b*. Watching the clock: studying variation in rates of molecular evolution. Trends in Ecology and Evolution 25:495–503.

Lavergne, S. B., N. J. Muenke, and J. Molofsky. 2010. Genome size reduction can trigger rapid phenotypic evolution in invasive plants. Annals of Botany 105:109–116.

Levin, D. A. 2003. The cytoplasmic factor in plant speciation. Systematic Botany 28:5–11.

Lleras, C. 2005. Path analysis. Pages 25–30 *in* K. Kempf-Leonard, ed. Encyclopedia of social measurement. Academic Press, San Diego, CA.

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*. Proceedings of the National Academy of Sciences of the USA 105:13965–13970.

Lourenço, J., S. Glémin, Y. Chiari, and N. Galtier. 2012. The determinants of the molecular substitution process in turtles. Journal of Evolutionary Biology 26:38–50.

Lucas-Lledo, J. I., and M. Lynch. 2009. Evolution of mutation rates: phylogenomic analysis of the photolyase/cryptochrome family. Molecular Biology and Evolution 26:1143–1153.

Luo, D. P., H. Xu, Z. L. Liu, J. X. Guo, H. Y. Li, L. T. Chen, C. Fang, et al. 2013. A detrimental mitochondrial-nuclear interaction causes cytoplasmic male sterility in rice. Nature Genetics 45:573–577.

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

———. 2010. Evolution of the mutation rate. Trends in Genetics 26:345–352.

Ma, H. 2013. A battle between genomes in plant male fertility. Nature Genetics 45:472–473.

Marcotrigiano, M. 2000. Herbivory could unlock mutations sequestered in stratified shoot apices of genetic mosaics. American Journal of Botany 87:355–361.

Matute, D. R., I. A. Butler, D. A. Turissini, and J. A. Coyne. 2010. A test of the snowball theory for the rate of evolution of hybrid incompatibilities. Science 329:1518–1521.

Mevik, B., R. Wehrens, and K. H. Liland. 2011. pls: partial least squares and principal component regression. Version 2.3-0. R package.

Moles, A. T., D. I. Warton, L. Warman, N. G. Swenson, S. W. Laffan, A. E. Zanne, A. Pitman, et al. 2009. Global patterns in plant height. Journal of Ecology 97:923–932.

Morgan, M. T. 2001. Consequences of life history for inbreeding depression and mating system evolution in plants. Proceedings of the Royal Society B: Biological Sciences 268:1817–1824.

Morton, C. M. 2011. Newly sequenced nuclear gene (*Xdh*) for inferring angiosperm phylogeny 1. Annals of the Missouri Botanical Garden 98:63–89.

Nosil, P., and D. Schluter. 2011. The genes underlying the process of speciation. Trends in Ecology and Evolution 26:160–167.

Ohta, T. 1993. An examination of the generation time effect on molecular evolution. Proceedings of the National Academy of Sciences of the USA 90:10676–10680.

Orr, H. A. 1995. The population genetics of speciation: the evolution of hybrid incompatibilities. Genetics 139:1805–1813.

Pagel, M., C. Venditti, and A. Meade. 2006. Large punctuational contribution of speciation to evolutionary divergence at the molecular level. Science 314:119–121.

Paradis, E., J. Claude, and K. Strimmer. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20:289–290.

Petit, R. J., and A. Hampe. 2006. Some evolutionary consequences of being a tree. Annual Review of Ecology, Evolution, and Systematics 37:187–214.

R Development Core Team. 2013. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.

Ries, G., W. Heller, H. Puchta, H. Sandermann, H. K. Seidlitz, and B. Hohn. 2000. Elevated UV-B radiation reduces genome stability in plants. Nature 406:98–101.

Rieseberg, L. H., and J. H. Willis. 2007. Plant speciation. Science 317:910–914.

Rohde, K. 1992. Latitudinal gradients in species-diversity: the search for the primary cause. Oikos 65:514–527.

Rosseel, Y. 2012. lavaan: an R package for structural equation modeling. Journal of Statistical Software 48:1–36.

Sanderson, M. J. 1990. Estimating rates of speciation and evolution: a bias due to homoplasy. Cladistics 6:387–391.

Sarich, V. M., and A. C. Wilson. 1973. Generation time and genomic evolution in primates. Science 179:1144–1147.

Schultz, S. T., and D. G. Scofield. 2009. Mutation accumulation in real branches: fitness assays for genomic deleterious mutation rate and effect in large-statured plants. American Naturalist 174:163–175.

Shipley, B. 2002. Cause and correlation in biology: a user's guide to path analysis, structural equations and causal inference. Cambridge University Press, Cambridge.

Smith, S. A., and M. J. Donoghue. 2008. Rates of molecular evolution are linked to life history in flowering plants. Science 322:86–89.

Sokal, R. R., and F. J. Rohlf. 1995. Biometry: the principles and practice of statistics in biological research. Freeman, San Francisco.

Soltis, D. E., S. A. Smith, N. Cellinese, K. J. Wurdack, D. C. Tank, S. F. Brockington, N. F. Refulio-Rodriguez, et al. 2011. Angiosperm phylogeny: 17 genes, 640 taxa. American Journal of Botany 98:704–730.

Strasburg, J. L., N. A. Sherman, K. M. Wright, L. C. Moyle, J. H. Willis, and L. H. Rieseberg. 2012. What can patterns of differentiation across plant genomes tell us about adaptation and speciation? Philosophical Transactions of the Royal Society B: Biological Sciences 367:364–373.

Sullivan, J. H., A. H. Teramura, and L. H. Ziska. 1992. Variation in UV-B sensitivity in plants from a 3,000-m elevational gradient in Hawaii. American Journal of Botany 79:737–743.

Sung, W., A. E. Tucker, T. G. Doak, E. Choi, W. K. Thomas, and M. Lynch. 2012. Extraordinary genome stability in the ciliate *Paramecium tetraurelia*. Proceedings of the National Academy of Sciences of the USA 109:19339–19344.

Tenaillon, M. I., J. D. Hollister, and B. S. Gaut. 2010. A triptych of the evolution of plant transposable elements. Trends in Plant Science 15:471–478.

Thomas, J. A., J. J. Welch, R. Lanfear, and L. Bromham. 2010. A generation time effect on the rate of molecular evolution in invertebrates. Molecular Biology and Evolution 27:1173–1180.

Venditti, C., A. Meade, and M. Pagel. 2008. Detecting the node-density artifact in phylogeny reconstruction. Systematic Biology 55:637–643.

Venditti, C., and M. Pagel. 2009. Speciation as an active force in promoting genetic evolution. Trends in Ecology and Evolution 25:14–20.

Warren, J. 2009. Extra petals in the buttercup (*Ranunculus repens*) provide a quick method to estimate the age of meadows. Annals of Botany 104:785–788.

Webster, A. J., R. J. H. Payne, and M. Pagel. 2003. Molecular phylogenies link rates of evolution and speciation. Science 301:478.

Welch, J. J. 2004. Accumulating Dobzhansky-Muller incompatibilities: reconciling theory and data. Evolution 58:1145–1156.

Welch, J. J., and D. Waxman. 2008. Calculating independent contrasts for the comparative study of substitution rates. Journal of Theoretical Biology 251:667–678.

Whitney, K. D., E. J. Baack, J. L. Hamrick, M. J. W. Godt, B. C. Barringer, M. D. Bennett, C. G. Eckert, et al. 2010. A role for nonadaptive processes in plant genome size evolution? Evolution 64: 2097–2109.

Whittle, C.-A., and M. O. Johnston. 2002. Male-driven evolution of mitochondrial and chloroplastidial DNA sequences in plants. Molecular Biology and Evolution 19:938–949.

Willis, K. J., K. D. Bennett, and H. J. B. Birks. 2009. Variability in thermal and UV-B energy fluxes through time and their influence on plant diversity and speciation. Journal of Biogeography 36: 1630–1644.

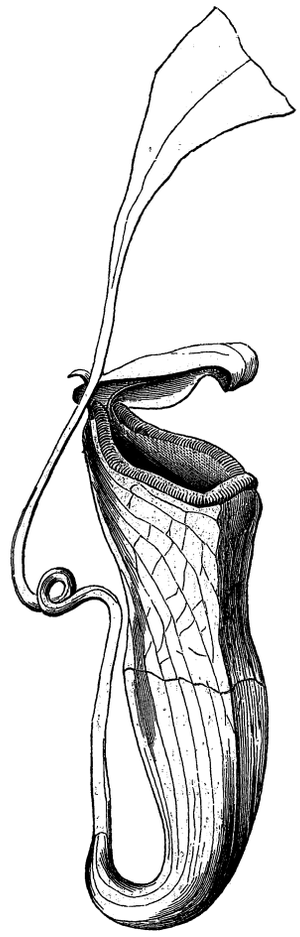
Woolfit, M. 2009. Effective population size and the rate and pattern of nucleotide substitutions. Biology Letters 5:417–420.

Woolfit, M., and L. Bromham. 2003. Increased rates of sequence evolution in endosymbiotic bacteria and fungi with small effective population sizes. Molecular Biology and Evolution 20:1545–1555.

———. 2005. Population size and molecular evolution on islands. Proceedings of the Royal Society B: Biological Sciences 272:2277–2282.

Wright, S. 1934. The method of path coefficients. Annals of Mathematical Statistics 5:161–215.
Wright, S. D., L. N. Gillman, H. A. Ross, and D. J. Keeling. 2010. Energy and the tempo of evolution in amphibians. Global Ecology and Biogeography 19:733–740.
Wright, S. D., J. Keeling, and L. Gillman. 2006. The road from Santa Rosalia: a faster tempo of evolution in tropical climates. Proceedings of the National Academy of Sciences of the USA 103: 7718–7722.
Wright, S. D., H. A. Ross, D. J. Keeling, P. McBride, and L. N. Gillman. 2011. Thermal energy and the rate of genetic evolution in marine fishes. Evolutionary Ecology 25:525–530.
Yang, Z. 2007. PAML: a program package for phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution 24:1586–1591. http://abacus.gene.ucl.ac.uk/software/paml.html.
Yoshimi, M., Y. Kitamura, S. Isshiki, T. Saito, K. Yasumoto, T. Terachi, and H. Yamagishi. 2013. Variations in the structure and transcription of the mitochondrial *atp* and *cox* genes in wild *Solanum* species that induce male sterility in eggplant (*S. melongena*). Theoretical and Applied Genetics 126:1851–1859.

Associate Editor: Charles F. Baer
Editor: Susan Kalisz



"The species whose minute anatomy we partially describe, is the *Nepenthes distillatoria*, found growing in China and at the Cape of Good Hope. [. . .] It bathes its roots in the hot swamps near the coast, but cannot lift its flowers very high in the sunshine, because its branching stem which bears many long and partly clasping leaves, and also its precious burthen of watercups, is too feeble to support the weight." From "The Structure of the Pitcher Plant" by J. G. Hunt (*The American Naturalist*, 1869, 3:13–17).