

ARTICLE

Received 20 Aug 2012 | Accepted 5 Apr 2013 | Published 21 May 2013

DOI: 10.1038/ncomms2836

# Taller plants have lower rates of molecular evolution

Robert Lanfear<sup>1,2</sup>, Simon Y. W. Ho<sup>3</sup>, T. Jonathan Davies<sup>4</sup>, Angela T. Moles<sup>5</sup>, Lonnie Aarssen<sup>6</sup>, Nathan G. Swenson<sup>7</sup>, Laura Warman<sup>5,8</sup>, Amy E. Zanne<sup>9,10</sup> & Andrew P. Allen<sup>11</sup>

Rates of molecular evolution have a central role in our understanding of many aspects of species' biology. However, the causes of variation in rates of molecular evolution remain poorly understood, particularly in plants. Here we show that height accounts for about one-fifth of the among-lineage rate variation in the chloroplast and nuclear genomes of plants. This relationship holds across 138 families of flowering plants, and when accounting for variation in species richness, temperature, ultraviolet radiation, latitude and growth form. Our observations can be explained by a link between height and rates of genome copying in plants, and we propose a mechanistic hypothesis to account for this—the 'rate of mitosis' hypothesis. This hypothesis has the potential to explain many disparate observations about rates of molecular evolution across the tree of life. Our results have implications for understanding the evolutionary history and future of plant lineages in a changing world.

<sup>1</sup>Ecology, Evolution, and Genetics, Research School of Biology, Building 116 Daley Road, Australian National University, Canberra, Australian Capital Territory 0200, Australia. <sup>2</sup>National Evolutionary Synthesis Center, Durham, North Carolina 27705-4667, USA. <sup>3</sup>School of Biological Sciences, Edgeworth David Building, University of Sydney, Sydney, New South Wales 2006, Australia. <sup>4</sup>Department of Biology, 1205 Avenue Docteur Penfield, McGill University, Montreal, Quebec, Canada H3A 1B1. <sup>5</sup>Evolution & Ecology Research Centre, School of Biological, Earth, and Environmental Sciences, The University of New South Wales, Sydney, New South Wales 2052, Australia. <sup>6</sup>Department of Biology, Queen's University, Kingston, Ontario, Canada K7L 3N6. <sup>7</sup>Department of Plant Biology, 612 Wilson Road, Michigan State University, East Lansing, Michigan 48824-1312, USA. <sup>8</sup>Institute of Pacific Islands Forestry, USDA Forest Service, 60 Nowelo Street, Hilo, Hawaii 96720, USA. <sup>9</sup>Department of Biological Sciences, 2023 G St. NW, George Washington University, Washington, District Of Columbia 20052, USA. <sup>10</sup>Center for Conservation and Sustainable Development, Missouri Botanical Garden, PO Box 299, St. Louis, Missouri 63110, USA. <sup>11</sup>Department of Biological Sciences, Macquarie University, Sydney, New South Wales 2109, Australia. Correspondence and requests for materials should be addressed to R.L. (email: rob.lanfear@anu.edu.au).

Many factors have been proposed to drive rates of molecular evolution in plants, including environmental energy, water availability, temperature, ultraviolet (UV) radiation, speciation rate, generation time and metabolic rate<sup>1–6</sup>, but disentangling these factors has been difficult. The results of many studies are equivocal or contradictory and there is confusion over the way in which some proposed mechanisms, especially the generation time effect, may operate<sup>1</sup>. Indeed, the vast majority of variation in rates of molecular evolution among plants remains both unexplored and unexplained.

Body size is fundamentally important to multiple aspects of the ecology, physiology and evolution of both animals and plants<sup>7–10</sup>. Because of this, analysing and understanding the relationship between body size and rates of molecular evolution has led to many important discoveries in animals<sup>10</sup>. However, the link between body size and rates of molecular evolution has not been examined in plants, despite its potential to elucidate many important but poorly understood aspects of plant molecular evolution.

Here, we use sister pairs and phylogenetically corrected linear regressions<sup>11</sup> to test for associations between rates of molecular evolution, height and other aspects of plant biology. Our results demonstrate that taller families of plants tend to have slower rates of molecular evolution in both the chloroplast and nuclear genomes. Furthermore, the association between height and rates of molecular evolution holds when accounting for variation in species richness, UV, temperature and latitude, and is also significant when considering exclusively herbaceous plant families, or those that contain only shrubs and trees.

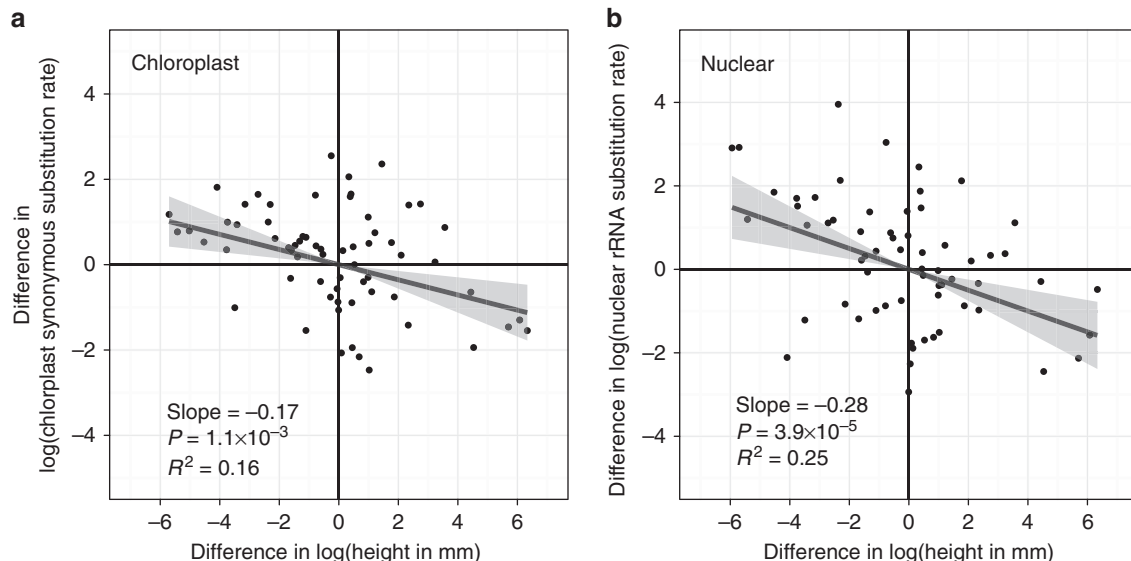
We discuss four possible mechanisms that might cause an association between rates of molecular evolution and height in plants: generation time, metabolic rate, population size and rates of mitotic cell division in the apical meristem. We conclude that metabolic rate and population size are unlikely to contribute to the association between rates and height, but that generation time and rates of mitosis may do so. This is because taller families of plants may comprise species with both longer generation times (and thus slower long-term rates of meiosis) and slower long-term rates of mitosis in their apical meristems.

## Results

**Data sets.** We assembled a data set of 138 families of flowering plants, grouped into 69 sister pairs. For each family, we estimated the amount of genetic change for the chloroplast and nuclear genomes from a large DNA sequence data set<sup>12</sup>. We estimated molecular branch lengths in substitutions per site, and absolute rates of molecular evolution in substitutions per site per million years. We did this using maximum-likelihood methods corrected to account for a known bias in branch length estimation (the node density effect; see Methods). We also estimated the average maximum height of the species in each family, using a database of maximum plant height records for over 20,000 species<sup>13</sup> collected to describe global patterns in plant height.

**Sister pairs analyses.** In our study, a sister pair comprises two families of plants that share a common ancestor to the exclusion of all other families in our data set. Both families are therefore the same age, and have had the same amount of time to accrue genetic differences. Consequently, any difference in the amount of genetic change that the two families have accumulated represents a difference in the underlying rate of molecular evolution, and differences in rates of molecular evolution can be calculated without knowing, or estimating, divergence dates of the sister pairs<sup>11</sup>.

Across all sister pairs, two distinct statistical tests indicate a significant association between the average maximum height a family and the rate of molecular evolution of that family. Two-tailed sign tests on sister pairs demonstrate that the taller plant family within each pair has the shorter branch length more often than expected by chance for both the chloroplast ( $P=0.024$ ) and nuclear genome ( $P=0.005$ ). Linear regressions of differences in height versus differences in branch lengths calculated from sister pairs confirm the significance of these associations and quantify their magnitude: height accounts for 16% of the variation in synonymous rates of molecular evolution in the chloroplast genome ( $B=-0.17$ ,  $P=0.001$ ; Fig. 1a) and 25% of the variation in rates of molecular evolution in nuclear rRNAs ( $B=-0.28$ ,  $P=4 \times 10^{-5}$ ; Fig. 1b).



**Figure 1 | The relationship between plant height and rates of molecular evolution in angiosperms.** Differences in the logarithms of plant height and rates of molecular evolution were measured for sister pairs of angiosperm families. Regressions through zero show a significant negative association between plant height and synonymous substitution rates measured from protein-coding genes in the chloroplast genome (a), and between plant height and substitution rates in rRNA genes in the nuclear genome (b). In each panel, the best-fit line (dark grey) was estimated using linear regression through the origin, with the 95% confidence intervals around this line shown in light grey (see main text).

**Table 1 | Pairwise regressions of chloroplast synonymous substitution rates.**

Predictor 1	Predictor 2	Height				Predictor 2			
		Slope	s.e.	<i>P</i> -value	Partial $r^2$	Slope	s.e.	<i>P</i> -value	Partial $r^2$
Height	Spp richness	-0.168	0.050	<b>0.001</b>	0.144	0.067	0.029	<b>0.023</b>	0.058
Height	Latitude	-0.183	0.053	<b>0.001</b>	0.186	-0.013	0.007	0.068	0.052
Height	Temperature	-0.180	0.054	<b>0.002</b>	0.183	6e-06	4e-06	0.157	0.041
Height	Ultraviolet	-0.178	0.054	<b>0.002</b>	0.171	2e-08	1e-08	0.158	0.027

Results of pairwise regressions through the origin of differences between chloroplast synonymous substitution rates and pairwise combinations of traits that include plant height. Significant *P*-values are highlighted in bold.

**Table 2 | Pairwise regressions of nuclear rRNA substitution rates.**

Predictor 1	Predictor 2	Height				Predictor 2			
		Slope	s.e.	<i>P</i> -value	Partial $r^2$	Slope	s.e.	<i>P</i> -value	Partial $r^2$
Height	Species richness	-0.271	0.062	<b>5e-05</b>	0.192	0.065	0.036	0.072	0.001
Height	Latitude	-0.279	0.064	<b>5e-05</b>	0.185	0.001	0.009	0.891	0.001
Height	Temperature	-0.278	0.065	<b>7e-05</b>	0.182	1e-06	5e-06	0.752	2e-04
Height	Ultraviolet	-0.278	0.064	<b>6e-05</b>	0.192	-4e-09	2e-08	0.774	9e-05

Results of pairwise regressions through the origin of differences between nuclear rRNA substitution rates and pairwise combinations of traits that include plant height. Significant *P*-values are highlighted in bold.

**Sister pairs analyses accounting for other variables.** To better understand the link between height and rates that we have observed, we assessed whether the relationship was affected by other known correlates of rates of molecular evolution in plants. In total, we were able to calculate family-level estimates of species richness, temperature, levels of UV radiation and latitude, each of which has been linked to rates of molecular evolution in previous studies<sup>2-4</sup>. We estimated differences in each of these variables for each of the sister pairs in our analysis and re-evaluated the relationship between height and molecular rates, including these additional variables as covariates (see Methods). We used the sister-pairs approach for these analyses because it makes the fewest assumptions of the available comparative methods about the ways in which rates and traits change over time<sup>11</sup>.

We estimated all pairwise models including height and each of the other variables as predictors of molecular rates in both the chloroplast and nuclear genomes (Tables 1 and 2, respectively). We also estimated the full model including all variables as predictors of molecular rates in the chloroplast and nuclear genomes (Fig. 2, Tables 3 and 4), and models including each of the additional variables as a predictor of height.

In all cases, height remained a significant predictor of molecular rates and consistently explained around one-fifth of the variation in rates of molecular evolution in both the chloroplast and nuclear genomes (Fig. 2, Tables 1-4). We also found a marginally significant ( $P=0.068$ ) negative association between height and latitude, and a significant ( $P=0.025$ ) positive association between height and temperature (see also ref. 13).

**Weighted sister-pairs analyses.** We performed weighted linear regressions to attempt to account for the differences in variance associated with our family-level height estimates (see Methods). The results of these analyses are qualitatively identical to the unweighted regressions. In linear models comparing height to rates of molecular evolution, weighted regressions show a significant negative association between height and rates of molecular evolution, and suggest that height accounts for 11-27% of the variation in rates of molecular evolution (chloroplast genome:

$B=-0.17$ ,  $P=0.005$ ,  $R^2=0.11$ ; nuclear genome:  $B=-0.29$ ,  $P=9 \times 10^{-6}$ ,  $R^2=0.27$ ). Similarly, weighted regressions of the full model including five predictors of rates of molecular evolution (height, species richness, latitude, temperature and UV) showed a significant negative correlation between height and the rate of molecular evolution in the chloroplast ( $B=-0.18$ ,  $P=0.004$ ) and nuclear genomes ( $B=-0.29$ ,  $P=3 \times 10^{-5}$ ). As with the unweighted regressions, all other predictors were non-significant except for species richness, which was significantly positively correlated to the rate of molecular evolution in the chloroplast genome ( $P=0.047$ ).

**Phylogenetic generalized least-squares analyses of height and rates of molecular evolution.** It is also possible to analyse the relationship between height and rates of molecular evolution by estimating the absolute rate of molecular evolution (in substitutions per site per million years) for each of the 138 families of plants in our data set (see Methods). This approach makes additional assumptions about the way that rates of molecular evolution change over time, and about the divergence dates of each of our sister pairs<sup>11</sup>. But it also provides more statistical power than a sister-pairs analysis, and allows us to straightforwardly account for uncertainty in the phylogenetic tree and molecular rate estimates using non-parametric bootstrapping (see Methods).

We used phylogenetic generalized least-squares (PGLS) regressions to compare log-transformed height to log-transformed absolute rates of molecular evolution, calculated from the combined nuclear and chloroplast data set. These analyses reveal highly significant negative associations between height and overall rates of molecular evolution ( $B=-0.05$ ,  $P=3 \times 10^{-7}$ ,  $R^2=0.41$ ). These relationships hold when accounting for uncertainty in the phylogeny and molecular rate estimates using 1,000 non-parametric bootstrap replicates (see Methods)—the 95% bootstrap confidence interval of the  $R^2$  value is 0.33-0.48, the 95% confidence interval of the slope is -0.04 to -0.06, and all 1,000 bootstrap replicates were statistically significant. The results are qualitatively identical when using weighted PGLS regressions.

**Table 3 | Multiple regression of chloroplast synonymous substitution rates.**

Predictor	Slope	s.e.	P-value	Partial r	Partial $r^2$
Species richness	7.890e-02	3.130e-02	<b>0.014</b>	0.248	0.062
Latitude	-2.921e-02	2.363e-02	0.221	-0.043	0.002
Temperature	1.082e-05	1.466e-05	0.463	0.116	0.013
Ultraviolet	-5.798e-08	4.769e-08	0.229	-0.086	0.007
Height	-1.823e-01	5.177e-02	<b>0.001</b>	-0.434	0.188

Results of multiple regression through the origin of differences between chloroplast synonymous substitution rates and five traits. Significant *P*-values are highlighted in bold. The *P*-value of the whole model is 0.002, and the  $R^2$  is 0.271.

**Table 4 | Multiple regression of nuclear rRNA substitution rates**

Predictor	Slope	s.e.	P-value	Partial r	Partial $r^2$
Species richness	7.585e-02	4.110e-02	0.0703	0.062	0.004
Latitude	-1.117e-03	2.732e-02	0.9675	0.091	0.008
Temperature	1.133e-05	1.842e-05	0.5410	0.016	2e-04
UV	-3.417e-08	5.566e-08	0.5418	0.088	0.008
Height	-2.772e-01	6.501e-02	<b>7.79e-05</b>	-0.417	0.174

Results of multiple regression through the origin of differences between nuclear rRNA substitution rates and five traits. Significant *P*-values are highlighted in bold. The *P*-value of the whole model is 0.001, and the  $R^2$  is 0.294.

**Analyses of herbaceous and woody plant families.** Previous studies have suggested that herbaceous and woody species show significantly different rates of molecular evolution<sup>6</sup>. Because of this, we conducted separate PGLS analyses for those families in our data set that are exclusively herbaceous ( $n=25$ ) and those that comprise only trees and/or shrubs ( $n=60$ ). In both cases, we observed significant negative associations between height and overall rates of molecular evolution calculated from the combined data set (herbs:  $B=-0.10$ ,  $P=0.003$ ,  $R^2=0.25$ ; trees/shrubs:  $B=-0.05$ ,  $P=0.019$ ,  $R^2=0.07$ ).

## Discussion

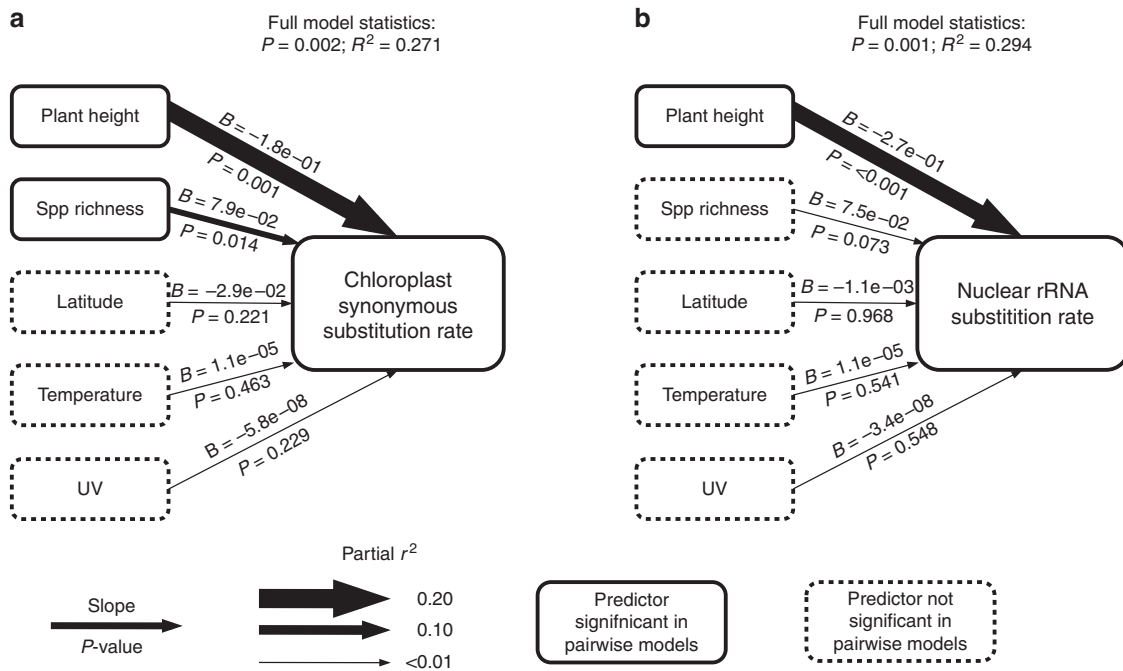
Our results demonstrate that taller families of plants evolve more slowly than shorter families, and that this association is independent of many other known correlates of rates of molecular evolution including temperature, latitude, UV radiation and species richness. The association between height and rates of molecular evolution holds for both the nuclear and chloroplast genomes, and across families that contain exclusively herbaceous species, as well as those that contain exclusively shrubs and trees. Our results are also robust to uncertainty in the underlying phylogenetic tree topology and molecular rate estimates.

The strength of the association between height and rates of molecular evolution is particularly striking because our family-level estimates of both height and rates of molecular evolution are necessarily crude due to incomplete sampling. Taking averages of traits for families of plants overlooks the fine-scale variation in traits that may occur between closely related species, and ignores the sometimes substantial variation among traits and rates within families. These factors are expected to introduce significant random error into our analysis, which suggests that the association between height and rates of molecular evolution may be even stronger than we have been able to detect in this study.

What might explain an association between height and rate of molecular evolution in plants? We consider four possible mechanisms below: generation time, metabolic rate, population size and rates of mitotic cell division in the apical meristem.

If taller families of plants tend to comprise species with longer generation times, our results could be explained by the effects of generation time on mutation or fixation rates in plant genomes. The generation time hypothesis states that species with shorter generation times copy their genomes more often, and consequently accrue more replication errors per unit time, resulting in higher mutation rates<sup>14</sup>. The generation time hypothesis assumes that generation time is correlated with the overall rate of genome replication. This is unlikely to be strictly true for plants because they grow from apical meristems, which undergo continual mitosis and from which reproductive tissues are derived late in development. Because of this, the number of mitotic cell divisions in plants can vary substantially between generations and among closely related species<sup>15</sup>. In contrast, many animals have deterministic development in which there is a fixed number of mitotic cell divisions in each generation. As a result, generation time can provide a useful proxy of the overall rate of genome replication in animals, but is unlikely to do so in plants. This may explain why the evidence for the generation time hypothesis is very strong for animals<sup>16–18</sup>, but mixed for plants<sup>1,19–21</sup>. Despite this, generation time will remain tightly associated with long-term rates of meiosis in plants, because an extant plant genome would have experienced one meiosis in each generation through which it has passed. Therefore, if a significant proportion of heritable mutations are associated with meiosis in plants, and if taller families of plants tend to have longer generation times, then a generation time effect on mutation rates might explain our results. A generation time effect on substitution rates could also explain our results, because all else being equal, generation times determines the absolute timescale of genetic drift.

The metabolic rate hypothesis proposes that organisms with higher mass-specific metabolic rates produce a higher concentration of damaging metabolic by-products (oxygen radicals), and thus accumulate more DNA damage and more mutations per unit time<sup>22</sup>. The metabolic rate hypothesis could explain our results if taller plants have lower mass-specific metabolic rates in their apical meristems. Although this is possible<sup>9</sup>, allometry theory predicts, and empirical data show, that cellular metabolic rates of certain tissues are essentially independent of plant height<sup>23</sup>, undermining a potential role for metabolic rate in driving rates of molecular



**Figure 2 | Biological predictors of rates of molecular evolution in plants.** Five potential predictors of substitution rates were measured for each plant family, log-transformed and included in two different linear models: **(a)** synonymous substitution rates in chloroplast protein-coding genes; and **(b)** substitution rates in nuclear rRNA genes. The  $P$ -value and  $R^2$  of the full model are shown above each diagram. The width of each arrow represents the partial  $r^2$  of each predictor, which estimates the effect of that predictor after controlling for all other predictors. The slope and significance of each predictor are shown above and below each arrow, respectively. Predictors that were significant in pairwise linear models with plant height are shown in solid boxes, and those that were not significant are shown in dashed boxes. Plant height was significant in all pairwise linear models (Tables 1 and 2).

evolution in plants. Furthermore, although correlative evidence for the metabolic rate hypothesis in animals is mixed<sup>24–26</sup>, experiments in animals have shown that oxygen radicals generated in organelles do not damage the nuclear DNA of the same cell<sup>27</sup>, and that germ-line mutations do not accumulate under increased oxidative stress<sup>28</sup>. If the same is true for plants, the metabolic rate hypothesis cannot explain the link we have observed between height and rates of molecular evolution in the nuclear genome.

If taller families of plants tend to comprise species with larger effective population sizes ( $N_e$ ), and if a significant proportion of the genomic changes we have measured are deleterious, then a population size effect could explain our results. In this case, selection would be more effective in taller plants, and so slightly deleterious mutations would be fixed less frequently in these lineages<sup>29</sup>. If taller plants have larger  $N_e$ , we would expect a negative correlation between height and the ratio of non-synonymous to synonymous substitutions ( $dN/dS$ <sup>29</sup>). To test this prediction, we compared differences in  $dN/dS$  to differences in height across all of our sister pairs. We found that height was significantly positively correlated to  $dN/dS$  ( $B = 0.16 \pm 0.07$ ,  $P = 0.02$ ,  $R^2 = 0.08$ ). This suggests that taller plants have smaller  $N_e$  and does not support a population size effect as an explanation for our results.

To account for the possibility that rates of mitosis could drive rates of molecular evolution in plants, we propose a novel explanation for our results here—the ‘rate of mitosis’ (ROM) hypothesis. The ROM hypothesis states that a substantial fraction of mutations that accumulate in the germ-line DNA of plants occur during mitotic cell divisions in the apical meristem. Such mutations are potentially heritable because plants do not sequester their germ lines until very late in development<sup>30,31</sup>. Thus, species with higher rates of mitotic cell division in the apical meristem will copy their DNA more

frequently, accrue more DNA replication errors per unit time, and have higher rates of mutation and substitution. Crucially, the long-term rate of mitosis in the apical meristem is likely to be lower in taller plants, because growth slows as plants increase in size and because there are physical limits to the delivery of water and nutrients to apical meristems as they increase in distance from the root system<sup>1,9,20,32–34</sup>. Thus, the ROM hypothesis predicts that taller plants will have lower mutation rates, potentially explaining the observations we have made in this study.

Interestingly, the ROM hypothesis could help to explain many other observations about rates of molecular evolution in plants and animals. For example, it can explain why perennials evolve more slowly than annuals, why woody plants evolve more slowly than herbs and why tree ferns evolve more slowly than other ferns<sup>1,6,35</sup>. In all three of these cases, the faster-evolving plants (annuals, herbs and non-arborescent ferns) are likely to have higher rates of mitosis than their more slowly evolving relatives (perennials, woody species and tree ferns), although we also note that the faster-evolving plants are also likely to have shorter generation times in all of these cases. The ROM hypothesis might also explain why generation times correlate with rates of molecular evolution in animals, but not in plants<sup>20,21,36</sup>, because generation times are a reliable predictor of rates of mitosis in animals, but not in plants. Finally, the ROM hypothesis might explain previously observed correlations between rates of molecular evolution and environmental energy<sup>3</sup>, latitude<sup>2</sup> and water availability<sup>5</sup> in plants. In all three of these cases, the faster-evolving taxa are associated with conditions more favourable for rapid plant growth (higher environmental energy, latitudes close to the tropics and higher water availability). Higher growth rates are likely to be associated with higher rates of mitosis in the apical meristem, and therefore with higher rates of mutation and substitution.



It is not possible from our data to distinguish the relative contributions of meiosis and mitosis to rates of molecular evolution in plants. However, this may be possible in systems where long-term rates of meiosis are constant across species, such as between closely related annual plants. In this case, the ROM hypothesis predicts that taller annual plants have higher rates of molecular evolution (the opposite to the pattern shown here) because their apical meristems will undergo more mitotic cell divisions per year. Thus, comparing the rates of molecular evolution to height in closely related annual plants may provide a useful test of the ROM hypothesis, and may allow future studies to more directly assess the relative contributions of meiosis and mitosis to rates of molecular evolution in plants.

Our results further underline the central role of height in the ecology and evolution of plants. We have shown that taller plants have lower rates of molecular evolution than shorter plants, and that this is likely to be driven by lower underlying mutation rates caused by lower rates of genome copying in taller plants. This has implications for reconstructing the evolutionary history of plants using techniques such as molecular dating, because it informs our expectations of rate variation among plant lineages. It also has implications for predicting the evolutionary future of plants, because the ability of species to adapt to a changing world depends critically on their underlying mutation rates<sup>37</sup>.

## Methods

**Phylogeny estimation and selection of sister pairs.** We estimated a phylogenetic tree of flowering plants using the DNA data set of Burleigh *et al.*<sup>12</sup>, comprising 567 species and 10,552 base pairs. We refined their alignments<sup>12</sup>, then partitioned the data set into MatK (which is often a pseudogene), 1st + 2nd codon positions, 3rd codon positions, RNA stems and RNA loops. We estimated a maximum-likelihood topology using RaxML<sup>38</sup>, applying an independent GTR + I + G model to each data partition. We then assigned each species in the tree to a family of plants<sup>29,39</sup>, and retained only those families that were recovered as monophyletic in our tree. Finally, we chose all sister pairs of monophyletic families in our tree for our final analysis, resulting in a data set of 138 monophyletic angiosperm families grouped into 69 sister pairs (Fig. 1). Input and output files for RAXML are available from DataDryad.

**Branch length estimation.** The sister-pairs method requires that branch lengths be estimated for each family of each sister pair. The more species from a family that can be included in each branch length estimate, the more representative that branch length estimate will be as an estimate of the branch length of the family in question. However, branch length estimation suffers from a known bias, whereby longer branch lengths tend to be inferred in clades with higher node density<sup>40</sup>. In our case, this could create an artefactual association between species richness and molecular branch lengths. To avoid this, we first equalized the number of representative species of the two members of each sister pair, by randomly deleting species from the family with more representative species. This ensures that the branch lengths for each sister pair are calculated from families with equal node density, thus removing the node density artefact from our data<sup>11</sup>. The removal of species in this way increases the variance of our branch lengths as estimators of the branch length of a given family of plants, but serves to remove any biases in these estimators. Thus, this method reduces the power of our approach to detect associations between rates of molecular evolution and traits such as height. This resulted in a tree of 196 taxa, with the majority of sister pairs comprising families represented by a single species in the tree (49/69 sister pairs), and the remaining sister pairs comprising families represented by two to four species per family.

Using this tree of 196 taxa, we estimated branch lengths using maximum-likelihood in HyPhy v2.0 (ref. 41). For families with more than one representative in our data set, we calculated the phylogenetic average of the family branch lengths by successively averaging branch lengths down the phylogeny for each family<sup>4</sup>. The result of this was an estimate of the nuclear ribosomal RNA and chloroplast synonymous branch lengths for each of the 138 families in our data set. These data are available from DataDryad.

**Estimation of absolute rates of molecular evolution.** We estimated absolute rates of molecular evolution in substitutions per site per million years across the entire 10,552-bp alignment using R8s<sup>42</sup>. To do this, we first selected at random one representative species from each of the 140 families in our data set for which we had height estimates, in order to avoid the node density effect in our rate estimations (138 families from sister pairs, plus two additional families which we excluded from the sister-pair analyses because they did not fall into sister pairs). We then used RAXML<sup>38</sup> to calculate the ML tree with branch lengths from the

10,552-bp, 140 taxon data set, as well as 1,000 bootstrap trees with branch lengths using non-parametric bootstrapping. We then used a Python script to calculate absolute rates of molecular evolution on each of the trees using a diverse set of 30 fossil calibrations<sup>43</sup>. This script and the input and output from R8s are available from DataDryad.

**Estimation of average maximum height for each family.** To calculate the average maximum height for each of the 138 families in our data set, we used a published database of over 32,737 maximum plant height records from 20,679 species<sup>13</sup>. This data set has been rigorously curated to ensure that taxonomic names are accurate, that synonyms are accounted for, that extreme values in the data set are legitimate and that all measurements represent maximum recorded height from a published source<sup>13</sup>. To calculate the average maximum height for each family, we first calculated the arithmetic mean of the log-transformed maximum heights of each genus in the family. We then calculated the arithmetic mean of the genus-level averages to provide an estimate of the average maximum height of each family. We are unable to release the raw height data set from Moles *et al.*<sup>13</sup>, because this would violate several data-sharing agreements, however, the family average height data on which the analyses in this study are based are available on DataDryad.

Using phylogenetic averaging in this way ensures that our estimates of plant height are, to the best of our ability, comparable to our estimates of molecular branch lengths. This is because branch length estimates represent the integral of the rate of molecular evolution over the entire history of a lineage, rather than an estimate of the rate of molecular evolution in the present day. On the other hand, traits such as plant height are measured exclusively from extant taxa. A phylogenetic average of a trait represents a measure of central tendency that is more comparable to the estimates of molecular rates that we use, and this approach has been shown to improve the power of comparative studies of molecular rate variation<sup>44</sup>.

Our estimates of plant height will in many cases be associated with high variance, particularly where our sampling of the extant taxa in a family is sparse, and/or the variance in height within a given family is large. This will introduce noise, but not bias, into our analysis, because the sampling of plant heights in the data set is not biased with respect to the rates of molecular evolution of the plants in question. To attempt to account for this sampling variance, we performed weighted regressions in which we used as weights the proportion of extant genera in each sister pair (for sister pairs analyses) or family (for PGLS analyses) for which we had at least one height record. Thus, the more of the extant genera we have sampled for a given sister pair or family, the higher the weight of that data point in the regression. We use the proportion of genera because genus-level sampling will have the largest effect on our phylogenetically averaged height estimates. Using these weights may help to account for the different variances associated with our height estimates, although we acknowledge that the approach is far from perfect.

**Estimation of other traits for each family.** Species numbers in each family were estimated from the Families of Flowering Plants database (<http://delta-intkey.com>) following Barraclough *et al.*<sup>4</sup>, as were the existence of species in each family that were herbs, shrubs or trees. All other variables (latitude, UV irradiation and temperature) were estimated by calculating the mean value of each trait per unit of area using digital distribution maps for each family of plants<sup>3</sup>. Trait data is available from DataDryad.

**Diagnostic tests on the data.** Parametric analyses (such as linear models) of sister pairs make certain assumptions about the data. To ensure that these assumptions were met, we performed three diagnostic tests on the differences in rates and traits that we calculated, following the advice given in Lanfear *et al.*<sup>11</sup>. These comprise: (i) a test to check that our log transformation of all rates and traits removed any association between the variance of data points and their absolute values<sup>45</sup>; (ii) a test to ensure that variance in rate and trait differences increases linearly with evolutionary time<sup>46</sup>; and (iii) a test to ensure that the variance in rate differences was not affected by low numbers of substitutions<sup>44</sup>. The first test indicated that log transformations of rate and trait data were appropriate for most traits and rates, except temperature and UV which were squared, and latitude which was left untransformed. The second test indicated that standardizing all differences by the square root of the sum of the branch lengths (as suggested in ref. 46) in each sister pair was appropriate. The third test indicated that four of the sister-pair differences in chloroplast synonymous branch lengths were unreliable (leaving a total of 65 sister pairs with reliable data), and that six of the sister-pair differences in nuclear rRNA branch lengths were unreliable (leaving 63 sister pairs with reliable data). Accordingly, we performed all statistical tests with and without these unreliable sister pairs, and the results are qualitatively identical in all cases. The results reported here are those with the unreliable data points removed, which is the approach suggested in previous studies<sup>11,44</sup>.

**Statistical analysis.** All statistical analyses were performed in R<sup>47</sup>. For the sister-pairs data, we performed linear regressions through zero<sup>46</sup> of the transformed trait data, after excluding sister pairs for which rate estimates were unreliable. The slope, s.e. of the slope and *P*-values for each variable in linear regressions were calculated using the *lm()* function in R. Partial correlation coefficients (partial *r*) and partial

coefficients of determination (partial  $r^2$ ) were calculated in R using matrix inversion of Pearson correlations between pairs of variables. For the absolute rate analyses, we performed PGLS regression using the CAPER and nlme packages. Analyses on the ML tree using CAPER indicated that the best fit of the model to the data was obtained when optimizing the 'delta' parameter during the PGLS. Therefore, we performed all PGLS analyses by optimizing 'delta'. R scripts are available from DataDryad.

**Data availability.** All data, including data on trees, traits, rates, fossil calibrations, R code and Python scripts used for this publication are available from DataDryad at <http://dx.doi.org/dryad.43mg3>.

## References

- Gaut, B., Yang, L., Takuno, S. & Eguarte, L. E. The patterns and causes of variation in plant nucleotide substitution rates. *Annu. Rev. Ecol. Evol. S* **42**, 245–266 (2011).
- Wright, S., Keeling, J. & Gillman, L. The road from Santa Rosalia: a faster tempo of evolution in tropical climates. *Pro. Natl Acad. Sci. USA* **103**, 7718–7722 (2006).
- Davies, T. J., Savolainen, V., Chase, M. W., Moat, J. & Barraclough, T. G. Environmental energy and evolutionary rates in flowering plants. *Proc. Biol. Sci.* **271**, 2195–2200 (2004).
- Barraclough, T. G. & Savolainen, V. Evolutionary rates and species diversity in flowering plants. *Evolution* **55**, 677–683 (2001).
- Goldie, X., Gillman, L., Crisp, M. & Wright, S. Evolutionary speed limited by water in arid Australia. *Pro. R. Soc. B* **277**, 2645–2653 (2010).
- Smith, S. A. & Donoghue, M. J. Rates of molecular evolution are linked to life history in flowering plants. *Science* **322**, 86–89 (2008).
- Reich, P. B., Tjoelker, M. G., Machado, J.-L. & Oleksyn, J. Universal scaling of respiratory metabolism, size and nitrogen in plants. *Nature* **439**, 457–461 (2006).
- Westoby, M. A leaf-height-seed (LHS) plant ecology strategy scheme. *Plant Soil* **199**, 213–227 (1998).
- Enquist, B., West, G., Charnov, E. & Brown, J. Allometric scaling of production and life-history variation in vascular plants. *Nature* **408**, 907–911 (1999).
- Bromham, L. The genome as a life-history character: why rate of molecular evolution varies between mammal species. *Philos. Trans. R. Soc. B* **366**, 2503–2513 (2011).
- Lanfear, R., Welch, J. J. & Bromham, L. Watching the clock: studying variation in rates of molecular evolution between species. *Trends Ecol. Evol.* **25**, 495–503 (2010).
- Burleigh, J. G., Hilu, K. W. & Soltis, D. E. Inferring phylogenies with incomplete data sets: a 5-gene, 567-taxon analysis of angiosperms. *BMC Evol. Biol.* **9**, 61 (2009).
- Moles, A. T. *et al.* Global patterns in plant height. *J. Ecol.* **97**, 923–932 (2009).
- Li, W. H., Ellsworth, D. L., Krushkal, J., Chang, B. H. & Hewett-Emmett, D. Rates of nucleotide substitution in primates and rodents and the generation-time effect hypothesis. *Mol. Phylogenet. Evol.* **5**, 182–187 (1996).
- Petit, R. J. & Hampe, A. Some evolutionary consequences of being a tree. *Annu. Rev. Ecol. Evol. S* **37**, 187–214 (2006).
- Thomas, J. A., Welch, J. J., Lanfear, R. & Bromham, L. A generation time effect on the rate of molecular evolution in invertebrates. *Mol. Biol. Evol.* **27**, 1173–1180 (2010).
- Nabholz, B., Glémin, S. & Galtier, N. Strong variations of mitochondrial mutation rate across mammals—the longevity hypothesis. *Mol. Biol. Evol.* **25**, 120–130 (2008).
- Welch, J. J., Bininda-Emonds, O. R. P. & Bromham, L. Correlates of substitution rate variation in mammalian protein-coding sequences. *BMC Evol. Biol.* **8**, 53 (2008).
- Gaut, B. S., Clark, L. G., Wendel, J. F. & Muse, S. V. Comparisons of the molecular evolutionary process at *rbcl* and *ndhF* in the grass family (Poaceae). *Mol. Biol. Evol.* **14**, 769–777 (1997).
- Whittle, C. -A. & Johnston, M. O. Broad-scale analysis contradicts the theory that generation time affects molecular evolutionary rates in plants. *J. Mol. Evol.* **56**, 223–233 (2003).
- Verdu, M., Pausas, J. G., Segarra-Moragues, J. G. & Ojeda, F. Burning phylogenies: fire, molecular evolutionary rates, and diversification. *Evolution* **61**, 2195–2204 (2007).
- Martin, A. P. & Palumbi, S. R. Body size, metabolic rate, generation time, and the molecular clock. *Proc. Natl Acad. Sci. USA* **90**, 4087–4091 (1993).
- Niklas, K. J. & Enquist, B. J. Invariant scaling relationships for interspecific plant biomass production rates and body size. *Proc. Natl Acad. Sci. USA* **98**, 2922–2927 (2001).
- Lanfear, R., Thomas, J. A., Welch, J. J., Brey, T. & Bromham, L. Metabolic rate does not calibrate the molecular clock. *Proc. Natl Acad. Sci. USA* **104**, 15388–15393 (2007).
- Santos, J. C. Fast molecular evolution associated with high active metabolic rates in poison frogs. *Mol. Biol. Evol.* **29**, 2001–2018 (2012).
- Lourenço, J. M., Glémin, S., Chiari, Y. & Galtier, N. The determinants of the molecular substitution process in turtles. *J. Evol. Biol.* **26**, 38–50 (2013).
- Hoffmann, S., Spitzkovsky, D., Radicella, J., Epe, B. & Wiesner, R. Reactive oxygen species derived from the mitochondrial respiratory chain are not responsible for the basal levels of oxidative base modifications observed in nuclear DNA of mammalian cells. *Free Radical Bio. Med.* **36**, 765–773 (2004).
- Joyner-Matos, J., Bean, L. C., Richardson, H. L., Sammel, T. & Baer, C. F. No evidence of elevated germline mutation accumulation under oxidative stress in *Caenorhabditis elegans*. *Genetics* **189**, 1439–1447 (2011).
- Woolfit, M. Effective population size and the rate and pattern of nucleotide substitutions. *Biol. Lett.* **5**, 417–420 (2009).
- Klekowski, E. J. Genetic load and its causes in long-lived plants. *Trees-Struct. Funct.* **2**, 195–203 (1988).
- Klekowski, E. J. & Godfrey, P. J. Ageing and mutation in plants. *Nature* **3**, 389–391 (1989).
- Kimura, M. *The Neutral Theory of Molecular Evolution* (Cambridge University Press, 1985).
- Gleeson, S. & Tilman, D. Plant allocation, growth-rate and successional status. *Funct. Ecol.* **8**, 543–550 (1994).
- Grime, J. & Hunt, R. Relative growth-rate: its range and adaptive significance in a local flora. *J. Ecol.* **63**, 393–422 (1975).
- Korall, P., Schuettelpelz, E. & Pryer, K. M. Abrupt deceleration of molecular evolution linked to the origin of arborescence in ferns. *Evolution* **64**, 2786–2792 (2010).
- Bromham, L. Why do species vary in their rate of molecular evolution? *Biol. Lett.* **5**, 401–404 (2009).
- Willi, Y. & Hoffmann, A. A. Demographic factors and genetic variation influence population persistence under environmental change. *J. Evol. Biol.* **22**, 124–133 (2009).
- Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690 (2006).
- AGP. II An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG II. *Bot. J. Linn. Soc.* **141**, 399–436 (2003).
- Hugall, A. F. & Lee, M. S. Y. The likelihood node density effect and consequences for evolutionary studies of molecular rates. *Evolution* **61**, 2293–2307 (2007).
- Pond, S. L., Frost, S. D. & Muse, S. V. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* **21**, 676–679 (2005).
- Sanderson, M. J. r8s: Inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* **19**, 301–302 (2003).
- Magallon, S. & Castillo, A. Angiosperm diversification through time. *Am. J. Bot.* **96**, 349–365 (2009).
- Welch, J. J. & Waxman, D. Calculating independent contrasts for the comparative study of substitution rates. *J. Theor. Biol.* **251**, 667–678 (2008).
- Freckleton, R., Harvey, P. & Pagel, M. Phylogenetic analysis and comparative data: a test and review of evidence. *Am. Nat.* **160**, 712–726 (2002).
- Garland, T., Harvey, P. & Ives, A. Procedures for the analysis of comparative data using phylogenetically independent contrasts. *Syst. Biol.* **41**, 18–32 (1992).
- R Development Core Team R. *A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2012).

## Acknowledgements

We thank the Australian Research Council-New Zealand Research Network for Vegetation Function for facilitating this collaboration. We also thank Michelle Leishman for assistance with data collection and Ainsley Seago for help with figures.

## Author contributions

R.L., S.Y.W.H., T.J.D., A.T.M. and A.A. conceived and designed the study. R.L., S.Y.W.H., T.J.D., A.M., L.A., N.G., L.W., A.Z. and A.A. collected/collated the data. R.L., S.Y.W.H., T.J.D. and A.M. analysed the data. R.L. drafted the manuscript, and R.L., S.Y.W.H., T.J.D., A.M., L.A., N.G., L.W., A.Z. and A.A. finalized the manuscript.

## Additional information

**Competing financial interests:** The authors declare no competing financial interests.

**Reprints and permission information** is available online at <http://npg.nature.com/reprintsandpermissions/>

**How to cite this article:** Lanfear, R. *et al.* Taller plants have lower rates of molecular evolution. *Nat Commun.* 4:1879 doi: 10.1038/ncomms2836 (2013).