Polyploid Formation Shapes Flowering Plant Diversity

Samuel V. Scarpino,[1,2,*] Donald A. Levin,[1] and Lauren Ancel Meyers[1,2]

1. Department of Integrative Biology, University of Texas, Austin, Texas 78712; 2. Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501

Submitted November 9, 2013; Accepted May 8, 2014; Electronically published August 18, 2014

Online enhancement: appendix. *Dryad data:* http://dx.doi.org/10.5061/dryad.35r23.

abstract: Polyploidy, or whole genome duplication, has been an important feature of eukaryotic evolution. This is especially true in flowering plants, where all extant angiosperms have descended from polyploid species. Here we present a broad comparative analysis of the effect of polyploidy on flowering plant diversity. We examine the widely held hypothesis that polyploid flowering plants generate more diversity than their diploid counterparts, by fitting stochastic birth/death models to observed ploidal frequency data from 60 extant angiosperm genera. Our results suggest the opposite, that diploids speciate at higher rates than polyploids, through a combination of simple diploid speciation and tetraploidy. Importantly, the estimated diploid advantage stemmed primarily from a higher rate of polyploidization in diploids than polyploids. Our model is also able to account for the empirically observed correlation between polyploidy and species richness without assuming that polyploids have a speciation advantage over diploids.

Keywords: polyploidy, angiosperms, ratchet model, speciation rates, approximate Bayesian computation.

Introduction

Recent genome level data has confirmed the ubiquity of polyploidy in plants (Soltis et al. 2009). Current estimates suggest all extant angiosperms have a polyploid ancestor, sharing at least two whole genome duplication events, and that 20%–50% are recently formed polyploids (Levin 2002; Soltis et al. 2009; *Amborella* Genome Project 2013). Although little doubt remains about the pervasiveness of polyploidy in flowering plants, there is considerable debate over whether diploids and polyploids differ in speciation and diversification rates.

Otto and Whitten (2000) reported a positive correlation between polyploidy and species richness. This, combined with evidence for polyploid-biased lineage survivorship through the Cretaceous-Tertiary boundary (Fawcett et al. 2009; Soltis and Burleigh 2009) and anecdotal support for polyploid superiority, has stimulated an ongoing discussion focused on possible biological and genetic (intrinsic) advantages of polyploids relative to diploids, as reviewed by Soltis et al. (2003). One hypothesis is that chromosome doubling itself may alter the ecological tolerances of populations, thereby allowing them to invade new habitats and rapidly establish (Stebbins 1980, 1985; Levin 1983, 2002). Another hypothesis is that phenotypic variation arising during polyploidy events enables the invasion of new habitats, mediated by a range of possible genetic and epigenetic mechanisms, as reviewed by te Beest et al. (2012).

In contrast, several recent studies estimating speciation and extinction rates of polyploids and diploids have failed to find evidence for a polyploid speciation advantage: Wood et al. (2009) found no evidence for higher polyploid diversification rates in twelve angiosperm genera, and Mayrose et al. (2011) found that net speciation rates in recently formed polyploids were lower than congeneric diploids. These studies have led some to conclude that polyploids lead to evolutionary dead ends (Arrigo and Barker 2012).

In response to the various hypotheses about polyploid advantages, Meyers and Levin (2006) proposed that the high frequency of polyploids in flowering plants might simply be an inevitable consequence of the directionality of polyploidy. Ploidal increases are largely irreversible over short evolutionary timescales, and therefore the abundance of polyploids should increase over time in a ratchet-like manner (Stebbins 1971). In their analysis, Meyers and Levin (2006) fit a simple, deterministic evolutionary model of polyploid evolution to data from 10 angiosperm genera. Although this model assumed that speciation in congeneric diploids and polyploids occurred at the same rate, it was still able to produce distributions of ploidal levels that were statistically similar to those observed in 9 of the 10 focal genera. This suggests that the irreversibility of polyploidy itself may explain the ubiquity of polyploids. These results also imply that this ratchet model should serve as a parsimonious baseline (null model) when considering other possible explanations for polyploid abundance.

* Corresponding author; e-mail: scarpino@utexas.edu.

Am. Nat. 2014. Vol. 184, pp. 456–465.
0003-0147/2014/18404-55085$15.00.
DOI: 10.1086/677752

Here, we address the three conflicting hypotheses about the evolutionary potential of polyploids—that they are drivers of diversification, evolutionary dead-ends, or neither—by examining the relationship between polyploidy and flowering plant diversity in a broad comparative context. Specifically, we extend the model introduced in Meyers and Levin (2006) and apply it to data from 60 angiosperm genera to assess whether (1) the simple ratchet model can explain ploidal level distributions across this phylogenetically broader set of taxa; (2) there is statistical evidence for differences in the net speciation rates of polyploids and diploids, and, if so, in which direction; and (3) allopolyploidy has contributed more to the formation of new polyploid lineages than autopolyploidy.

Models, Methods, and Data

Data

Our analysis was based on the observed ploidal level distributions of 60 flowering plant genera chosen from several issues of the Missouri Botanical Garden Index to Plant Chromosome Numbers, spanning the years 1967 to 2000; data are deposited in Dryad Digital Repository, http://dx.doi.org/10.5061/dryad.35r23 (Moore 1973; Goldblatt 1981; Missouri Botanical Garden 2005; Scarpino et al. 2014). We sought genera for which chromosome counts were available for large numbers of species and included at least three ploidal levels. In total, we identified the ploidal level of 7,271 species, with an average of 121 species per genus, ranging from 35 identified species in *Rosa* to 336 identified species in *Senecio*. All 10 genera used in Meyers and Levin (2006) were included in our data set; however, the species counts and ages were updated to reflect the most current data.

To determine the ploidal level of a species, we followed the rationale of Grant (1981). For example, if a genus contains species with 20, 40, or 60 chromosomes species, then 2n = 20 would be counted as diploid, 4n = 40 tetraploid, and 6n = 60 hexaploid. Where minor intraspecific variation was present and one number prevailed, that number was used. We were unable to unambiguously assign eight species to a ploidal level and they were excluded, these include two species from *Festuca*, four species from *Poa*, and one species from *Aster*. Importantly, these excluded species represent <0.1% of the 7,271 species in our data set.

The Polyploid Ratchet Model

The model assumes that a genus is founded by a single diploid species and then tracks the evolutionary dynamics of the genus in terms of changing numbers of species at each ploidal level. The ploidal level of each species in the genus is given with respect to the original founder. We let x_k denote the number of species of ploidal level k, and consider only even values of k. For a species with ploidal level k, r_k denotes the within-ploidal level diversification rate, which is the rate species give rise to daughter lineages with the same ploidal level, and h_k denotes the rate at which the species gives rise to new species through successful polyploidy events. The resulting polyploids are autopolyploids with probability a and allopolyploids with probability $(1 - a)$. Extinction happens at a background rate μ. Using these parameters, the net speciation rate for ploidal level k is $\lambda_k = r_k + h_k - \mu$. The model assumes that polyploidy is irreversible over short evolutionary timescales, specifically, the estimated or assumed age of each genus.

In this study, we consider simple and complex versions of this model. "Simple ratchet" (two parameters) is the original Meyers and Levin (2006) model where all species in a genus share a single diversification rate r and a single polyploidization rate h, all polyploids are allopolyploids ($a = 0$), and there is no extinction ($\mu = 0$). In the "complex ratchet" (five parameters), polyploid species share a single polyploidization rate ($h_p = h_4 = h_6 = h_8 = ...$) and diversification ($r_p = r_4 = r_6 = r_8 = ...$) rate that can be different from the diploid polyploidization ($h_d = h_2$) and diversification rates ($r_d = r_2$). The fraction of autopolyploids (a) can take on nonzero values. Extinction is set at a fixed rate for all species in a genus ($\mu = \mu_2 = \mu_4 = \mu_6 = ...$).

Simulation of the Model

Here, we outline a numerical algorithm for simulating the complex ratchet model, henceforth called the "simulation." This continuous-time stochastic algorithm has evolutionary parameters specific to each genus: the diversification rates (r), polyploidization rates (h), the probability that a polyploidy event yields an autopolyploid species (a), the final size of the genus (number of species = N_g), the estimated age of the genus (T_g million years), and extinction rate (μ). Each genus is simulated separately, beginning with a single diploid species at time zero and tracking the number of species in each ploidal class as the genus diversifies through speciation and polyploidization.

In the continuous-time version of the model, the times between both speciation and polyploidization events for a single lineage are distributed exponentially. The simulation has a constantly updating queue of speciation and polyploidization events, and iteratively performs the first event in the queue until the genus reaches its final size. Each event has three pieces of information: type (speciation or polyploidization), time (t_e, million years), and spe-

cies (A_e). Speciation events occur as follows. S1: the simulation clock is updated to the time of the event (t_e). S2: the parent species A persists and a new offspring species B is created. S3: species A is assigned a new time until speciation (σ_a). This value is a random deviate from an exponential distribution with rate r_d or r_p if A is a diploid or polyploid, respectively. S4: species B is assigned a time until speciation (σ_b) and a time until polyploidization (γ_b). These values are random deviates from an exponential distribution with rates r_d or r_p and h_d or h_p, depending on whether the species is a diploid or higher ploid. S5: event times σ_a, σ_b, and γ_b are inserted into the queue and the queue is sorted in ascending order, based on the timing of the event.

For polyploidization events, a random deviate from a Bernoulli distribution with probability of success (a) is drawn, and the event will be an autopolyploid event if a one is selected and an allopolyploid event otherwise. If it is an allopolyploid event, a second parent species is required. The simulation checks whether another species is waiting to form an allopolyploid. If so, polyploidization occurs, and if not, the parent species is wait-listed for polyploidization (at any time, there is at most a single species waiting). During an allopolyploidization event the following operations occur. P1: the simulation clock is updated to the time of the event (t_e; the latter of the two parental polyploidization times). P2: the parent species A_1 and A_2 with ploidal levels k_1 and k_2 persist, and new offspring B is created with ploidal level $k_1 + k_2$. P3: species A_1 and A_2 are assigned new times until polyploidization (γ_{a1} and γ_{a2}). These times are random deviates from an exponential distribution with rate h_k. P4: The offspring species (B) is assigned a time until speciation (σ_b) and a time until polyploidization (γ_b), as per S4. P5: The event times γ_{a1}, γ_{a2}, γ_b, and σ_b are added to the queue such that the queue remains sorted from the earliest to the latest event.

For autopolyploidy events, the daughter species will have a ploidal level double that of the parent species. The parent species will be given a new polyploidization waiting time and the daughter species both a polyploidization and speciation waiting time as described above.

Extinction is modeled using an exponential waiting time with rate $\mu = 0.1$. When each new species is formed, a random deviate from an exponential distribution with rate μ is drawn, and that plus the current simulation clock time becomes the extinction time for that species. Once the simulation clock reaches the extinction time, that species is removed from all queues. If a species is waiting to form an allopolyploid when extinction occurs, the daughter species are not created.

Parameter Estimation

Of the parameters included in the model (r_k, a, h_k, T_g, N_g, and μ), we used fixed values for T_g, N_g, and μ. The total genus size (N_g) and age of the genus (T_g) were taken from the literature; where no T_g was reported, we assumed that species had an age equal to $T_g = 22$ million years; our data are deposited in Dryad Digital Repository, http://dx.doi.org/10.5061/dryad.35r23 (Scarpino et al. 2014). Ages were taken from the literature and the Missouri Botanical Garden Index to Plant Chromosome Numbers, spanning the years 1967 to 2000 (Moore 1973; Goldblatt 1981; Missouri Botanical Garden 2005), data deposited in Dryad Digital Repository, http://dx.doi.org/10.5061/dryad.35r23 (Scarpino et al. 2014). To estimate the remaining parameters, we employed the aforementioned simulation and the model-fitting procedure approximate Bayesian computation—sequential Monte Carlo (ABC-SMC; Beaumont et al. 2002; Toni et al. 2008); data deposited in Dryad Digital Repository, http://dx.doi.org/10.5061/dryad.35r23 (Scarpino et al. 2014). We used ABC-SMC because closed-form expressions for the likelihood equation necessary to calculate the joint-posterior distribution of r_k, h_k, and a do not exist. The lack of a closed-form likelihood equation derives primarily from the presence of allopolyploidy in the model.

We now briefly describe the ABC-SMC parameter estimation procedure (see Toni et al. 2008 for a complete discussion of the methodology). If a model M with key parameter θ, which has a prior distribution $\pi(\theta)$, generates some data D, then the posterior distribution for θ is given by

$$f(\theta|D) = \frac{\Pr(D|\theta)\pi(\theta)}{\int \Pr(D|\theta)\pi(\theta)d\theta}.$$

Since it is not possible to calculate the likelihood $\Pr(D \mid \theta)$ directly for our model, we implement the following approximate method for estimating $f(\theta \mid D)$. (1) Select a vector of decreasing acceptance levels, $\mathbf{e} = \{\varepsilon_1, \varepsilon_2, ..., \varepsilon_n\}$, such that $\varepsilon_1 > \varepsilon_2 > ... > \varepsilon_n$. (2) Select a final number of acceptances, S, for each level in vector $\mathbf{e}$ (3) Generate a parameter set Θ by selecting a random deviate from the joint prior distribution $\Theta = \pi(\theta_{h_k}) \circ \pi(\theta_{r_k}) \circ \pi(\theta_a)$ (the prior distributions were all uniform over a broad range of values; see table A1, available online). (4) Simulate a data set D_r with parameter set Θ. (5) Calculate the distance between the observed data, D_{obs}, and the simulated data, D_r, using a set of prespecified distance metrics $\delta(D_{obs}, D_r)$. In our case, we calculated the Euclidean distance on nine summary statistics that have been transformed using partial least squares regression. For a description of the summary statistics, see the appendix, available online, and for the value of each summary statistic in each genus, see data

deposited in the Dryad Digital Repository, http://dx.doi.org/10.5061/dryad.35r23 (Scarpino et al. 2014). (6) Accept Θ if $\delta(D_{obs}, D_r) < \varepsilon_i$ (7); repeat steps 3–5 until S sets of parameters are accepted, each time recording the parameter set and the corresponding distance, $\delta(D_{obs}, D_r)$. (8) Once S sets of parameters are accepted, update the prior distributions to be equivalent to the distribution of accepted parameters, and add noise to the distributions using a uniform kernel (see appendix for a description of the noise kernel). (9) Update the tolerance to ε_{i+1} and repeat steps 2–6. (10) Once S sets of parameters are accepted at tolerance ε_n, the algorithm stops and the accepted parameter sets are taken as the joint posterior distribution over the parameters Θ.

Comparative Analyses

To assess whether there is evidence of a polyploid advantage or disadvantage in terms of relative evolutionary rates across our study genera, we considered parameter estimates under the complex ratchet model with $\mu = 0$. The reason for setting the extinction rate to zero is discussed below, but importantly, this was a statistically conservative decision. To evaluate the fit of the simple ratchet model, we used the goodness-of-fit test described below. We controlled for phylogenetic nonindependence using a method proposed by Lajeunesse (2009) for use in meta-analyses. First, we constructed a phylogeny of angiosperm genera using data from Jansen et al. (2007). To fill in taxa not represented in the Jansen et al. (2007) data set, we used family level relationships, while maintaining the original topology. We then calculated the difference between the diploid and polyploid evolutionary rates; for example, the diversification rate difference for each genus would be $r_d - r_p$, and the pooled variance $\sigma^2(r_d - r_p)$. Using a random effects model that incorporates a variance/covariance matrix derived from the phylogenetic relationships, we determined the phylogenetically independent effect size and 95% confidence interval (Lajeunesse 2009). Finally, we determined whether each difference in evolutionary rate was significantly different from zero. These calculations were performed using the software package PHYLOMETA (Lajeunesse 2011).

Polyploidy and Species Richness

To investigate the effect of polyploidy on within-genus species richness (the total number of species), we compared two models: one where $h = 0$ and another where $h > 0$. Under each model, we simulated going forward in time, recording the absolute number of species at time T_G. The net speciation rates of diploid and higher ploids were set to be equal in both models ($\lambda_p = \lambda_d$), allowing us to focus on the relative contribution of polyploid formation to within-genus species richness.

Goodness-of-Fit Test

We used a goodness-of-fit test to determine whether, for each of the 60 genera included in this study, the simple ratchet model could generate distributions of ploidal levels statistically similar to the empirical distributions. Because the number of observations in individual ploidal classes is often too low to perform a standard χ^2 test, null distributions must be generated via Monte Carlo simulation, originally described in Meyers and Levin (2006). Briefly, the estimated parameters values are used to simulate 10,000 ploidal level distributions for each genus. For every simulated distribution we calculate a χ^2 statistic,

$$\chi^2_g = \sum_{k=2}^{16} \frac{(O_k - E_k)^2}{E_k}$$

where O_k is the simulated or observed number of species in ploidal class k and E_k is the actual or expected number. By aggregating each of the 10,000 χ^2 statistics, we can calculate a P value, which is the proportion of this distribution that is greater than or equal to the χ^2 statistic calculated using the simple ratchet model.

Results

Estimating and Comparing Evolutionary Rates

Using the goodness-of-fit test, we find that the simple ratchet model of polyploid evolution introduced in Meyers and Levin (2006) is sufficient to account for the distribution of ploidal levels in 47 of the 60 genera considered (fig. 1). This model assumes that congeneric diploids and polyploids speciate at equal rates, polyploid formation is irreversible, and all polyploids are allopolyploids. While we find less support for the simple ratchet model than Meyers and Levin (2006), it can account for the diversity observed in a majority of genera considered.

Using the complex ratchet model with no extinction, which allows different diversification rates for diploids and higher ploidal species, we estimated and compared these rates. We found statistical support for a diploid speciation advantage (fig. 2). Using phylogenetic group as a random effect, we calculated a phylogenetically independent difference between diploid and polyploid net speciation rates: $\text{diff} = (r_d + h_d) - (r_p + h_p)$. The estimated difference was 0.101 (95% confidence interval, 0.093 – 0.110, $P < 1 \times 10^{-4}$), indicating a diploid advantage. Not surprisingly, given the broad phylogenetic distribution of our genera, there was no significant effect of phylogenetic nonindependence on the results. These results were further

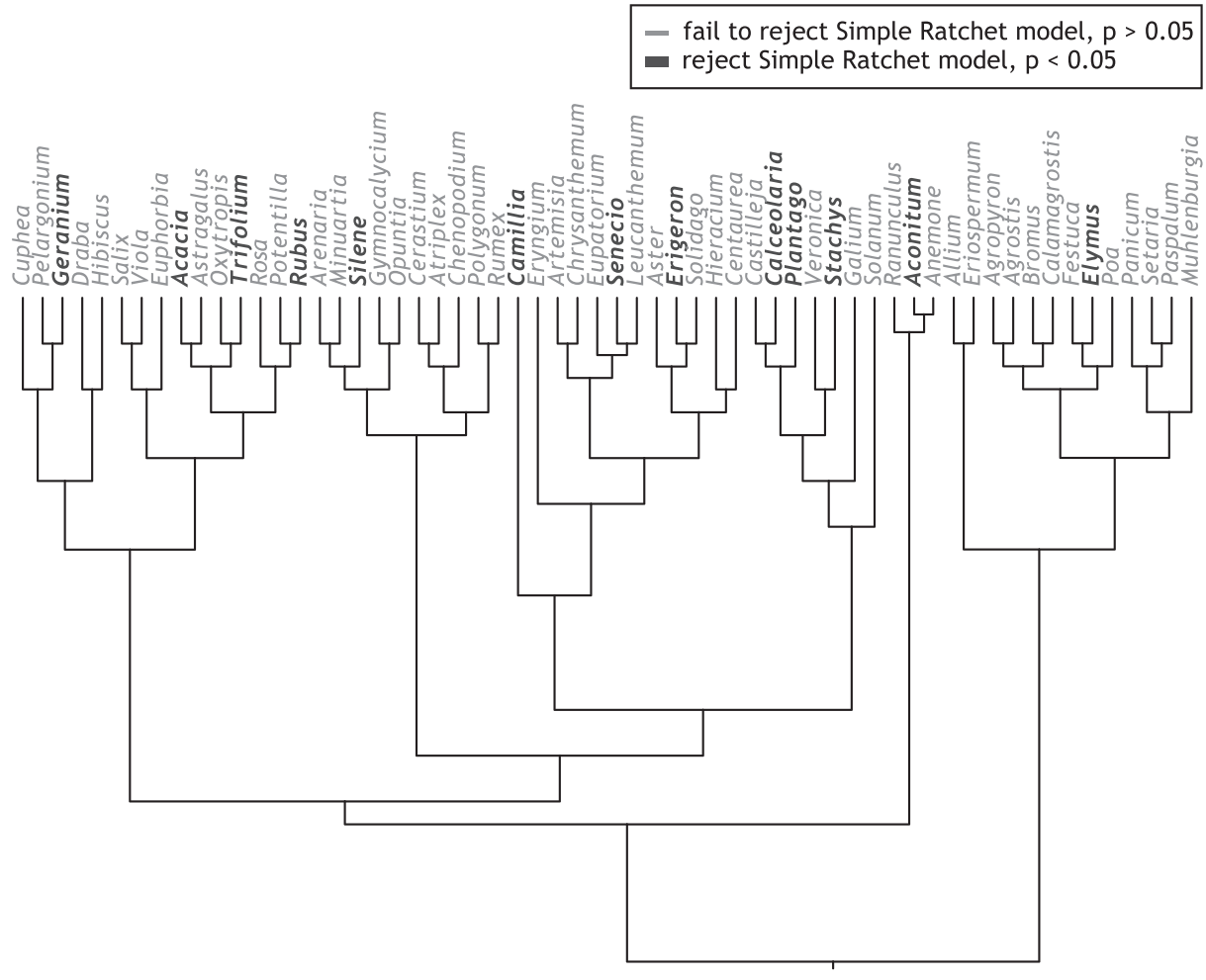

Figure 1: The phylogeny of genera included in this study, with genera supporting the simple ratchet model indicated with a thinner gray font and genera failing to support the simple ratchet model indicated with thicker black font. Using a simulation-based goodness-of-fit test, the simple ratchet model was sufficient in 47 out of 60 genera at the $\alpha = 0.05$ level.

supported using a Welch's two-sample *t*-test (ignoring phylogeny; $t = 4.944$, $P < 3.7 \times 10^{-6}$). This diploid advantage was driven primarily by a difference in the polyploidization rates ($r_d - r_p > 0$, $P = 0.054$; $h_d - h_p > 0$, $P < .003$).

To investigate the effect of extinction, we compared versions of the complex ratchet model with and without extinction, $\mu = 0.1$ and $\mu = 0$. Intuitively, extinction caused an increase in the estimates of all speciation rates. However, extinction yielded a larger discrepancy between the estimated net speciation rate of diploids and polyploids, driven primarily by differences in polyploidization rates, as determined using the proportional increase in diversification rate (Welch *t*-test comparing $r_d^{\mu=0.1}/r_d^{\mu=0.0}$ and $r_p^{\mu=0.1}/r_p^{\mu=0.0}$, $t = 1.325$, $P = .19$) and polyploidization rate: (Welch *t*-test comparing $h_d^{\mu=0.1}/h_d^{\mu=0.0}$ and $h_d^{\mu=0.1}/h_d^{\mu=0.0}$, $t = 5.336$, $P = 1 \times e{-}6$).

We also considered models where the extinction rate was either a free parameter or a function of the diversification rate, *r*. However, this introduced a substantial amount of uncertainty in all parameter estimates, and we were unable to estimate the relative speciation rates of diploids and polyploids.

The Role of Allopolyploidy and Autopolyploidy

We estimated that allopolyploidy (as opposed to autopolyploidy) was responsible for the majority of polyploidy events in most genera (fig. 3). Importantly, individual species are not identified as being allo- or autopolyploids; instead, we estimate the fraction of autopolyploids in each genus. However, this approach suffered from low statistical power to estimate the fraction autopolyploid. Most of the statistical information on the fraction of polyploidy events

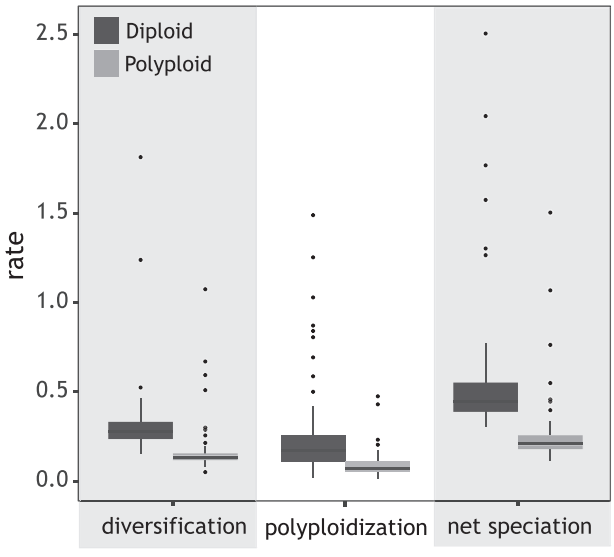

Figure 2: Boxplots with the median, interquartile, and range for the estimated diversification (r), polyploidization (h), and net speciation rates ($\lambda = r + h - \mu$) from each of the 60 genera. Using phylogenetic group as a random effect, we calculated a phylogenetically independent difference between the mean diploid and polyploid net speciation rates across all 60 genera. The estimated difference was 0.101 (95% credible interval 0.093–0.110, $P < 1 \times 10^{-4}$), indicating a diploid advantage. This diploid advantage was driven primarily by a difference in the polyploidization rates ($r_d - r_p$, $P = .054$; $h_d - h_p$, $P < .003$). These tests include uncertainty in the estimated model parameters, which is not presented in the figure.

due to autopolyploidy derives from the number of 6n–14n species. The model assumes diploid foundry, and therefore under a model with autopolyploidy alone, $a = 1$, no species with 6n–14n genomes can be formed. The vast majority of species considered have ploidal levels less than 6n.

Polyploidy and Species Richness

Polyploidization increases diversity by directly generating new species and creating lineages able to undergo further diversification and thus can fuel increases in species richness even if polyploids diversify at the same rate as (or even slower than) diploids. For example, in *Eriospermum*, the polyploid net speciation rate is estimated to be equal to the diploid net speciation rate, and the simple ratchet model projects that polyploid species will grow to dominate the genus (fig. 4). In this way, polyploidization itself can explain the increased species richness associated with clades containing more polyploids, without invoking an intrinsic evolutionary advantage of polyploids.

Discussion

We have analyzed ploidal distribution data from 7,271 species in 60 flowering plant genera representing 25 families across 17 orders and found evidence that diploids speciate faster than their congeneric polyploids. This diploid advantage stems primarily from a higher rate of polyploidization in diploids than polyploids. Our model also accounts for the empirically observed, positive correlation between average ploidal-level and species richness, without assuming a polyploid speciation advantage.

In their original article, Meyers and Levin (2006) introduced a simple quantitative model of angiosperm evolution in which polyploids held no evolutionary advantage over diploids and polyploidy events were irreversible. This model was able to account for the distribution of ploidal classes across nine angiosperm genera. Although this model is quite general, Meyers and Levin (2006) made a number of simplifying assumptions: (1) congeneric diploids and higher ploids shared a single diversification and polyploidization rate, (2) all polyploids were allopolyploids, (3) no extinction, and (4) deterministic evolution. Here, we have advanced our understanding of angiosperm evolution by relaxing these assumptions and analyzing the broadest phylogenetic distribution of species yet considered in a study of polyploid evolution.

Forty-seven of the 60 flowering plant genera considered have ploidal distributions statistically consistent with the simple Meyers and Levin (2006) ratchet model. These genera included 5,435 species, representing 23 of the 25 families and 16 of the 17 orders contained in the full data set. We therefore conclude that this provides support for the utility of the ratchet model of polyploid evolution. Many of the genera not supporting the simple ratchet model clearly violate one or more of the assumptions made by that model. For example, *Silene* had the highest ratio of diploids to tetraploids and *Geranium* the lowest, and the simple ratchet model was rejected for both. The simple ratchet model was also rejected for *Acacia* and *Rubus*, which both have a high prevalence of autopolyploids, violating a key assumption of the model, that polyploids are allopolyploids (Mandal and Ennos 1995; Thompson 1997). Although variation in diversification rates has undoubtedly contributed to the expansion of some angiosperm genera, the abundance of polyploids observed in many lineages can be explained without assuming that polyploids are superior to diploids. This finding highlights again the importance of irreversibility in shaping evolutionary trajectories (Bull and Charnov 1985; Gray et al. 2010).

At first glance, our assumption about polyploidy irreversibility may seem at odds with the recent suggestion that all extant angiosperms have descended from a polyploid ancestor (Levin 2002; Soltis et al. 2009; Amborella

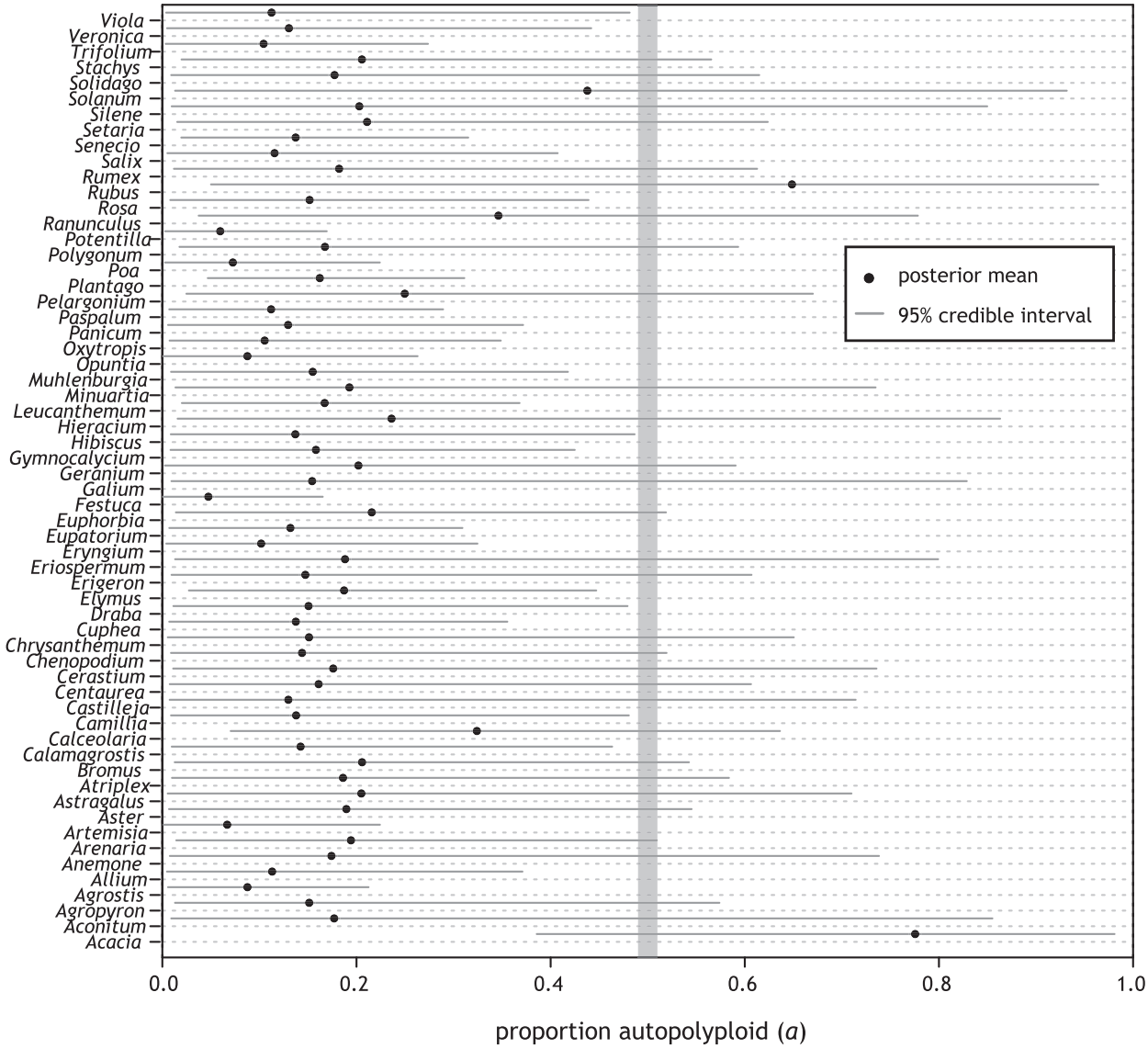

Figure 3: The estimated proportion of polyploid lineages formed due to autopolyploidy (rather than allopolyploidy) for each genus (*a*). Error bars represent the 95% credible intervals of the posterior distribution. Only *Acacia* and *Rubus* are estimated to have a proportion of autopolyploids greater than 0.5 (gray vertical bar), indicating that more than half of the polyploidy events in those genera were due to autopolyploidy.

Genome Project 2013). However, this assumption need hold only over the evolutionary timescales considered in each simulation. For example, Jiao et al. (2011) estimated that two ancestral whole genome duplications shared by all angiosperms occurred between 100–300 and 200–450 million years ago, whereas the genera in our study have a mean age of 16 million years. Thus, our study assumes irreversibility over only a relatively short and recent period of angiosperm evolution. This assumption is further supported by Mayrose et al. (2011), who found statistical support for the irreversibility of polyploidy.

Successful polyploidization is infrequent, in part because fertilization events in which unreduced gametes unite to form viable and fertile polyploids are rare (Soltis and Soltis 1999). However, unreduced gamete formation is more prevalent in interspecific crosses, a finding in support of our result that allopolyploids may be more common than autopolyploids (Ramsey and Schemske 1998; Brownfield and Kohler 2011; De Storme and Geelen 2013; Tayale and Pariso 2013). These nascent polyploids then must overcome a substantial minority disadvantage (Levin 1975). Consequently, newly formed polyploids frequently

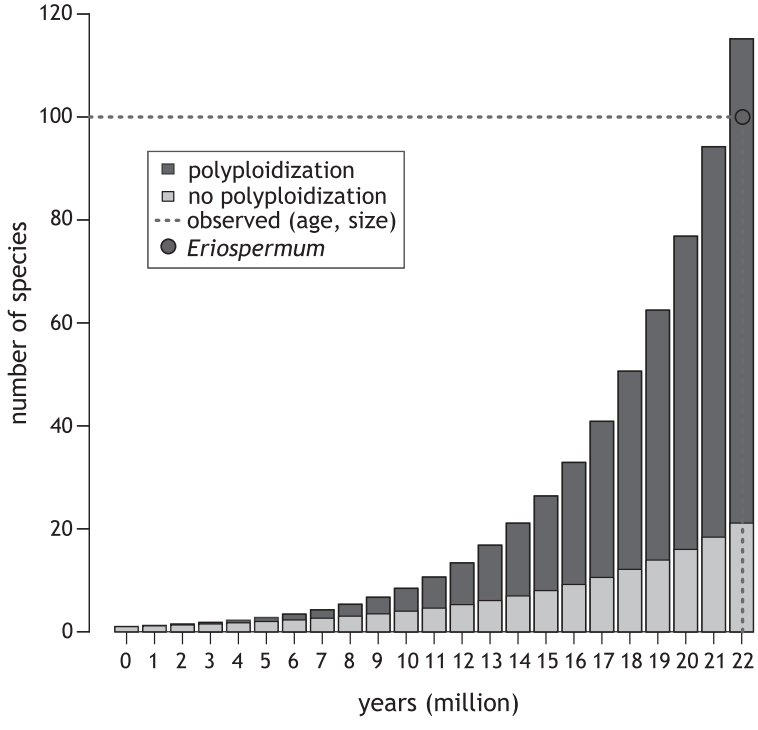

Figure 4: The correlated ascent of polyploids and species richness in the genus *Eriospermum*. Using the estimated values for the diversification and polyploidization rates for *Eriospermum*, we projected the number of species over time (in millions of years), both with (dark) and without (gray) polyploids. The greater than twofold difference in the number of species at 22 Ma suggests that the simple ratchet model can generate the observed correlation between species richness and average ploidal level (without assuming any polyploid superiority).

exhibit characteristics such as high levels of self-fertilization, assortative mating, divergent habitat preferences, or a substantial fitness advantage over their progenitors (Soltis and Soltis 1999; Otto 2007). Given these restrictive conditions surrounding polyploid emergence in the presence of their lower-ploid progenitors, our model distinguishes the polyploid lineage formation process from subsequent diversification.

Importantly, our model considers only random extinction. Since we assume all genera are founded by diploids, random extinction ultimately reduces diploid diversity proportionally more than polyploid diversity. For this reason, we presented results from a version of the complex ratchet model without extinction. However, as stated earlier, the diploid speciation advantage increased in the presence of random extinction. Incorporating nonrandom extinction; for example, an upper bound on viable genome size may impact the evolutionary rate estimates, which we cannot predict a priori. Such extensions of our model to include more complex evolutionary processes should prove insightful.

We lacked statistical power to precisely estimate the proportion of polyploid lineages founded by autopolyploidy versus allopolyploidy. Our method gauges the balance between the two forms of polyploidy primarily from observed numbers of 6n, 10n, 12n, and 14n species, which cannot be created via autopolyploidy alone in our model. Because the model assumes diploid foundry, if all polyploids are autopolyploids, then only 2n, 4n, 8n, and 16n species can be formed. However, there were too few higher ploidal species in most of our focal genera to estimate the relative importance of autopolyploidy, and it thus remains an important topic for future investigation.

Our results are consistent with an emerging consensus that polyploids do not have a speciation advantage over related diploids. Wood et al. (2009) failed to detect increased diversification rates in lineages with higher ploidal levels using a method of nonnested sister group contrasts, and Mayrose et al. (2011) found evidence for a decrease in the speciation rate of recently formed polyploids. Interestingly, five of our genera were included in the Mayrose et al. (2011) analysis, and despite different data and methods, our qualitative conclusions agree (see appendix and fig. A1, available online). Importantly, the Mayrose et al. (2011) method considered phylogenetic relationships within genera, while ours does not.

Speculation about the advantage of polyploids has been motivated, in part, by the positive relationship between

species diversity and the incidence of polyploidy. We demonstrate that this relationship arises naturally from polyploidy irreversibility, and can occur even if polyploids have an evolutionary disadvantage. Concordantly, Vamosi and Dickinson (2006) reported a correlation between polyploid incidence and species richness in Rosaceae, yet found no evidence that polyploids diversified faster than their diploid counterparts. They concluded that ploidal evolution alone could account for the pattern. Additionally, when Mayrose et al. (2009) recreated chromosome number evolution in *Helianthus*, they hypothesized that major polyploid events are followed by depressed speciation rates. These studies have led some to conclude that polyploids should, in fact, be considered evolutionary dead-ends (Arrigo and Barker 2012).

Our results suggest that a simple ratchet model for polyploid evolution can explain the within-genus distribution of ploidal levels across many angiosperm groups, with the caveat that our data set contained 44 core eudicots, 3 noncore eudicots, and 13 monocots. We argue that this model should serve as a null model for future studies on polyploidy and diversity. We have taken this approach in evaluating more complex evolutionary drivers and find statistical evidence for a diploid advantage driven by the relatively frequent emergence of new tetraploid species. We conclude that the rise of polyploids and the concomitant rise of biodiversity do not require the evolutionary superiority of polyploids. Nonetheless, polyploidy as a process should be considered a central driver of evolutionary diversification (Freeling 2009; Soltis et al. 2009; Freeling et al. 2012; Huminiecki and Conant 2012; Garsmeur et al. 2014; Mayfield-Jones et al. 2013).

Acknowledgments

We thank R. Guerrero, M. Kirkpatrick, S. Otto, and several anonymous reviewers for helpful comments on earlier versions of the manuscript and insightful discussions of polyploid evolution. We acknowledge funding support from a National Science Foundation Graduate Research Fellowship, the Santa Fe Institute, and the Omidyar Foundation to S.V.S.

Literature Cited

Amborella Genome Project. 2013. The *Amborella* genome and the evolution of flowering plants. Science 342:1241089.

Arrigo, N., and M. S. Barker. 2012. Rarely successful polyploids and their legacy in plant genomes. Current Opinion in Plant Biology 15:140–146.

Beaumont, M., W. Zhang, and D. Balding. 2002. Approximate Bayesian computation in population genetics. Genetics 162:2025–2035.

Brownfield, L., and C. Köhler. 2011. Unreduced gamete formation in plants: mechanisms and prospects. Journal of Experimental Botany 62:1659–1668.

Bull, J. J., and E. L. Charnov. 1985. On irreversible evolution. Evolution 39:1149–1155.

De Storme, N., and D. Geelen. 2013. Sexual polyploidization in plants: cytological mechanisms and molecular regulation. New Phytologist 198:670–684.

Fawcett, J. A., S. Maere, and Y. Van der Peer. 2009. Plants with double genomes might have had a better chance to survive the Cretaceous-Tertiary extinction event. Proceedings of the National Academy of Sciences of the USA 106:5737–5742.

Freeling, M. 2009. Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. Annual Review of Plant Biology 60:433–453.

Freeling M., M. R. Woodhouse, S. Subramaniam, G. Turco, D. Lisch, and J. C. Schnable. 2012. Fractionation mutagenesis and similar consequences of mechanisms removing dispensable or less expressed DNA in plants. Current Opinion in Plant Biology 15:131–139.

Garsmeur O., J. C. Schnable, A. Almeida, C. Jourda, A. D'Hont, and M. Freeling. 2014. Two evolutionarily distinct classes of paleopolyploidy. Molecular Biology and Evolution 31:448–454.

Goldblatt, P., ed. 1981. Index to plant chromosome numbers 1975–1978. Monographs in Systematic Botany. Missouri Botanical Garden, St. Louis.

Grant, V. 1981. Plant speciation. 2nd ed. Columbia University Press, New York.

Gray, M. W., J. Lukes, J. M. Archibald, P. J. Keeling, and W. F. Doolittle. 2010. Irremediable complexity? Science 330:920–921.

Huminiecki, L., and Conant, G. C. 2012. Polyploidy and the evolution of complex traits. International Journal of Evolutionary Biology 2012-292068:1–2.

Jansen, R. K., Z. Cai, L. A. Raubeson, H. Daniell, C. W. Depamphilis, J. Leebens-Mack, K. F. Müller, et al. 2007. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. Proceedings of the National Academy of Sciences of the USA 104:19369–19374.

Jiao, Y., N. J. Wickett, S. Ayyampalayam, A. S. Chanderbali, L. Landherr, P. E. Ralph, L. P. Tomsho, et al. 2011. Ancestral polyploidy in seed plants and angiosperms. Nature 473:97–100.

Lajeunesse, M. J. 2009. Meta-analysis and the comparative phylogenetic method. American Naturalist 174:369–381.

———. 2011. *phyloMeta*: a program for phylogenetic comparative analyses with meta-analysis. Bioinformatics 27:2603–2604.

Levin, D. A. 1975. Minority cytotype exclusion in local plant populations. Taxon 24:35–43.

———. 1983. Polyploidy and novelty in flowering plants. American Naturalist 122:1–25.

———. 2002. The role of chromosomal change in plant evolution. Oxford University Press, New York.

Mandal, A., and R. Ennos. 1995. Mating system analysis in a natural population of *Acacia nilotica* subspecies *kraussiana*. Forest Ecology and Management 79:235–240.

Mayfield-Jones D., J. D. Washburn, T. Arias, P. P. Edger, J. C. Pires, and G. C. Conant. 2013. Watching the grin fade: tracing the effects of polyploidy on different evolutionary time scales. Seminars in Cellular and Developmental Biology 24:320–331.

Mayrose, I., M. S. Barker, and S. P. Otto. 2009. Probabilisitic models

of chromosome number evolution and the inference of polyploidy. Systematic Biology 59:132–144.
Mayrose, I., S. H. Zhan, C. J. Rothfels, K. Magnuson-Ford, M. S. Barker, L. H. Rieseberg, and S. P. Otto. 2011. Recently formed polyploid plants diversify at lower rates. Science 333:1257.
Meyers, L. A., and D. A. Levin. 2006. On the abundance of polyploids in flowering plants. Evolution 60:1198–1206.
Missouri Botanical Garden. 2005. Tropicos: Botanical Information System at the Missouri Botanical Garden. Accessed March 2014. www.tropicos.org.
Moore, R. J., ed. 1973. Index to plant chromosome numbers 1967–1971. Monographs in Systematic Botany. Missouri Botanical Garden, St. Louis.
Otto, S. P. 2007. The evolutionary consequences of polyploidy. Cell 131:452–462.
Otto, S. P., and J. Whitton. 2000. Polyploid incidence and evolution. Annual Revue of Genetics 34:401–437.
Ramsey, J., and D. W. Schemske. 1998 Pathways, mechanisms, and rates of polyploid formation in flowering plants. Annual Review of Ecology and Systematics 29:467–501.
Scarpino, S. V., D. A. Levin, and L. A. Meyers. 2014. Data from: Polyploid formation shapes flowering plant diversity. American Naturalist, Dryad Digital Repository, http://dx.doi.org/10.5061/dryad.35r23.
Soltis, D. E., V. Albert, J. Leebens-Mack, C. Bell, A. Paterson, C. Zheng, D. Sankoff, C. Depamphilis, P. Wall, and P. S. Soltis. 2009. Polyploidy and angiosperm diversification. American Journal of Botany 96:336–348.
Soltis, D. E., and J. G. Burleigh. 2009. Surviving the K-T mass extinction: new perspectives of polyploidization in angiosperms. Proceedings of the National Academy of Sciences of the USA 106: 5455–5456.
Soltis, D. E., and P. S. Soltis. 1999. Polyploidy: recurrent formation and genome evolution. Trends in Ecology and Evolution 14:348–352.
Soltis, D. E., P. S. Soltis, and J. A. Tate. 2003. Advances in the study of polyploidy since *Plant Speciation*. New Phytologist 161:173–191.
Stebbins, G. L. 1971. Chromosomal evolution in higher plants. Addison-Wesley, Reading, MA.
———. 1980. Polyploidy: future prospects. Pages 495–520 *in* W. H. Lewis, ed. Polyploidy: biological relevance. Springer, New York.
———. 1985. Polyploidy, hybridization, and the invasion of new habitats. Annals of the Missouri Botanical Gardens 72: 824–832.
Tayale, A., and C. Parisod. 2013. Natural pathways to polyploidy in plants and consequences for genome reorganization. Cytogenetic and Genome Research 140:79–96.
te Beest, M., J. J. Le Roux, D. M. Richardson, A. K. Brysting, J. Suda, M. Kubešová, and P. Pyšek. 2012. The more the better? the role of polyploidy in facilitating plant invasions. Annals of Botany 109: 19–45.
Thompson, M. M. 1997. Survey of chromosome numbers in *Rubus* (*Rosacea: Rosoidea*). Annals of the Missouri Botanical Garden 84: 128–164.
Toni, T., D. Welch, N. Strelkowa, A. Ipsen, and M. P. H. Stumf. 2008. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. Journal of the Royal Society Interface 6:1–16.
Vamosi, J. C., and T. A. Dickinson. 2006. Polyploidy and diversification: a phylogenetic investigation in *Rosaceae.* International Journal of Plant Sciences 167:349–358.
Wood, T. E., N. Takebayashi, M. S. Barker, I. Mayrose, P. B. Greenspoon, and L. H. Rieseberg. 2009. The frequency of polyploid speciation in vascular plants. Proceedings of the National Academy of Sciences of the USA 106:13875–13879.

Associate Editor: Tia-Lynn Ashman
Editor: Judith L. Bronstein

"At first I thought that the plant could be nothing more than a curious form of *Leucanthemum vulgare* Lam., that it was nothing more than the result of a mere freak of nature; and when, on the 10th inst., I went in search of more specimens, I half expected to find the new form and the common one growing on one and the same stem. But although I found specimens by scores, not a stem among them all had the two forms upon it." From "A Supposed New Columbine, and a New Ox-eye Daisy" by Sanborn Tenney (*The American Naturalist*, 1867, 1:388–389).