Evolution, 60(6), 2006, pp. 1198–1206

ON THE ABUNDANCE OF POLYPLOIDS IN FLOWERING PLANTS

LAUREN ANCEL MEYERS[1,2,3,4] AND DONALD A. LEVIN[1,5]
[1]*Section of Integrative Biology, University of Texas at Austin, Austin, Texas 78712*
[2]*Institute for Cellular and Molecular Biology, University of Texas at Austin, Austin, Texas 78712*
[3]*Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501*
[4]*E-mail: laurenmeyers@mail.utexas.edu*
[5]*E-mail: dlevin@uts.cc.utexas.edu*

Abstract.—The wide distribution of polyploidy among plants has led to a variety of theories for the evolutionary advantages of polyploidy. Here we claim that the abundance of polyploidy may be the result of a simple ratcheting process that does not require evolutionary advantages due to the biological properties of organisms. The evolution of polyploidy is a one-way process in which chromosome number can increase but not decrease. Using a simple mathematical model, we show that average ploidal level within a plant lineage can continually increase to the levels observed today, even if there are ecological or physiological disadvantages to higher ploidy. The model allowed us to estimate the average net speciation and polyploidy rates for ten angiosperm genera. Based on these estimates, the model predicts distributions of ploidal levels statistically similar to those observed in nine of the 10 genera.

Key words.—Evolution, flowering plants, irreversible, polyploidy, speciation.

Received November 9, 2005. Accepted April 7, 2006.

Polyploidy is a common feature of flowering plants. Over 70% of all angiosperm species have a ploidal level increase somewhere in their evolutionary histories (Levin 2002). Some polyploids are of recent origin, while others are tens of millions of years old (Stebbins 1971; Levin 2002). The proportion of polyploid species has increased over time, as indicated by increasing guard cell size (a correlate of ploidal level) in the Lauraceae, Magnoliaceae, and Platanaceae over the past hundred million years (Masterson 1994). Otto and Whitton (2000) estimate that the rate of polyploidization is between two and four percent of the rate of speciation. Thus, polyploidy is an active, continuous process contributing to an ever-increasing percentage of polyploids. Ramsey and Schemske (1998) estimate that autopolyploids are formed at a higher rate than allopolyploids. Given that the majority of angiosperm polyploids have hybridity in their ancestry (Levin 2002), it appears that allopolyploids have an advantage over their autoploid relatives in forming long-lived species.

Stebbins (1971) proposed that young polyploid complexes would contain many diploids and a few tetraploids. As the complex aged, a greater proportion of species would be polyploid, and higher ploidal levels would develop. As time progresses, the ancestors of the polyploids would gradually die out, leaving taxa whose ancestry would be difficult to document. Although Stebbins' proposition is widely accepted, the mechanistic basis for the prevalence of polyploids in contemporary floras and for their ascendance over time remains unclear.

There has been considerable speculation on the properties of polyploids that afford them an advantage over their diploid counterparts. Given that the majority of polyploids are the products of hybridization as well as of ploidal change, there is one school of thought that holds that fixed heterozygosity is the key to polyploid success (Thompson and Lumaret 1992; Soltis and Soltis 1993; Jiang et al. 1998). This view is supported by evidence that vigor is a positive function of heterozygosity in autopolyploids (e.g., rye, Lunquist 1966; alfalfa, Bingham 1980; *Dactylis glomerata*, Tomekpe and Lumaret 1991).

Another possibility is that the emergence and prevalence of polyploids arises from their novel phenotypes, and expanded and divergent ecological tolerances. These attributes may be due to ploidal level change per se, the union of disparate genomes, recombination and trangressive segregation, and nucleocytoplasmic interactions (Stebbins 1980, 1985; Levin 2000; Otto and Whitton 2000; Ramsey and Schemske 2002). Recent studies have demonstrated that divergent morphology, physiology, and life-history traits also may emerge in newly constituted polyploids as a result of mutations and epigenetic effects (Levy and Feldman 2004; Pires et al. 2004; Schranz and Osborn 2004; Adams and Wendel 2005).

It is difficult to determine the extent to which heterozygosity and/or novelty explain the ascendance of polyploids within lineages or in angiosperms as a whole. Perhaps these factors have only been of minor importance. Indeed without knowing the evolutionary histories of diploid and related polyploidy lineages, we do not actually know that (in general) polyploids are superior to their diploid relatives. If they were, then polyploid lineages would survive longer than diploid lineages, and polyploid lineages would thus give rise to more new species.

Here, we consider the hypothesis that the key to polyploid ascendance lies not in their vigor, nor in their novel ecological attributes, nor in other properties of their biologies. Rather, polyploids have been increasing over time primarily because polyploidy is largely irreversible (Stebbins 1971; Grant 1981). For example, diploid species beget diploid and tetraploid species, whereas the latter beget tetraploids and higher ploidal types. Although chromosome loss can occur, reversions to original ploidal level are probably quite rare. As discussed by Raven and Thompson (1964) and De Wet (1968), polyploidy reversals will result in polyhaploids that are apt to be sterile, and thus evolutionary dead ends (Stebbins 1980; Grant 1981). In a recent review of the subject, Ramsey and Schemske (2002) showed that polyhaploids have reduced growth, survivorship, and/or reproduction, further supporting the irreversibility of polyploidy.

Using a simple mathematical model, we show that the av-

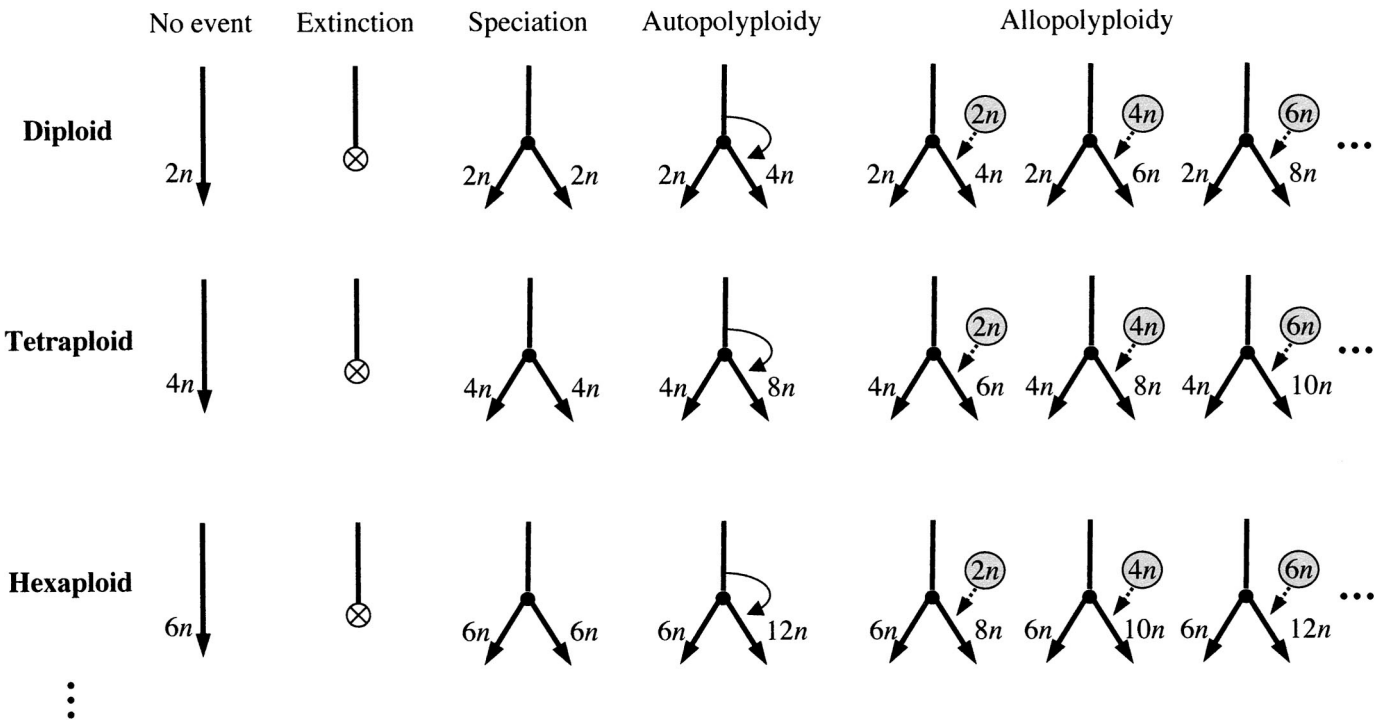

FIG. 1. A simple model for evolution of polyploidy. Within a small time interval, a species can remain unchanged, go extinct, diverge into two species with the same chromosome numbers, speciate via autopolyploidy, or speciate via allopolyploidy. The product of allopolyploidy events depends on the relative frequencies of ploidal levels within the genus.

erage ploidal level within lineages can increase continually when the net speciation rate of polyploids is greater than or equal to that of diploids. Moreover, polyploids may increase to substantial frequencies even if they have an evolutionary disadvantage (lower speciation rates) to their diploid counterparts.

THE MODEL AND METHODS

Polyploid Ratchet Model

Here we introduce a simple ratchet model for the evolution of ploidal levels within a genus. During a small unit of time, a species can go extinct, speciate (split into two species), undergo autopolyploidy or hybridize to produce a higher ploidal species, or not change at all (Fig. 1). We assume that the outcome of a hybridization event depends on the relative frequencies of all ploidal levels in the genus at the time of the event. For example, if there are equal numbers of diploid and tetraploid species, then a hybrid event on a diploid lineage is equally likely to produce a tetraploid or hexaploid. We begin by introducing a general model for the evolution of polyploidy in angiosperms, and then simplify it into a more tractable two-type birth-only (Yule) branching process. Birth-only processes have been used previously for the estimation of macroevolutionary rates in angiosperms (Sanderson and Donohue 1994; Baldwin and Sanderson 1998; Nee 2001).

We assume that a genus is founded by a single diploid species and track the changing numbers of species at each ploidal level through time. Let x_k denote the number of species of ploidal level k, for all even values of k. We assume that species are sexual and ignore odd values of k (e.g., triploids), even when they exist. In a given time unit, a species of ploidal level k will go extinct with probability μ_k and speciate by means other than polyploidization with probability λ_k. We model this as a birth-only process in which speciation (not including polyploidization) occurs at a net rate of $s_k = \lambda_k - \mu_k$, and refer to s_k as the net speciation rate. Furthermore, a species will undergo autopolyploidy with probability t per time unit, and hybridize to produce an allopolyploid with probability p per time unit, regardless of ploidal level (Fig. 1). The expected numbers of species at each ploidal level will change as given by the following recursive equations:

$$x_k' = \sum_{i=1}^{(k-2)/2} \left(x_{2i} \cdot p \cdot \frac{x_{k-2i}}{\sum_{j=1}^{\infty} x_{2j}} \right) + x_{k/2}t + x_k(1 + s_k) \qquad (1)$$

for all ploidal levels (even values of k). For example, the expected number of octoploid species will change according to

$$x_8' = (2x_2x_6 + x_4x_4)\frac{p}{\sum_{j=1}^{\infty} x_{2j}} + x_4t + x_8(1 + s_8). \qquad (2)$$

For the purposes of analysis, we simplify the model by assuming that all higher ploidy species share the same net speciation rate ($s_p = s_4 = s_6 = s_8 = \cdots$) that may or may not differ from the net speciation rate of diploids (s_d), and then consider only changes in the number of diploids and the total number of polyploids. This yields a 2×2 matrix in which the ijth element m_{ij} is the expected number of species

of type i that will arise from a species of type j in a single time unit:

$$M = \begin{pmatrix} 1 + s_d & p + t \\ 0 & 1 + s_p + p + t \end{pmatrix}. \quad (3)$$

In this matrix, the first row and column correspond to diploid species while the second row and column correspond to higher ploidal species. For example m_{11} is the expected number of diploid species arising from a single diploid species during a time unit and m_{12} is the expected number of higher ploidal species arising from a single diploid species during a time unit. Although the number of diploids increases exclusively through speciation, the number of polyploids grows through a combination of speciation and polyploidy events.

We can find the equilibrium expected proportions of diploid and polyploid species within a genus by calculating the leading eigenvalue of the matrix M. These proportions depend on the net speciation and polyploidy rates. If extinction outpaces speciation and polyploidization across all ploidal levels ($s_d < 0$ and $s_p + p + t < 0$), then the entire genus will ultimately go extinct. When the net evolutionary increase of polyploids exceeds that of diploids ($s_p + p + t \geq s_d \geq 0$), then the ratio of polyploids to diploids diverges to infinity, that is, the proportion of species that are polyploids asymptotes to one. If instead polyploids have an evolutionary disadvantage ($s_d > s_p + p + t \geq 0$), then the equilibrium expected proportions of diploids and polyploids will be

$$\frac{s_d - (s_p + p + t)}{s_d - s_p} \quad \text{and} \quad \frac{p + t}{s_d - s_p}, \quad (4)$$

respectively. Thus, polyploids are predicted to eventually become more common than diploids if $s_p > s_d - 2(p + t)$.

We now simplify the model even further by assuming that all ploidal levels share the same net speciation rate, that is, $s_p = s_d = s$. This reduces the number of parameters to three (s, p, and t). Suppose a genus is born at time $n = 0$ with a single diploid species. Then the expected numbers of diploids and polyploids at time n are given by

$$x_{d_n} = (1 + s)^n \quad \text{and} \quad (5)$$

$$x_{p_n} = [p + t + (1 + s)]^n - (1 + s)^n. \quad (6)$$

Equation (6) can be proven inductively. Assume that $x_{p_0} = 0$ and that $x_{p_{n-1}} = [p + t + (1 + s)]^{n-1} - (1 + s)^{n-1}$. At the next time step, the expected number of higher ploidal species will be $x_{p_0} = x_{d_{n-1}}(p + t) + x_{p_{n-1}}(1 + s + p + t)$.That is, all ploidy events occurring on diploid lineages will necessarily yield higher ploidal species, as will all speciation and ploidy events occurring on higher ploidal lineages. This expression then simplifies to equation (6).

Statistical Methods

Parameter estimation

For the purposes of exploration, we consider only allopolyploidy and not autopolyploidy (that is, we set $t = 0$), which reduces the simple model to two parameters (s and p). We use the model to estimate the net speciation rate and rate of polyploidy based on an observed ploidal distribution within a genus and the approximate age of the genus, n. In particular, we set the observed number of diploid and higher ploidal species to the expected number of such species at time n and then solve for s and p. Equations (5) and (6) imply

$$s = \sqrt[n]{x_{d_n}} - 1 \quad \text{and} \quad (7)$$

$$p = \sqrt[n]{x_{p_n} + (1 + s)^n} - (1 + s). \quad (8)$$

For each of ten different genera, we have both an estimate of the total number of species in the genus (G_g) (Mabberley 1997) and the number of species of each ploidal level ($o_{g,k}$, for $k = 2, 4, 6, \ldots$) within a representative sample from each genus containing N_g species–(Osmond et al. 1980; Graham and Cavalcanti 2001; Goldblatt and Johnson 2003; Missouri Botanical Garden 2005; Hijmans et al. 2006). We estimate the total numbers of species at each ploidal level within a genus g by $y_{k,g} = o_{k,g}/N_g \cdot G_g$, which yields the following approximate numbers of diploids and higher ploidal species in the genus

$$\hat{x}_{d_n} = y_{2,g} \quad \text{and} \quad (9)$$

$$\hat{x}_{p_n} = \sum_{j>1} y_{2j,g}. \quad (10)$$

Given these values and an approximate age of the genus, n, equations (7) and (8) yield estimates for the genus-specific net speciation and polyploidy rates.

Generating hypothetical species distributions

We stochastically simulate the evolution of polyploids under the model described above to calculate both confidence intervals for the evolutionary rate estimates and goodness-of-fit scores for the model itself. In particular, speciation is modeled as a birth-only stochastic process occurring at a specified net speciation rate (speciation rate minus extinction rate). Thus, extinction of the genus cannot occur. Allopolyploidy occurs along each lineage at a specified polyploidization rate for which the resulting ploidal level is the sum of the parent ploidal level and the ploidal level of a randomly chosen second parent from among all extant species.

As just described, equations (7) and (8) allow us to estimate net speciation ($s_{g,a}$) and polyploidy ($p_{g,a}$) rates, respectively, from the observed distribution of ploidal levels within a genus g of presumed age a. For each pair of parameter estimates, we perform 1000 stochastic simulations of the evolutionary model described in equation (1) for a period of a years and record the resulting species distributions. (The model assumes discrete time steps of one per 10^5 years.) We generate a hypothetical sample by randomly choosing a fraction N_g/G_g from the final simulated species distribution, which corresponds to the size of our original sample of ploidal levels within genus g. Let $O^n_{g,a}$ denote the n^{th} simulated sample for genus g of estimated age a. We use each sample $O^n_{g,a}$ to re-estimate the net speciation rate ($\sigma^n_{g,a}$) and polyploidy rate ($\pi^n_{g,a}$) following the method described in the previous section. This procedure yields two distributions of parameter estimates, $\Sigma_{g,n}$ and $\Pi_{g,n}$, for each combination of g and a.

Confidence intervals

The simulated distributions ($O^n_{g,a}$) reflect the variability produced simply by chance, assuming the ratchet model to

be true. Thus we can use these distributions to gauge the precision of our rate estimates. In particular, the 2.5th and 97.5th percentiles of $\Sigma_{g,n}$ and $\Pi_{g,n}$ serve as 95% confidence bounds for the original, empirically based estimates for net speciation ($s_{g,a}$) and polyploidy ($p_{g,a}$) rates.

Confidence intervals for genera with no observed polyploids

If no higher ploidal species have been observed for a genus, then these methods would lead us to conclude incorrectly that the rate of polyploidization is zero with confidence interval containing only the value zero. Bayesian statistics addresses the problem of unseen elements by assigning a small prior probability of each type. This is achieved by adding a small number ϕ, typically ranging between zero and one, to the observed numbers of each of the possible types (Lidstone 1920; Krichevsky and Trofimov 1981; Orlitsky et al. 2003). In our model, one can take a similar approach to estimating speciation and polyploidization rates of genera that lack higher ploidal species. In particular, one would modify the approximate numbers of diploids and higher polyploids given in equations (9) and (10) as follows:

$$\hat{x}_{d_n} = \frac{o_{2,g} + \phi}{N_g} \cdot G_g \quad \text{and}$$

$$\hat{x}_{p_n} = \frac{\left(\sum_{j>1} o_{2j,g}\right) + \phi}{N_g} \cdot G_g \qquad (11)$$

setting $\phi = 1$ (Laplace method) or $\phi = 0.5$ (Jeffreys-Perks Law, Krichevsky-Trofimov Estimator, or Schurmann-Grassberger Law) and then proceed exactly as described above. All of the genera considered in our study have at least one higher ploidal species, thus it was not necessary for us to use these modified estimators. We did, however, use these formulae (with $\phi = 1$) to estimate speciation and polyploidization rates for each simulated genus with no polyploids (that is, for each $O^n_{g,a}$ containing only diploids).

Goodness-of-fit

Given evolutionary rate estimates for a specific genus and genus age ($s_{g,a}$ and $p_{g,a}$), we can use the deterministic model defined by equation (1) to predict the number of species at each ploidal level after a years of evolution. We multiply the predicted numbers of each ploidal class by N_g/G_g to arrive at an expected number of species at each ploidal level in a sample of size N_g. The similarity between the observed empirical distribution ($O_{g,n}$) and this expected distribution ($E_{g,n}$) indicates the compatibility of our model with the original data, and can be quantified using the chi-square statistic,

$$\chi^2_{g,a} = \sum \frac{(O_{g,a} - E_{g,a})^2}{E_{g,a}}.$$

Because the expected frequencies can be quite small (less than five), we cannot simply perform a chi-square goodness-of-fit test. Instead, we generate a null distribution for this statistic using the simulated ploidal distributions described above. For each $O^n_{g,a}$, we use the simulation-specific evolutionary rate estimates $\sigma^n_{g,a}$ and $\pi^n_{g,a}$ in conjunction with equation (1) to attain an expected distribution $E^n_{g,a}$, and then calculate a chi-square value for the hypothetical data,

$$\chi^2_{n,g,a} = \sum \frac{(O^n_{g,a} - E^n_{g,a})^2}{E^n_{g,a}}.$$

For each combination of g and a, these 1000 values provide a null distribution for assessing the fit between the empirical data and the model. The P-value is then the proportion of this distribution that is greater than or equal to the test statistic $\chi^2_{g,a}$.

Table 1. Estimates for evolutionary rate parameters (and standard errors) for flowering plants (Levin and Wilson 1976).

	s Net speciation rate in lineage per my	p Mean increase in ploidal diversity per my
Herbs	1.05 ± 1.19	0.05 ± 0.09
Shrubs	0.24 ± 0.45	0.01 ± 0.03
Trees	0.09 ± 0.03	0.001 ± 0.001

Results

Although our model includes both allopolyploidy and autopolyploidy, we have studied the scenario in which polyploidization occurs exclusively by allopolyploidy. That is, we have set the rate of autopolyploidization (t) equal to zero.

Speciation, ploidization, and extinction rates appear to vary considerably across the plant kingdom. Levin and Wilson (1976) estimated these evolutionary parameters for angiosperms (Table 1). Assuming that species with different levels of ploidy do not differ in their average speciation and extinction rates, the model predicts that the proportion of polyploids in the genus should asymptote to one, with the numbers of polyploids increasing at a higher rate than the numbers of diploids (Fig. 2). If polyploids have a net evolutionary advantage expressed as higher speciation rates ($s_p + p > s_d \geq 0$), then the proportion of polyploids increases even more quickly. If, instead, polyploids have a net evolutionary disadvantage ($s_d > s_p + p \geq 0$), then this proportion approaches a value less than one.

Let us consider some examples based on the speciation rates presented in Table 1. If the net speciation rate for polyploids is 10% lower than that of diploids (assuming that the net speciation rates for diploids are the values given in Table 1), then equation (4) predicts that herb and shrub polyploids eventually will be almost as numerous as their diploid counterparts, reaching 47.6% and 41.5% of all species, respectively. In contrast, hardwood polyploids with a 10% speciation disadvantage will only climb to 11.5% of species in the genus. For an herb lineage, polyploids are predicted to become as common as diploids if they have a speciation disadvantage of 9.52%. The same is true for shrub and woody lineages when the polyploid speciation disadvantage is 8.33% and 2.22%, respectively. The incidence of polyploidy may be substantial even if they have a 25% disadvantage to their diploid relatives. For example, herb polyploids with such a disadvantage will eventually make up 19.4% of species in the genus (Fig. 2). Thus, the presence of polyploids in a genus says nothing about their evolutionary success relative to diploids.

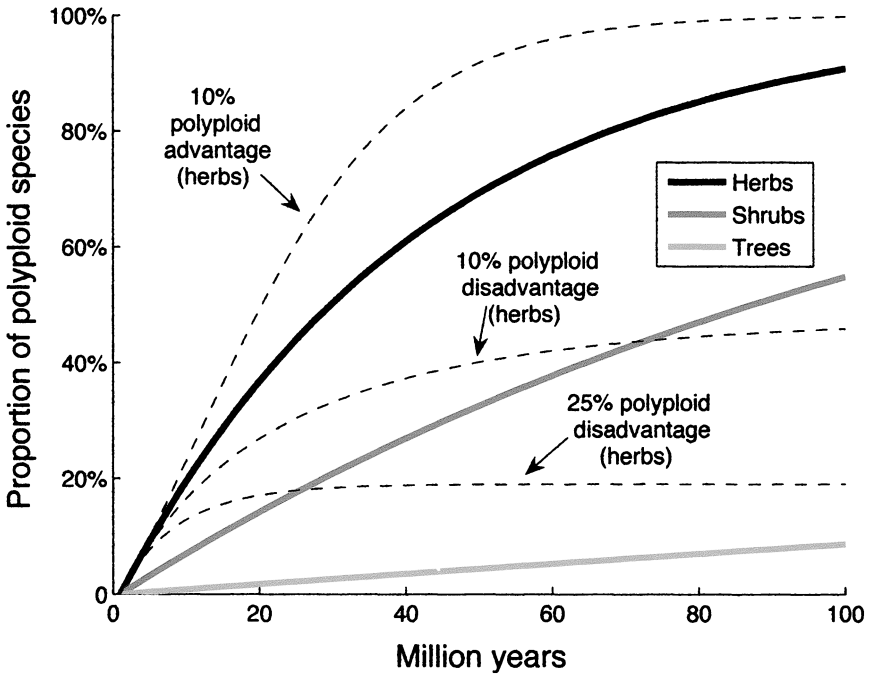

FIG. 2. The ascendance of polyploids in flowering plants. Assuming the parameter values given in Table 1, the model given by equation (1) predicts that the proportion of polyploids should increase to one when polyploids have an evolutionary growth rate greater than that of diploids, $s_p + p > s_d$. This assumes that all higher ploidal levels share the same net speciation rate, s_p. A 10% polyploid advantage or disadvantage assumes that polyploid net speciation rates are 110% or 90% that of diploids ($s_p = 1.1 \cdot s_d$ or $s_p = 0.9 \cdot s_d$), respectively. Similarly, a 25% disadvantage means $s_p = 0.75 \cdot s_d$.

To assess the validity of the model, we consider observed ploidal distributions from 10 genera—*Artemisia* (Missouri Botanical Garden 2005), *Atriplex* (Osmond et al. 1980; Goldblatt and Johnson 2003), *Cuphea* (Graham and Cavalcanti 2001), *Draba* (Goldblatt and Johnson 2003; Missouri Botanical Garden 2005), *Festuca* (Goldblatt and Johnson 2003; Missouri Botanical Garden 2005), *Galium* (Goldblatt and Johnson 2003; Missouri Botanical Garden 2005), *Poa* (Missouri Botanical Garden 2005), *Salix* (Goldblatt and Johnson 2003; Missouri Botanical Garden 2005), *Silene* (Goldblatt and Johnson 2003; Missouri Botanical Garden 2005), and *Solanum* (Hijmans et al. 2006) (Table 2). (*Salix* is a tree genus and the remaining nine are herb genera.) Given the large error bars on the values in Table 1, it is unlikely that the rates of speciation and ploidization for any particular genus will be equal to the average rates for herbs, shrubs, or woody plants. Therefore we use the observed distributions to estimate genus-specific evolutionary rates. Assuming that all ploidal levels within a genus have the same net speciation rate, Equations (7) and (8) predict this rate and the frequency of polyploid events, respectively, for each genus.

Angiosperm genera are thought to range in age from 2.5 to 110 million years (my; Levin and Wilson 1976). Assuming a genus age of 25 my, the net speciation rate estimates range from 0.181 to 0.348 new species per lineage per my, and the polyploidy rate estimates range from 0.004 to 0.066 ploidal events per lineage per my. If we increase the presumed genus age to 50 my, these rate estimates drop by approximately 55%. These values lie well within the error bars for the previous estimates given in Table 1.

Assuming these genus-specific estimates for net speciation and polyploidy rates, we apply the deterministic ratchet model given by equation (1) to calculate the expected current distribution of ploidal levels within each genus and then ask whether the predictions resemble the observed data (Table 2). To estimate the speciation and polyploidization rates, recall that we set the expected number of diploids and expected total number of polyploids equal to the observed numbers of diploids and polyploids, respectively, and then use equations (7) and (8) to solve for s and p. Thus, when we input these values into the model and calculate the expected numbers of diploids and polyploids, we predict exactly the observed values. However, the specific distribution of species among the higher ploidal levels is not completely constrained by our parameter estimation method. Note that the expected distributions are quite insensitive to the age of the genus.

Although the predicted and observed numbers of higher ploidy species are by no means identical, the model produces qualitatively similar distributions to those observed. Figure 3 illustrates the models predictions for a genus evolving for 200 million years, starting from a single diploid species. This assumes the evolutionary rates estimated from the observed distribution of New Zealand members of the genus *Atriplex* (Osmond et al. 1980) and a presumed genus age of 25 my (Table 2). The four bold lines correspond to the expected proportions of diploids, tetraploids, hexaploids, and octoploids in the genus. The ratchet works its way up ploidal levels. That is, all polyploids progressively gain on diploids, with lower polyploids initially growing more quickly than higher polyploids. After 150 million years of evolution (far into the future for *Atriplex*), the ploidal hierarchy is predicted to reverse, with octoploids becoming more numerous than

TABLE 2. Genus-specific estimates, observations, and predictions. The genus-specific polyploidy (p) and net speciation (s) rates are estimated using the observed numbers of diploids ($2n$) and the sum of the observed numbers of higher ploidy species ($4n + 6n + \cdots + 16n$). We use the model given by equation (1) to predict species numbers at each ploidal level. As described in the text, 95% confidence intervals for evolutionary rates and P-values for model goodness-of-fit are based on comparisons to data generated by stochastic evolutionary simulation. "Obs" (in bold) and "Exp" denote observed and predicted species counts, respectively. P-values in bold denote significant deviations from the model. Total species in each genus are from (Mabberley 1997) and distributions of ploidal levels within samples from each genus are from (Osmond et al. 1980; Graham and Cavalcanti 2001; Goldblatt and Johnson 2003; Missouri Botanical Garden 2005; Hijmans et al. 2006).

			Species distribution within sample										
Genus	Data	Total	$2n$	$4n$	$6n$	$8n$	$10n$	$12n$	$14n$	$16n$	P-value	Estimate for s	Estimate for p
Artemisia	Obs	**350**	**72**	**37**	**12**	**1**	—	—	—	—			
	Exp (25 my)	350	72	29.6	12.1	4.9	2.0	0.8	0.3	0.1	0.387	0.248 (0.120, 0.307)	0.028 (0.004, 0.079)
	Exp (50 my)	350	72	29.5	12.1	4.9	2.0	0.8	0.3	0.1	0.392	0.115 (0.054, 0.141)	0.012 (0.002, 0.033)
Atriplex	Obs	**300**	**39**	**17**	**5**	**1**	—	—	—	—			
	Exp (25 my)	300	39	14.5	5.4	2.0	0.7	0.3	0.1	0.0	0.586	0.244 (0.116, 0.303)	0.024 (0.002, 0.072)
	Exp (50 my)	300	39	14.5	5.4	2.0	0.7	0.3	0.1	0.0	0.574	0.112 (0.058, 0.139)	0.011 (0.001, 0.029)
Cuphea	Obs	**260**	**65**	**60**	**17**	**11**	—	—	—	—			
	Exp (25 my)	260	65	37.5	21.6	12.4	7.1	4.1	2.3	1.3	0.258	0.217 (0.084, 0.268)	0.044 (0.008, 0.094)
	Exp (50 my)	260	65	37.4	21.5	12.4	7.1	4.1	2.3	1.3	0.277	0.101 (0.043, 0.126)	0.019 (0.004, 0.046)
Draba	Obs	**300**	**32**	**10**	**5**	**9**	**3**	—	—	**1**			
	Exp (25 my)	300	32	15.0	7.0	3.2	1.5	0.7	0.3	0.2	0.165	0.236 (0.112, 0.294)	0.033 (0.004, 0.079)
	Exp (50 my)	300	32	15.0	7.0	3.2	1.5	0.7	0.3	0.2	0.165	0.109 (0.054, 0.133)	0.014 (0.001, 0.039)
Festuca	Obs	**450**	**52**	**45**	**43**	**10**	**2**	—	—	—			
	Exp (25 my)	450	52	34.4	22.6	14.8	9.7	6.4	4.2	2.8	0.196	0.234 (0.098, 0.289)	0.056 (0.012, 0.113)
	Exp (50 my)	450	52	34.3	22.6	14.8	9.7	6.4	4.2	2.8	0.203	0.108 (0.045, 0.134)	0.025 (0.006, 0.057)
Galium	Obs	**300**	**71**	**36**	**4**	**4**	**2**	—	—	—			
	Exp (25 my)	300	71	28.0	11.0	4.3	1.7	0.7	0.3	0.1	0.389	0.242 (0.110, 0.300)	0.026 (0.003, 0.071)
	Exp (50 my)	300	71	27.9	11.0	4.3	1.7	0.7	0.3	0.1	0.349	0.112 (0.051, 0.138)	0.011 (0.002, 0.027)
Poa	Obs	**200**	**22**	**26**	**20**	**11**	**2**	—	—	—			
	Exp (25 my)	200	22	16.1	11.7	8.5	6.2	4.5	3.3	2.4	0.311	0.181 (0.066, 0.235)	0.066 (0.013, 0.127)
	Exp (50 my)	200	22	16.1	11.7	8.5	6.2	4.5	3.3	2.4	0.288	0.085 (0.033, 0.109)	0.029 (0.006, 0.058)
Salix	Obs	**400**	**40**	**12**	**5**	**1**	—	—	—	—			
	Exp (25 my)	400	40	12.4	3.8	1.2	0.4	0.1	0.0	0.0	0.651	0.264 (0.131, 0.324)	0.020 (0.001, 0.063)
	Exp (50 my)	400	40	12.4	3.9	1.2	0.4	0.1	0.0	0.0	0.687	0.122 (0.065, 0.148)	0.009 (0.001, 0.024)
Silene	Obs	**700**	**138**	**7**	**2**	**1**	—	—	—	—			
	Exp (25 my)	700	138	9.3	0.6	0.04	0.0	0.0	0.0	0.0	**0.026**	0.310 (0.169, 0.368)	0.004 (0, 0.012)
	Exp (50 my)	700	138	9.3	0.6	0.04	0.0	0.0	0.0	0.0	**0.031**	0.141 (0.083, 0.166)	0.002 (0, 0.006)
Solanum	Obs	**1700**	**142**	**30**	**15**	—	—	—	—	—			
	Exp (25 my)	1700	142	34.2	8.2	2	0.5	0.1	0.0	0.0	0.246	0.348 (0.200, 0.407)	0.016 (0.002, 0.045)
	Exp (50 my)	1700	142	34.2	8.2	2	0.5	0.1	0.0	0.0	0.219	0.157 (0.094, 0.183)	0.007 (0.001, 0.017)

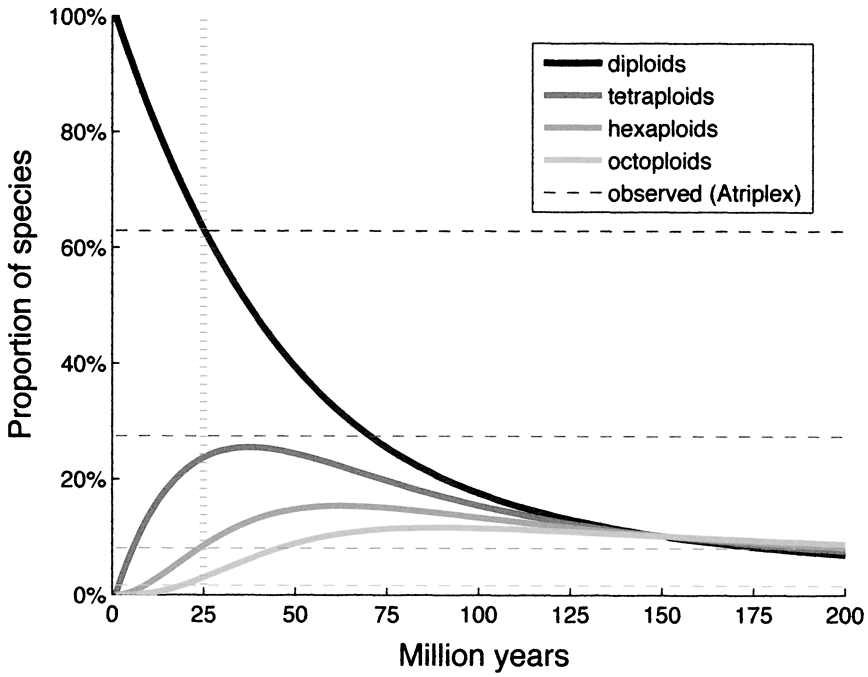

FIG. 3. The evolution of polyploidy in *Atriplex*. Using the net speciation and polyploidy rates estimated for *Atriplex* (Table 2), the model predicts that proportion of species that are diploid should decline as the proportions of tetraploids, hexaploids, and octoploids increase (assuming that all ploidal levels have the same net speciation rates). The vertical dotted line at 25 million years indicates the age of the genus that we assumed in estimating the net speciation and polyploidy rates. The dashed horizontal lines indicate observed proportions within a sample from the *Atriplex* genus (Osmond et al. 1980). From top to bottom, these lines show the observed proportions of diploids, tetraploids, hexaploids, and octoploids. The horizontal lines are close to the theoretical predictions at approximately 25 million years.

hexaploids, hexaploids becoming more numerous than tetraploids, and tetraploids becoming more numerous than diploids. The horizontal dashed lines in Figure 3 indicate the observed proportions of *Atriplex* diploids, tetraploids, hexaploids, and octoploids (from top to bottom), from which the evolutionary parameters were estimated. The model predicts that the genus will exhibit roughly the observed distribution after 25 million years (vertical dotted line), assuming all species evolved from a single diploid colonizer of New Zealand.

We quantitatively assess the fit of the model by calculating the similarity between the predicted and observed values, and then comparing this value to comparable values measured on simulated data. With the exception of *Silene*, the predicted and observed distributions based on the empirical data matched better than the corresponding distributions for at least 15% of simulated data. This suggests that for almost all of the genera, the discrepancies between the model predictions and the data can be attributed to chance rather than a fundamental incompatibility with the model. We therefore have reasonable support for the ratchet model across nine of the ten genera, and suggest that it may offer a parsimonious explanation for the abundance of higher ploidal levels within flowering plants.

DISCUSSION

We have introduced a model that shows that the proportion of polyploids in a genus increases over time. This proportion may reach a substantial level even if polyploids are at a disadvantage, even if the disadvantage is sizeable. Although the latter may seem counterintuitive, it is a direct consequence of the irreversibility of ploidal level increases. Accordingly, efforts to explain the overall success of polyploids *only* on the basis of some set of contemporary attributes may be somewhat misplaced.

Although our analysis demonstrates that one need not invoke an advantage to polyploids for their relative numbers to increase in time, it does not mean that such an advantage does not exist overall, or in some genera, or in some phylads within genera. This model can be used to compare the simple ratchet hypothesis to hypotheses that assume differences in net speciation and polyploidy rates among different ploidal levels. One can also address the hypothesis that polyploids have a speciation advantage by looking for correlations between species richness and the incidence of polyploids within genera. Otto and Whitton (2000) reported that the average species richness of 200 dicot genera increased modestly as the degree of polyploidy within genera increased. A survey of 50 genera from the flora of the Pyrenees showed a somewhat similar relationship (Petit and Thompson 1999). As our model suggests, however, the total number of species should grow in tandem with both the proportion and level of polyploids within a genus, even if polyploids have a net speciation disadvantage. By the simple ratchet, older genera will have both greater species richness and larger proportions of higher polyploids than younger genera. Therefore, as Otto and Witton (2000) also noted, correlations between species richness and ploidal levels do not necessarily imply that polyploids

enjoy greater rates of speciation. Ideally one would compare species richness in pairs of sister clades in which a polyploidy event occurred early in the history of only one of the two clades. If the clades with higher ancestral ploidal level tend to have greater species richness, then one may infer a higher rate of lineage splitting among polyploids. A recent phylogenetic study of Rosaceae found that clades with more polyploids tended to be more speciose, but found no evidence that polyploids diversified faster than their diploid counterparts (in our model, $s_p = s_d$) (Vamosi and Dickinson 2006). The authors concluded that the rate of polyploidization itself (in our model, p), could account for the higher diversity.

Until now, we have discussed the advantages and disadvantages of polyploidy in terms of the net speciation rate of polyploids after they have been established. We now turn to the issue of the initial increase in ploidal level. There is clearly variation in the ability to produce polyploids in the first place. The rate of chromosome number increase in herbs is about twenty times greater than that in hardwood trees (Levin and Wilson 1976). Stebbins (1938) speculated that cell enlargement produced by ploidal increase might prevent the proper formation of wood fibers by the vascular cambium, and thus retard the evolution of woody polyploids. This may explain why gymnosperms have lower levels of polyploidy than angiosperms, although they are older (Khoshoo 1959; Delevoryas 1980; Murray 1998). In this study, we estimated the rates of polyploid events for 10 genera and found some variation, with the single woody plant genus (*Salix*) on the lower end of the range estimated for the nine herb genera (Table 2). Importantly, even if a genus only very rarely produces higher ploidal levels, once a polyploid lineage is established, there is no turning back. If polyploids have a net speciation rate roughly similar to that of diploids, then the proportions of polyploids are expected to increase indefinitely, although perhaps slowly (Fig. 1).

Given that ploidal increases occur by a ratchet process, we surmise that the early angiosperms had lower ploidal levels than contemporary species. This idea is supported by the work of Masterson (1994) on the Lauraceae, Magnoliaceae, and Platanaceae, which showed that ploidal levels were lower millions of years ago than they are now.

As species with lower ploidal levels die out, the base number of genera will increase. The high base numbers of whole families of woody dicots (e.g., Magnoliacae, $x = 19$; Salicaceae, $x = 19$; Platanaceae, $x = 21$) apparently represent ancient polyploid conditions (Stebbins 1971, 1980). This condition certainly seems to be the case in ferns and other pteridophytes, which predate the angiosperms by hundreds of millions of years (Manton 1950; Wagner and Wagner 1980). Consider the base of numbers of *Polypodium* ($x = 37$), *Psilotum* ($x = 50$), *Tmesipterus* ($x = 104$), and *Ophioglossum* ($x = 120$). These and other such lineages must represent the survivors of ancient polyploid complexes. Many pteridophytes contain polyploids built on very high base numbers. For example *Ophioglossum reticulatum* has a haploid number of 630 (Stebbins 1971). A very high proportion of ferns and fern allies appear to be ancient polyploids (Grant 1981).

If the premise that young phylads should have low levels of polyploidy is correct, it follows that the genera comprising oceanic island floras of recent vintage also would have low levels of chromosomal evolution. This indeed is the case. In situ polyploidy in Hawaiii is very limited, with clear examples in *Peperomia*, *Portulaca*, and *Wikstromia*, and less certain examples in a few other genera (Carr 1998). The same trends hold for other islands as well (Stuessy and Crawford 1998).

Ploidal level is not the only type of chromosomal change that is mostly irreversible. Translocations and inversions also follow a ratchet process. Thus, the longer the time since lineages have diverged from a common ancestor, the larger the number of chromosomal rearrangements by which they will differ. Also, the longer the time to common ancestry the more rearrangements will have become established within a genus. Some of the rearrangements will be small, and evident only from dense genetic maps (Levin 2002). Others will be large, and establish partial sterility barriers between populations, in some cases being associated with species formation.

Polyploidy is but just one type of irreversible (or rarely reversible) evolution. Bull and Charnov (1985) discuss other types of irreversible evolution including all-female parthenogenesis, heteromorphic sex chromosomes, and haplo-diploidy (arrhenotoky) in arthropods. These characteristics also appear independently in many lineages. The inability of a population to revert to a recent ancestral state is sometimes referred to as Dollo's law (Dobzhansky 1970; Gould 1970). As noted by Bull and Charnov (1985), awareness of state irreversibility is paramount when using comparative methods to test evolutionary hypotheses including the construction of phylogenetic trees.

Acknowledgments

We thank M. Kirkpatrick, M. Lachmann, S. Otto, C. Wilke, and an anonymous reviewer for insightful comments on earlier versions of this manuscript and statistical advice.

Literature Cited

Adams, K. L., and J. F. Wendel. 2005. Polyploidy and genome evolution in plants. Curr. Opin. Plant Biol. 8:135–141.

Baldwin, B. G., and M. J. Sanderson. 1998. Age and rate of diversification of the Hawaiian silversword alliance (Compositae). Proc. Natl. Acad. Sci. USA. 95:9402–9406.

Bingham, E. T. 1980. Maximizing heterozygosity in polyploids. Pp. 471–490 *in* W. H. Lewis, ed. Polyploidy—biological relevance. Plenum, New York.

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

Carr, G. D. 1998. Chromosome speciation in Hawaiian flowering plants. Pp. 5–47 *in* T. F. Stuessy and M. Ono, eds. Evolution and speciation of island plants. Cambridge Univ. Press, New York.

Delevoryas, T. 1980. Polyploidy in gymnosperms. Pp. 215–218 *in* W. H. Lewis, ed. Polyploidy—biological relevance. Plenum, New York.

De Wet, J. M. J. 1968. Diploid-tetraploid-haploid cycles and the origin of of variability in *Dichanthium* agamospecies. Evolution 22:394–397.

Dobzhansky, T. 1970. The genetics of the evolutionary process. Columbia Univ. Press, New York.

Goldblatt, P., and D. E. Johnson. 2003. Index to plant chromosome numbers. 1998–2000. Monogr. Syst. Bot. Missouri Bot. Gard. No. 94.

Gould, S. J. 1970. Dollo on Dollo's law: irreversibility and the status of evolutionary laws. J. Hist. Biol. 3:189–212.
Graham, S., and T. B. Cavalcanti. 2001. New chromosome counts in the Lythraceae and a review of chromosome numbers in the family. Syst. Bot. 26:445–458.
Grant, V. 1981. Plant speciation. 2d ed. Columbia Univ. Press, New York.
Hijmans, R. J., T. Gavrilenko, S. Stephenson, J. Bamberg, A. Salas, and D. M. Spooner. 2006. Polyploidy in wild and cultivated potatoes (*Solanum* Section *Petota*). Am. J. Bot: *In press*.
Jiang, C., R. J. Wright, K. J. El-Zik, and A. H. Patterson. 1998. Polyploid formation created unique avenues for response to selection in *Gossypium*. Proc. Natl. Acad. Sci. USA 95:4419–4424.
Khoshoo, T. N. 1959. Polyploidy in gymnosperms. Evolution 13: 24–39.
Krichevsky, R., and V. Trofimov. 1981. The performance of universal encoding. IEEE Trans. Inform. Theory 27:199–207.
Levin, D. A. 2000. The origin, demise, and expansion of plant species. Oxford Univ. Press, New York.
———. 2002. The role of chromosomal change in plant evolution. Oxford Univ. Press, New York.
Levin, D. A., and A. C. Wilson. 1976. Rates of evolution in seed plants: Net increase in diversity of chromosome numbers and species numbers through time. Proc. Natl. Acad. Sci. USA 73: 2086–2090.
Levy, A. A., and M. Feldman. 2004. Genetic and epigenetic reprogramming of the wheat genome upon allopolyploidization. Biol. J. Linn. Soc. 82:607–613.
Lidstone, G. J. 1920. Note on the general case of the Bayes-Laplace formula for inductive or a posteriori probabilities. Trans. Fac. Act. 8:182–192.
Lundquist, A. 1966. Heterosis and inbreeding depression in autotetraploid rye. Hereditas 56:317–366.
Mabberley, D. J. 1997. The plant book. 2nd ed. Cambridge, Univ. Press, Cambridge, U.K.
Manton, I. 1950. Problems of cytology and evolution in the Pteridophyta. Cambridge Univ. Press, Cambridge, U.K.
Masterson, J. 1994. Stomatal size in fossil plants: evidence for polyploidy in majority of flowering plants. Science 264: 421–424.
Missouri Botanical Garden. 2005. w3TROPICOS index to plant chromosome number data base. Available from http:/mobot.mobot.org/W3T/search/ipcn.html.
Murray, B. G. 1998. Nuclear DNA amounts in gymnosperms. Ann. Bot. 82(Suppl. A):3–15.
Nee, S. 2001. Inferring speciation rates from phylogenies. Evolution 55:661–668.
Otto, S. P., and J. Whitton. 2000. Polyploid incidence and evolution. Annu. Rev. Genet. 34:401–437.
Orlitsky, A., N. P. Santhanam, J. Zhang. 2003. Always Good Turing: Asymptotically optimal probability estimation. Science 302: 427–431.
Osmond, C. B., O. Björkman, and D. J. Anderson. 1980. Physiological processes in plant ecology: towards a synthesis with *Atriplex*. Springer-Verlag, Berlin.
Petit, C., and J. D. Thompson. 1999. Species diversity and ecological range in relation to ploidal level in the flora of the Pyrenees. Evol. Ecol. 13:45–66.
Pires, J. C., J. Zhao, M. E. Schranz, E. J. Leon, P. A. Quijada, L. N. Lukens, and T. C. Osborn. 2004. Flowering time divergence and genomic rearrangements in resynthesized *Brassica* polyploids (Brassicaceae). Biol. J. Linn. Soc. 82:675–688.
Ramsey, J., and D. W. Schemske. 1998. Pathways, mechanisms, and rates of polyploid formation in flowering plants. Annu. Rev. Ecol. Syst. 29:467–501.
Ramsey, J., and D. W. Schemske. 2002. Neopolyploidy in flowering plants. Annu. Rev. Ecol. Syst. 33:589–639.
Raven, P. H., and H. J. Thompson. 1964. Haploidy and angiosperm evolution. Am. Nat. 98:251–252.
Sanderson, M. J., and M. J. Donohue. 1994. Shifts in the diversification rate with the origin of angiosperms. Science. 264: 1590–1593.
Schranz, M. E., and T. C. Osborn. 2004. De novo variation in life-history traits and responses to growth conditions of resynthesized polyploid *Brassica napus* (Brassicaceae). Am. J. Bot. 91: 174–183.
Soltis, D. E., and P. S. Soltis. 1993. Molecular data and the dynamic nature of polyploidy. Crit. Rev. Plant Sci. 12:243–273.
Stebbins, G. L. 1938. Cytological characteristics associated with different growth habits in dicotyledons. Am. J. Bot. 25:189–198.
———. 1971. Chromosomal evolution in higher plants. Addison-Wesley, London.
———. 1980. Polyploidy: future prospects. Pp. 495–520 *in* W. H. Lewis, ed. Polyploid—biological relevance. Plenum, New York.
———. 1985. Polyploidy, hybridization, and the invasion of new habitats. Ann. Missouri Bot. Gard. 72:824–832.
Stuessy, T. F., and D. J. Crawford. 1998. Chromosomal stasis during speciation in angiosperms of oceanic islands. Pp. 307–324 *in* T. F. Stuessy and M. Ono, eds. Evolution and speciation of islands plants. Cambridge Univ. Press, Cambridge, U.K.
Thompson, J. D., and R. Lumaret. 1992. The evolutionary dynamics of polyploid plants: origins, establishment and persistence. Trends Ecol. Evol. 7:302–307.
Tomekpe, K., and R. Lumaret. 1991. Association between quantitative traits and allozyme heterozygosity in a tetrasomic species, *Dactylis glomerata*. Evolution 45:359–370.
Vamosi, J. C., and T. A. Dickinson. 2006. Polyploidy and diversification: a phylogenetic investigation in Rosaceae. Int. J. Plant Sci. 167:349–358.
Wagner, W. H., Jr.,and F. S. Wagner. 1980. Polyploidy in pteridophytes. Pp. 199–214 *in* W. H. Lewis, ed. Polyploid—biological relevance. Plenum, New York.

Corresponding Editor: M. Geber