Ploidy frequencies in plants with ploidy heterogeneity: fitting a general gametic model to empirical population data

Jan Suda¹,² and Tomáš Herben¹,²

¹Faculty of Science, Department of Botany, Charles University in Prague, Benešová 2, 128 01 Prague 2, Czech Republic
²Academy of Sciences of the Czech Republic, Institute of Botany, 252 43 Pražské Příhonice, Czech Republic

Genome duplication (polyploidy) is a recurrent evolutionary process in plants, often conferring instant reproductive isolation and thus potentially leading to speciation. Outcome of the process is often seen in the field as different cytotypes co-occur in many plant populations. Failure of meiotic reduction during gametogenesis is widely acknowledged to be the main mode of polyploid formation. To get insight into its role in the dynamics of polyploidy generation under natural conditions, and coexistence of several ploidy levels, we developed a general gametic model for diploid–polyploid systems. This model predicts equilibrium ploidy frequencies as functions of several parameters, namely the unreduced gamete proportions and fertilities of higher ploidy plants. We used data on field ploidy frequencies for 39 presumably autopolyploid plant species/populations to infer numerical values of the model parameters (either analytically or using an optimization procedure). With the exception of a few species, the model fit was very high. The estimated proportions of unreduced gametes (median of 0.0089) matched published estimates well. Our results imply that conditions for cytotype coexistence in natural populations are likely to be less restrictive than previously assumed. In addition, rather simple models show sufficiently rich behaviour to explain the prevalence of polyploids among flowering plants.

1. Introduction

Polyploidy (i.e. the multiplication of entire chromosome sets above the diploid level) has played a key role in evolution and diversification of many eukaryotic organisms, including yeasts, insects, vertebrates and plants [1]. Genome multiplication is especially common in monilophytes (horsetails and ferns) and angiosperms, with estimated proportions of species with one or more polyploidization events in their evolutionary history of up to 95 per cent and 70 per cent, respectively [2]. Such events can be relatively recent (i.e. neopolyploidization), but genomic data have shown that virtually all angiosperms underwent one or more rounds of ancient whole genome duplication [3]. Because polyploidization, unlike other speciation processes, may confer instant reproductive isolation, it is considered a major mode of sympatric speciation, at least in the plant kingdom [4]. Quantitative estimates of the contribution of polyploidy to speciation have only recently begun to appear; it has been estimated that 15 per cent of speciation events in angiosperms and 31 per cent in ferns are associated with ploidy increases [5]. Considering the ubiquity and significance of polyploidy, understanding this mode of speciation basically means understanding the single most common evolutionary transition in the evolutionary history of angiosperms.

Whereas polyploid derivate may originate through somatic chromosome doubling, unreduced gamete formation is undoubtedly the major driving force in the formation of polyploids [6]. The origin via the production of unreduced gametes has been originally applied to autopolyploids, but it may also be valid for allopolyploids that have relatively frequently been hypothesized...
to originate as diploid homoploid hybrids that polyploidize afterwards. The frequency of recent autopolyploids in natural conditions is co-determined by the rates of their formation, establishment success, longer-term persistence, reproductive success and further spread. In addition to studies that have addressed one or more of these factors experimentally [7], the last few decades have seen several theoretical models aiming to evaluate the fates of new polyploids within the populations of their diploid/low-ploid ancestors [8–15]. Most models are based on the assumption that the newly formed polyploid must cope with the minority disadvantage (i.e. the frequency-dependent process in which the rare ploidy is evolutionary disadvantaged unless it develops breeding barriers or shows higher fitness or fertility [16]). However, some data indicate that odd-numbered interploidy hybrids (such as triploids) may actually enhance the rate of polyploid formation and facilitate rather than diminish their fixation [10]. The outcome of these two opposing, but not mutually exclusive, processes (i.e. triploid block versus triploid bridge) is largely affected by the reproductive and fitness characteristics of triploid crosses. In general, the fate of a newly arisen tetraploid in a diploid population depends on the rate of triploid formation (or, conversely, the strength of triploid block), triploid fitness (viability and fertility) relative to diploids and tetraploids, and the ploidy of functional gametes of triploids. Simulations have shown that even partially fit triploids can increase the likelihood of diploid–tetraploid coexistence and, in some cases, facilitate tetraploid fixation [10].

Despite the development of more complex models that take into account some key parameters that may affect the formation and establishment of polyploids, including ploidy fitness, ploidy fertility and the rate of unreduced gamete production, their outcomes were not checked against actual ploidy frequencies observed under natural conditions. Consequently, the relative role of triploid block versus triploid bridge cannot be assessed for existing field systems. Analysis of actual ploidy frequencies was for a long time impractical because the number of cytotype samples in conventional karyological studies was limited by methodological constraints that precluded large-scale population screening. Broad insights into ploidy variation in situ only became possible with the advent of flow cytometry, which transformed plant karyology from the examination of single individuals to the analysis of whole populations across multiple species and different spatial scales [17]. Ploidy data for several thousand individuals are now commonplace in cytogeographic studies, which provide an ideal framework for relating in silico results of ploidy dynamics to empirical population data.

In this paper, we develop a simple and general gametic model of ploidy-level frequencies, on the basis of the models of Husband [10] and Yamauchi et al. [12]. Such models predict equilibrium frequencies of individual ploidy levels using a few basic parameters (probabilities of formation of unreduced gametes and fertilities of higher ploidy levels). In our approach, we have developed different versions of the model for systems comprising different ranges of ploidy levels. We first present an analysis of the model’s behaviour for systems comprising individuals ranging from: (i) 2x to 3x, (ii) 2x to 4x, and (iii) 2x to 6x, exploring the model’s general predictions in light of current knowledge of the dynamics of polyploid series and other models. We then fit the model to observed ploidy frequency data to:

(i) find out whether the empirical frequencies are compatible with such a simple gametic generating process, and (ii) obtain estimates of these key parameters of the model and to compare their numerical values with independent estimates available from other sources.

Although there are two basic modes of polyploidy establishment (i.e. auto- and allopolyploidy), for the sake of simplicity, we deal only with autopolyploidy (also called intraspecific or intrapopulation polyploidy) to avoid problems related to the involvement of different parental species in the polyploidy’s origin. Recent data show that autopolyploids are more common than previously considered and have important evolutionary and ecological roles in natural populations [18].

2. Material and methods

(a) A gametic model of polyploidy: general framework

Let us assume that the ploidy level of an individual fully determines its gametic structure. We further assume that diploids produce mostly reduced gametes and a certain proportion of unreduced gametes (b), triploids produce pairs of reduced (s and 2s) gametes, in addition to a certain proportion of unreduced gametes (c). We assume that aneuploid gametes either do not form or that aneuploid progeny does not survive.

We further assume that populations of individual cytotypes are fully homogeneous (i.e. gametic structure of all individuals of a given cytotype is identical) and that gametes combine freely in individuals of the same ploidy level (irrespective of the actual mode of reproduction, whether outcrossing or selfing). Thus, we do not assume any a priori minority cytotype disadvantage. The progeny of each individual are thus fully determined by the probabilities of individual gametic types (see [12]). (We refer to this later as the ‘gametic process’.) For example, a diploid individual that produces haploid gametes with probability 1-b and unreduced gametes with probability b, will have diploid progeny if two reduced gametes combine (i.e. with the probability \((1 - b)^2\)), triploid with the probability \(2b(1 - b)\), and tetraploid with the probability \(b^2\). Using the same logic, the progeny of a triploid plant with c proportion of unreduced gametes will range from diploid (if two reduced gametes with the haploid chromosome number merge, with probability \((1 - c)^2/4\)) to hexaploid (if two fully unreduced gametes merge, with probability \(c^2\); see also electronic supplementary material, table S2 and figure S2).

We further assume that individuals of all ploidy levels reproduce only generatively, have identical lifespans and survival, and have non-overlapping generations. For the sake of simplicity, we also disregard the fact that hybridization between different ploidy levels can take place. In addition, we assume that the system is closed and that the maximum viable ploidy level is 6x (plant groups with ploidy variation ranging from diploid to above hexaploid are quite rare under natural conditions). Changes in numbers of individuals of ploidy i from time t to time \(t + 1\) can then be approximated by using a set of recurrent equations:

\[
P_i(t+1) = sn[i\mu_2F_2P_2(t) + \mu_3F_3P_3(t) + \cdots],
\]

where \(P_i(t)\) is the abundance of plants of ploidy i (the subscript thus equalling the particular multiple of the basic chromosomal complement, and in the general case ranging from 2 to 6) at the time t, \(F_j\) is the proportion of progeny with ploidy i of mother individuals of ploidy j as determined by the gametic process, \(\mu_i\) is the relative fertility of individuals of ith ploidy; and s and n are, respectively, survival to maturity of and number of seeds produced by a diploid plant. For convenience, we assume, without loss of generality, \(\mu_2 = 1\).
Table 1. Models used for fitting of different subsets of the data. (c, estimated proportion of unreduced gametes of the triploid; \( \mu \), relative fertility of higher ploidy levels.)

<table>
<thead>
<tr>
<th>model</th>
<th>constraints against the model II</th>
<th>no. free parameters</th>
<th>fit to</th>
<th>fitting</th>
</tr>
</thead>
<tbody>
<tr>
<td>minimal model</td>
<td>( c = 0, \mu = 0 )</td>
<td>1</td>
<td>( 2x ) – ( 3x ) species</td>
<td>analytic solution</td>
</tr>
<tr>
<td>model I</td>
<td>( c = 0 )</td>
<td>2</td>
<td>( 2x ) – ( 3x ) species, ( 2x ) – ( 4x ) species</td>
<td>analytic solution</td>
</tr>
<tr>
<td>model II</td>
<td></td>
<td>3</td>
<td>( 2x ) – ( 4x ) species, ( 2x ) – ( 6x ) species</td>
<td>parameter optimization</td>
</tr>
</tbody>
</table>

\( F_{ij} \)'s follow directly from the gametic process outlined above (see also the electronic supplementary material). For example,

\[
F_{22} = (1 - b)^2, \tag{2.2a}
\]

\[
F_{23} = \frac{(1 - c)^2}{4}, \tag{2.2b}
\]

\[
F_{32} = 2b(1 - b) \tag{2.2c}
\]

and

\[
F_{33} = \frac{(1 - c)^2}{2}, \tag{2.2d}
\]

where \( b \) is the proportion of unreduced gametes in the diploid and \( c \) is the proportion of unreduced gametes in the triploid.

The equilibrium abundances of individual ploidy levels are then given by the dominant eigenvector of the matrix of coefficients of the recurrent equation (2.1):

\[
M = \begin{pmatrix} F_{22} & \mu F_{23} & \cdots \\ F_{32} & \mu F_{33} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}, \tag{2.3}
\]

which depends only on the values of \( \mu \) and \( F_{ij} \). The latter in turn depends on probabilities associated with the formation of unreduced gametes in diploids (\( b \)) and triploids (\( c \)).

(b) Ploidy-level frequency data

Ploidy estimates using DNA flow cytometry were performed as a part of other projects during 2000–2011. In addition, data on species and population ploidy composition were excerpted from published articles. In order to guarantee comparability of the results, only species meeting the following criteria were included in the study: (i) the occurrence of at least two ploidy levels from the range between \( 2x \) and \( 6x \), (ii) ploidy data available for a total of at least 90 individuals from at least three populations, (iii) the existence of a sexual mode of reproduction, and (iv) (presumed) autopolyploid origin of higher ploidy levels. Most data originated from contact zones or were collected across the entire distributional range of a species. In addition to species-level data, individual mixed-ploidy populations within these taxa with more than 90 cytотyped individuals were analysed separately, and are referred to here as species_name P#. While population-level data provide more accurate estimates of frequency levels, they can suffer from random shifts in ploidy composition owing to stochastic processes. Our total data set consisted of 26 angiosperm species and additional 13 populations within these species, and included ploidy data for nearly 40 000 individuals (see the electronic supplementary material, table S1). Species lacking diploids and having all ploidy levels divisible by one small integer (Cenestium spp., Elytrigia spp.) were analysed so that the lowest ploidy level was assumed to be diploid.

(c) Fitting the model to the data: choice of models to be fit

We fit the gametic model to empirical data for each species/population. For the fitting, we introduced some simplifying assumptions. First, we assumed that only diploids and triploids are capable of producing unreduced gametes. Further, unless otherwise stated, we assumed that all \( \mu_5 = \mu_4 = \mu_2 \ldots \), i.e. fertility of all higher ploidy to be equal over all odd and even ploidy levels. Although fertility in triploids and pentaploids is often markedly lower [7,19], analysis of the model behaviour shows that the effect of \( \mu_5 \) and \( \mu_6 \) on model predictions is rather low relative to the fertility and survival of even-ploid plants.

For species with ploidy levels up to tetraploid, we fit a model with the parameter \( c \) (i.e. proportion of unreduced gametes in a triploid) set to zero (‘model I’, which is essentially the model of Husband [10]; table 1). The model thus has only two parameters, and the \( 2x \)–\( 3x \)–\( 4x \) system has two degrees of freedom, meaning that the system is just-identified. Further, to examine whether the assumption that \( c = 0 \) is reasonable in a system that contains all three ploidy levels, we relaxed the condition \( c = 0 \), re-fit the model and checked the value of \( c \). For species having ploidy levels only up to triploid, we also fit model I with the additional constraint, \( \mu_3 = 0 \) (referred to as the ‘minimal model’). In species/populations with ploidy levels from \( 2x \) up to \( 6x \), we fit the model with parameters \( b \), \( c \) and \( \mu \) (‘model II’).

(d) Fitting the model to the data: techniques

Parameters of model I and the minimal model were determined exactly by finding their analytic solutions by hand or using Mathematica v. 8 (Wolfram Research, Inc., Champaign, IL, USA; see the electronic supplementary material S3). Model I has no meaningful analytic solution if the number of individuals of any ploidy level is zero (except for the trivial solutions at \( b = 0 \) or \( \mu = 0 \)); in such cases, we replaced the zero relative frequency by \( 1/(2N) \), where \( N \) is the sample size for a given species (i.e. assuming that one individual of the missing ploidy level would be found if the sample size were doubled). The values obtained thus have to be viewed as upper/lower limits of true values. The model can always be solved, as in both these cases the number of parameters to be estimated equals the number of independent data points (proportion of triploids in the minimal model, and proportions of triploids and tetraploids in model I). We examined whether the values found are biologically meaningful.

In model II, parameter values were estimated using a parameter optimization procedure in which we compared observed frequencies of individual ploidy levels with the models’ predictions, using

\[
\sum \left( \frac{n_i}{\sum n_j} - \frac{n_i}{\sum n_j} \right)^2 \]

as a criterion to be minimized, where \( n_i \) is the observed number of individuals of the \( i \)th ploidy in a given species and \( n_j \) is the element of the dominant eigenvector corresponding to the \( i \)th ploidy of the matrix \( M \) (equation (2.3)). Parameter optimization was done by a quasi-Newton method of Broyden, Fletcher, Goldfarb and Shanno (BFGS) as implemented in the function \texttt{optim} in R v. 2.9.1 (module stats; [20]). All the parameters we work with have natural constraints: probabilities of unreduced gamete formation are bound by zero and unity; relative fertility has a lower bound of zero, and although it can theoretically have values greater than unity, in accordance with the empirical data we set its upper limit to unity as well. We therefore used a modified version of
3. Results

(a) Predictions of the gametic model with only diploids capable of producing unreduced gametes (model I)

The behaviour of the model shows a region of coexistence of all three cytotypes (when \( \mu < 1 - b \)) and a region where only tetraploids exist (figure 1). It can be shown analytically that it has a threshold at \( \mu = 1 - b \), close to which equilibrium proportions of individual ploidies are extremely sensitive to the parameter values (figure 1). Equilibrium proportions of cytotypes depend on both parameters \( \mu \) and \( b \) and are primarily affected by the frequency of unreduced gametes in the diploid (figure 1). The effect of the relative fertility of tetraploids is highly nonlinear: close to the threshold, proportions of ploidies are highly sensitive to \( \mu \), but otherwise the effect of \( \mu \) is rather weak (figure 1).

Triploid plants are never predicted by the model to have high frequencies. The proportion of triploids is always affected by the proportion of unreduced gametes in the diploid \( b \), and, in the variant of model I with a general value of triploid fertility \( \mu_3 \) (i.e. \( \mu_3 \not= \mu_4 \)), also, but to a lesser degree, by fertility of triploids \( \mu_3 \) relative to tetraploids \( \mu_4 \). Because triploids are inherently unstable (only half of their progeny are triploid if unreduced gametes are not generated, and an even lower proportion if they are generated; see the electronic supplementary material, table S2 and figure S2), their frequency is governed primarily by the parameters of their formation, not by their own fertility.

(b) Predictions of the gametic model with both diploids and triploids capable of producing unreduced gametes (model II)

In this scenario, there can be five ploidy levels, from diploid to hexaploid. In principle, the system has six parameters, but assuming \( \mu_3 = \mu_4 = \mu_5 = \mu_6 \) (which we will denote \( \mu \) leaves a system with three parameters. If relative survival rates of higher ploidy levels (\( \mu \)) are sufficiently far from unity, the system has a stable ploidy structure, which is determined by all three parameters. In this case, either diploids or tetraploids can dominate the system; proportions of individual ploidies are largely determined by the proportion of unreduced gametes produced by diploids \( b \) and much less by the proportion of unreduced gametes from triploids \( c \); figure 2).

Similar to the previous case, there is a threshold value of the relative fertility of higher ploidies (\( \mu \)), above which the ploidies are not able to coexist (figure 2). In contrast to the previous case, however, above the threshold the matrix \( M \) has two identical eigenvalues corresponding to two different stable structures: tetraploids only and hexaploids only. The incidence and proportions of these two ploidies are thus not determined by the structure of the matrix only, but also by the initial conditions. No other ploidy is able to persist above the threshold. The position of the threshold is affected only by the values of \( b \) (proportion of unreduced gametes in the diploid) and relative fertility of higher ploidies (\( \mu \)), and is
the transition matrix has two identical eigenvalues. In species with 2\(x\) and 3\(x\) ploidy levels, the predicted values of relative fertility of higher ploidies were considerably smaller, although typically non-zero. The minimal model (i.e. keeping the relative fertility of higher ploidies at zero), yielded higher values of \(b\) and equally good (if not better) fit.

When model II was fit to species/populations with 2\(x\), 3\(x\) and 4\(x\) ploidies, it mostly predicted the value of \(c\) (proportion of unreduced gametes of the triploid) to be zero, which confirmed that model I is sufficient to explain the ploidy structure in these species. In a few cases, however, it predicted a non-zero value of \(c\) and consequently also a low proportion of hexaploids (which were not found in the field using our sample sizes). In Chamerion angustifolium, Knautia arvensis \(P3\) and Melampodium leucanthum, data seem to be compatible both with model I with \(c = 0\) (i.e. no unreduced gametes of the triploid) and model II with \(c > 0\) (table 2). Relative fertilities of higher ploidies were predicted by model II to be close to unity in most species. Only in the two populations of Ga. arcelata to which model I could not be fit was the relative fertility of tetraploids estimated by model II considerably smaller than 0.9.

A satisfactory fit using model II could be obtained in nine out of 11 species/populations with 2\(x\)–6\(x\) ploidy levels (table 2). Values of \(b\) fell into a similar range as in the 2\(x\)–3\(x\)–4\(x\) species, but proportions of unreduced gametes in triploids (\(c\)) were (rather expectedly) predicted to be considerably higher than proportions of unreduced gametes in diploids (\(b\)). They ranged widely, from almost zero (in K. arvensis \(P4\)) to over 0.7 (in Solidago altissima). The fertility of tetraploids was again predicted to be rather close to unity in most of the species. Pilosella echioides, although its ploidy structure was predicted very well (table 2), had quite exceptional predicted parameter values (very high \(b\) and very low fertility of higher ploidies). Oxycoccus palustris and P. echioides \(P3\) could not be fit by model II at all. In Oxycoccus, high frequency of pentaploids is not compatible with a complete absence of triploids (and hence diploids) under the simple gametic model.

**4. Discussion**

The last decades have seen several attempts to model the conditions, allowing establishment of polyploid derivatives within a population of diploid/lower-ploid ancestors [8–15]. Our work adds to these efforts and presents a gametic model for different diploid-autopolyploid systems occurring under natural conditions. We show that simple assumptions of the gametic process yield a model that predicts a number of patterns observed in the field, and parameters of this process can be estimated (either analytically or using an optimization procedure) from proportions of individual ploidy levels under natural conditions. While the model involves a number of simplifications (absence of crosses between ploidy levels, and identical fertility of all higher ploidies), it does capture the key processes in autopolyploid generation and

**Figure 2.** Equilibrium proportions of ploidy levels if both diploid and triploid can produce unreduced gametes (model II). (a) \(b = 0.05, c = 0.1\); (b) \(\mu = 0.8, c = 0.1\); (c) \(\mu = 0.8, b = 0.05\). Note that beyond the threshold (proportion of unreduced gametes of the diploid > 0.157 in this case), two equilibrium values (tetraploids only or hexaploids only) exist, as the transition matrix has two identical eigenvalues.

(c) Fitting gametic models to the empirical data

In species with 2\(x\), 3\(x\) and 4\(x\) ploidy levels, model I could be successfully fit to all species/populations in the dataset with the exception of two populations of Galax arcelata (table 2; see also figure 3). Aside from one other population of Ga. arcelata, the estimated values of \(b\) range from \(5 \times 10^{-4}\) to approximately \(5 \times 10^{-2}\). Relative fertilities of higher ploidies in model I were very high, close to unity (but smaller than 1–\(b\), which is the condition for coexistence of all ploidies). In species with only 2\(x\) and 3\(x\) ploidy levels, the predicted values of relative fertility of higher ploidies were considerably smaller, although typically non-zero. The minimal model (i.e. keeping the relative fertility of higher ploidies at zero), yielded higher values of \(b\) and equally good (if not better) fit.
Table 2. Estimated parameter values for species/populations and the values' fit to the data. (Parameters of the minimal model and model I were determined analytically, i.e. no measure of fit was used. For the definition of the fit measure see §2. b, estimated proportion of unreduced gametes of the diploid; c, estimated proportion of unreduced gametes of the triploid; µ, relative fertility of higher ploidy levels. n.a., no meaningful analytic solution exists or no meaningful values were found by the optimization procedure. Empty cells mean cases in which a given model was not fit to the data for a priori reasons.)

<table>
<thead>
<tr>
<th>species</th>
<th>proportion of odd ploidy individuals</th>
<th>no. ploidy levels</th>
<th>max. ploidy level*</th>
<th>minimal model</th>
<th>model I</th>
<th>model II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cerastium alsinifolium</td>
<td>0.005</td>
<td>2</td>
<td>3</td>
<td>0.0023</td>
<td>0.0015</td>
<td>0.6669</td>
</tr>
<tr>
<td>Cerastium arvense</td>
<td>0.026</td>
<td>2</td>
<td>3</td>
<td>0.0132</td>
<td>0.0108</td>
<td>0.3822</td>
</tr>
<tr>
<td>Elytrigia intermedia</td>
<td>0.022</td>
<td>2</td>
<td>3</td>
<td>0.0110</td>
<td>0.0084</td>
<td>0.4685</td>
</tr>
<tr>
<td>Elytrigia repens</td>
<td>0.010</td>
<td>2</td>
<td>3</td>
<td>0.0049</td>
<td>0.0033</td>
<td>0.6169</td>
</tr>
<tr>
<td>Galax urceolata P2</td>
<td>0.270</td>
<td>2</td>
<td>3</td>
<td>0.1561</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>P. hieracioides</td>
<td>0.011</td>
<td>2</td>
<td>3</td>
<td>0.0056</td>
<td>0.0042</td>
<td>0.4931</td>
</tr>
<tr>
<td>Campanula rotundifolia agg.</td>
<td>0.002</td>
<td>3</td>
<td>4</td>
<td>0.0009</td>
<td>0.9981</td>
<td>0.0115</td>
</tr>
<tr>
<td>Cardamine amara</td>
<td>0.025</td>
<td>3</td>
<td>4</td>
<td>0.0166</td>
<td>0.9724</td>
<td>0.0284</td>
</tr>
<tr>
<td>Chamerion angustifolium</td>
<td>0.033</td>
<td>3</td>
<td>4</td>
<td>0.0156</td>
<td>0.9648</td>
<td>0.0164</td>
</tr>
<tr>
<td>Empetrum nigrum</td>
<td>0.018</td>
<td>3</td>
<td>4</td>
<td>0.0155</td>
<td>0.9777</td>
<td>0.0263</td>
</tr>
<tr>
<td>Ga. urceolata</td>
<td>0.110</td>
<td>3</td>
<td>4</td>
<td>0.0553</td>
<td>0.8644</td>
<td>0.0337</td>
</tr>
<tr>
<td>Ga. urceolata P1</td>
<td>0.228</td>
<td>3</td>
<td>4</td>
<td>0.3348</td>
<td>0.5271</td>
<td>0.3341</td>
</tr>
<tr>
<td>Ga. urceolata P3</td>
<td>0.220</td>
<td>2</td>
<td>4</td>
<td>n.a.</td>
<td>n.a.</td>
<td>0.8800</td>
</tr>
<tr>
<td>Ga. urceolata P4</td>
<td>0.000</td>
<td>2</td>
<td>4</td>
<td>0.0019</td>
<td>0.9946</td>
<td>0.0071</td>
</tr>
<tr>
<td>Heuchera grossulariifolia</td>
<td>0.014</td>
<td>3</td>
<td>4</td>
<td>0.0041</td>
<td>0.9638</td>
<td>0.0041</td>
</tr>
<tr>
<td>Knautia arvensis</td>
<td>0.001</td>
<td>3</td>
<td>4</td>
<td>0.0010</td>
<td>0.9984</td>
<td>0.0178</td>
</tr>
<tr>
<td>K. arvensis P1</td>
<td>0.000</td>
<td>2</td>
<td>4</td>
<td>0.0008</td>
<td>0.9989</td>
<td>0.0083</td>
</tr>
<tr>
<td>K. arvensis P2</td>
<td>0.000</td>
<td>2</td>
<td>4</td>
<td>0.0005</td>
<td>0.9986</td>
<td>0.0061</td>
</tr>
<tr>
<td>K. arvensis P3</td>
<td>0.000</td>
<td>2</td>
<td>4</td>
<td>0.0012</td>
<td>0.9949</td>
<td>0.0044</td>
</tr>
<tr>
<td>Melampodium cinerereum</td>
<td>0.006</td>
<td>3</td>
<td>4</td>
<td>0.0031</td>
<td>0.9935</td>
<td>0.0107</td>
</tr>
<tr>
<td>Melampodium leucanthum</td>
<td>0.002</td>
<td>3</td>
<td>4</td>
<td>0.0007</td>
<td>0.9967</td>
<td>0.0039</td>
</tr>
<tr>
<td>Oxalis hirta P1</td>
<td>0.028</td>
<td>3</td>
<td>4</td>
<td>0.0027</td>
<td>0.9963</td>
<td>0.0371</td>
</tr>
<tr>
<td>Oxalis pes-caprae</td>
<td>0.000</td>
<td>2</td>
<td>4</td>
<td>0.0023</td>
<td>0.9971</td>
<td>0.0280</td>
</tr>
<tr>
<td>Parasenecio auriculata</td>
<td>0.004</td>
<td>3</td>
<td>4</td>
<td>0.0017</td>
<td>0.9963</td>
<td>0.0093</td>
</tr>
<tr>
<td>Ranunculus saxatilis</td>
<td>0.001</td>
<td>3</td>
<td>4</td>
<td>0.0006</td>
<td>0.9988</td>
<td>0.0134</td>
</tr>
<tr>
<td>Plantago media</td>
<td>0.002</td>
<td>3</td>
<td>4</td>
<td>0.0008</td>
<td>0.9982</td>
<td>0.0089</td>
</tr>
</tbody>
</table>

(Continued.)
### Table 2. (Continued)

<table>
<thead>
<tr>
<th>species</th>
<th>proportion of odd ploidy individuals</th>
<th>no. ploidy levels</th>
<th>max. ploidy level&lt;sup&gt;a&lt;/sup&gt;</th>
<th>minimal model</th>
<th>model I</th>
<th>model II</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>b</td>
<td>b</td>
<td>μ</td>
</tr>
<tr>
<td>Ranunculus adonaeus P1</td>
<td>0.017</td>
<td>3</td>
<td>4</td>
<td>0.0093</td>
<td>0.9829</td>
<td>0.0173</td>
</tr>
<tr>
<td>Vicia cracca</td>
<td>0.001</td>
<td>3</td>
<td>4</td>
<td>0.0008</td>
<td>0.9988</td>
<td>0.0194</td>
</tr>
<tr>
<td>Actinidia chinensis</td>
<td>0.006</td>
<td>4</td>
<td>6</td>
<td>0.0063</td>
<td>0.6367</td>
<td>0.9665</td>
</tr>
<tr>
<td>Gentiana jacea</td>
<td>0.000</td>
<td>3</td>
<td>6</td>
<td>0.0159</td>
<td>0.0000</td>
<td>0.9730</td>
</tr>
<tr>
<td>Gymnadenia conopsea agg.</td>
<td>0.022</td>
<td>5</td>
<td>6</td>
<td>0.0151</td>
<td>0.0031</td>
<td>0.9709</td>
</tr>
<tr>
<td>Gy. conopsea P1</td>
<td>0.045</td>
<td>5</td>
<td>6</td>
<td>0.0273</td>
<td>0.0043</td>
<td>0.9510</td>
</tr>
<tr>
<td>Gy. conopsea P2</td>
<td>0.017</td>
<td>5</td>
<td>6</td>
<td>0.0075</td>
<td>0.0091</td>
<td>0.9741</td>
</tr>
<tr>
<td>K. arvensis P4</td>
<td>0.000</td>
<td>4</td>
<td>6</td>
<td>0.0284</td>
<td>0.0000</td>
<td>0.9644</td>
</tr>
<tr>
<td>Oxalis purpurea</td>
<td>0.000</td>
<td>3</td>
<td>6</td>
<td>0.0083</td>
<td>0.3473</td>
<td>0.9710</td>
</tr>
<tr>
<td>Oxycoccus palustris</td>
<td>0.169</td>
<td>3</td>
<td>6</td>
<td>n.a.</td>
<td>n.a.</td>
<td>0.3603</td>
</tr>
<tr>
<td>Pilosella echoides</td>
<td>0.444</td>
<td>5</td>
<td>6</td>
<td>0.4364</td>
<td>0.1038</td>
<td>0.1249</td>
</tr>
<tr>
<td>P. echoides P1</td>
<td>0.741</td>
<td>5</td>
<td>6</td>
<td>n.a.</td>
<td>n.a.</td>
<td>0.0870</td>
</tr>
<tr>
<td>Solidago altissima</td>
<td>0.000</td>
<td>3</td>
<td>6</td>
<td>0.0060</td>
<td>0.7279</td>
<td>0.9658</td>
</tr>
</tbody>
</table>

*aUnder assumption that the lowest ploidy level is diploid.*
than relative production of
dies only if survival/fertility of higher ploidies is lower
ploids, and is thus partly arbitrary. The same type of
m) depends on the upper estimate of the frequency of tetra-
fitness parameters (such as the proportions of unreduced
gametes and survival rates of higher ploidy plants) can be
substituted these parameters into the model and can be
ascertaining whether (i) predicted ploidy frequencies match
with the empirical ones, and (ii) the inferred parameter
values are compatible with values obtained using more
direct (experimental) approaches.

With a few exceptions (Ga. urceolata, O. palustris and
P. echoides, which often showed low fit), the predicted fre-
cuencies of unreduced gamete production for diploid
plants (5 × 10^{-4} – 5 × 10^{-7}) matched the empirical data
well. The mean production of diploid gametes across non-
hybrid flowering plants is 5.6 × 10^{-3} [6], which is not
very far from the values predicted by our models (mean
1.66 × 10^{-2}; median 8.9 × 10^{-3}, across all plants/populations
and models with good fit; the unrealistically high values for
the above-mentioned species excluded). Under natural con-
ditions, the observed frequencies of diploid gametes vary
substantially among species as well as among conspecific
individuals. For example, the proportion of unreduced
pollen in individual clones of Vaccinium sect. Cyanococcus
varied between 0.01 and 0.28 [22], while mean population-
level frequencies of unreduced pollen in Achillea borealis
ranged from 3 × 10^{-4} to 5.4 × 10^{-3} but were as high as
0.158 in individual plants [23].

Conversely, external knowledge of the reproductive and
fitness parameters (such as the proportions of unreduced
gametes and survival rates of higher ploidy plants) can be
used as a further test of the model realism. In such a case, pre-
dicted frequencies of minority cytotypes follow directly from
substituting these parameters into the model and can be
compared with the field-observed values.

Sample size permitting, the model was fit both to species
(i.e. cumulative ploidy frequencies across all populations)
and individual populations. Despite the greater effect of sto-
chastic processes (which can cause the elimination of a
particular cytotype and preclude model fitting; 3x and 4x
populations of Ga. urceolata P2 may be an example), the
latter show slightly better fit perhaps owing to more accurate
field data on ploidy composition at finer spatial scales
(cf. population and species data for Gymnadenia conopsea or
K. arvensis). In a few cases, some species/populations could

**Figure 3.** Examples of observed (grey bars) and predicted (black bars) proportions of ploidy levels for 2x–4x ((a) Chamerion angustifolium) and 2x–6x ((b) Solidago gigantea) species (model II). The 2x–4x system has complete fit as the number of parameters equals the number of independent values to be fit; the 2x–6x system has fewer parameters than independent values, resulting in incomplete fit.
not be fit with biologically meaningful parameters, indicating incompatibility between what is occurring in nature and one or more model assumptions. First, individual cytotypes, especially in mixed-ploidy populations and/or in contact zones, are not necessarily fully reproductively isolated and can hybridize. Whereas there often is a strong breeding barrier between diploids and tetraploids (i.e. the so-called triplid block; [19]), increase in genome copy number can be associated with relaxation of intercytotype reproductive barriers [24]. Consequently, interploidy interactions can play more important roles in shaping ploidy composition than do gamete-driven processes. Oxycoccus palustris [25] and P. echioides [26] probably fall into this category and both show poor fit or unrealistic predicted parameter values. Moreover, the model predicts, in agreement with most field data, large prevalence of even ploidy levels. Therefore, fit was usually unsatisfactory (or yielded unrealistic predicted parameter values) in groups with a high proportion of odd ploidy levels (Ga. urceolata, P. echioides). Additionally, the current formulation always predicts a non-zero proportion of diploids in species with ploidy heterogeneity, therefore entailing poor or even no fit in species/populations lacking diploids (e.g. O. palustris). In general, poor fit to the gametic model can be used as a strong indication of action of other processes (hybridization, strongly unequal fertility or survival of higher ploidy levels, etc.), in the species in question.

(c) Implications
The main feature of the gametic model is the requirement that higher ploidy levels must be less fertile relative to diploids if all ploidy species/populations needs to be better understood, a simple gametic model may also be used to identify such cases from field-collected ploidy frequency data.

The support was provided by the Academy of Science of the Czech Republic (long-term research development project no. RVO 67985939), institutional resources of Ministry of Education, Youth and Sports of the Czech Republic for the support of science and research, and the Czech Science Foundation (project 206/09/0843). The William J. Fulbright Commission and University of Michigan supported T. H. during his sabbatical stay at the Department of Ecology and Evolutionary Biology. We thank all our colleagues for permission to use their unpublished data on ploidy frequencies (Martin Hanzl, Kinutia; Jana Krejčíková, Campovalo, Osaka; Karol Marhold, Cardamim; Petr Vit, Cerastium). Jonathan Rosenthal kindly improved the English. We thank two anonymous referees for their comments that helped us to clarify the paper.

References


