Genetic Variation and Random Drift in Autotetraploid Populations

Michael E. Moody,* Laurence D. Mueller† and Douglas E. Soltis‡

**Department of Pure and Applied Mathematics and Department of Genetics and Cell Biology, ‡Department of Botany, Washington State University, Pullman, Washington 99164-4234, and †Department of Ecology and Evolutionary Biology, University of California, Irvine, California 92717*

Manuscript received August 24, 1992
Accepted for publication February 25, 1993

ABSTRACT

The rate of decay of genetic variation is determined for randomly mating autotetraploid populations of finite size, and the equilibrium homozygosity under mutation and random drift is calculated. It is shown that heterozygosity is lost at a slower rate than in diploid populations, and that the equilibrium heterozygosity with mutation and random drift is higher than for diploids. Outcrossing populations as well as populations that randomly self are analyzed. A method of comparing genetic variation between autotetraploid and diploid populations is proposed. Our treatment suggests that the "gametic homozygosity" provides a unified approach for comparing genotypes within a population as well as comparing genetic variation between populations with different levels of ploidy.

IT is well accepted that polyploidy is of major importance in the evolution of plants. This is reflected by the large number of species of polyploid origin: estimates for the percentage of polyploid angiosperms range from 30% to 52% (STEBBINS 1950; GRANT 1981). Two broad categories of polyploids are recognized, autopolyploid and allopolyploid. Although these two types of polyploidy are different, they can be considered as extremes in a spectrum (STEBBINS 1947, 1950; CLAUSEN, KECK and HIESEY 1945).

Polyploid inheritance is complicated by the large number of modes of gamete formation (MATHER 1935, 1936). These modes are a consequence of recombination between loci as well as of recombination between loci and the centromere, and have been classified and extensively studied by FISHER (1947, 1962). The population genetic consequences of polysomic inheritance are considerable. An immediate casualty of an increase in ploidy is the familiar Hardy-Weinberg law. Hardy-Weinberg proportions are attained only gradually in panmictic polyploid populations (HALDANE 1930; BENNETT 1953, 1954), and, unless there is no double reduction, equilibrium frequencies of zygotes are not proportional to the corresponding product of allele frequencies (LI 1955; SEYFFERT 1960; ELANDT-JOHNSON 1967; BENNETT 1968). Consideration of linkage in polysomic inheritance adds to the complexity (FISHER 1943, 1947; GEIRINGER 1949a,b,c; BENNETT 1953; CROW 1954; GRIFFING 1957). Inasmuch as a majority of naturally occurring polysomic populations are not strictly panmictic (STEBBINS 1980), as well as many cultivated polysomic crop plants, the effect of the mating system on the evolution of polyploids is important. This effect has been relatively well studied, especially for self-fertilization and regular systems of inbreeding (KEMPTHORNE 1957; SEYFFERT 1959; BENNETT 1968, 1976; MCCONNELL and FYFE 1975). KEMPTHORNE (1955, 1957) has calculated the correlation between relatives for autotetraploids and autohexaploids, and the probabilistic methods of MALÉCOT (1966) for the analysis of inbreeding have been extended to polysomics (GALLAIS 1967, 1969; GUY 1970, 1972; GUY and GRENIER 1967). This theory has been applied to study inbreeding depression in polysomic crop plants (*e.g.*, BUSBICE and WILSIE 1966; RICE and DUDLEY 1974). To a limited extent the effect of selection in polysomic populations has been investigated (WRIGHT 1938; PARSONS 1959), as well as the quantitative genetics of autotetraploids (LI 1957; KEMPTHORNE 1957; BUSBICE and WILSIE 1966; KILLICK 1971).

Whereas allopolyploidy is known to be widespread in many plant groups, including angiosperms and pteridophytes, autopolyploid speciation has generally been considered to be rare and of little evolutionary importance (CRAWFORD 1985; LEVIN 1983; LEWIS 1980; SOLTIS and RIESEBERG 1986; STEBBINS 1980). As a result of this long-standing view, relatively few studies have addressed the practical or theoretical problems regarding autopolyploid speciation, although there are noteworthy exceptions (*e.g.*, FOWLER and LEVIN 1984). The genetic and ecological attributes of naturally occurring autopolyploids and their presumed diploid progenitors have rarely been investigated in detail.

Recent studies indicate, however, that autopolyploid speciation is more common and evolutionarily important than heretofore appreciated (CRAWFORD 1985; EHRENDORFER 1980; LEVIN 1983; SAMUEL, PIN-

SKER and EHRENDORFER 1990; SOLTIS and RIESEBERG 1986; SOLTIS and SOLTIS 1989a,b; RIESEBERG and DOYLE 1989). Autotetraploids, at least newly formed ones, are characterized by tetrasomic inheritance (BARBER 1970; HALDANE 1930, MULLER 1914; STEBBINS 1947, 1950) and, as described above, present special theoretical and practical complications for both modeling and data analysis; diploids and allopolyploids are simpler in this regard, both of which exhibit disomic inheritance. Although much population genetics theory has been developed for autopolyploid evolution, many questions remain unanswered. In particular, relatively little is known about random drift in finite polysomic populations (ROWE 1986) and the equilibrium levels of genetic variation in finite populations with mutation has not been characterized.

By the very nature of polysomy, we expect that greater heterozygosity would be maintained in an autotetraploid population than in a diploid population of the same size. This argument has been invoked to explain the evolutionary success (or potential success) of autopolyploids in nature (BARBER 1970; HALDANE 1930; MULLER 1914; STEBBINS 1947, 1950, 1980). Electrophoretic investigations of autopolyploid speciation suggest that heterozygosity is significantly greater in populations of autopolyploids than in those of their diploid progenitors (SOLTIS and RIESEBERG 1986). It is unclear, however, exactly how comparisons of heterozygosity should be made between polyploid and diploid populations and how tetrasomic inheritance intrinsically affects the level of genetic variation. In this paper we propose a measure of genetic variation in finite autotetraploid populations, describe its evolution, and discuss a means of comparing such populations to their diploid counterparts.

THE MODEL

No mutation: Consider an autosomal locus in an autotetraploid population of N individuals. For simplicity we will ignore double reduction and therefore suppose that the probability of segregating homoallelic gametes is zero (CROW and KIMURA 1970). We assume that generations are discrete and nonoverlapping and that neither natural selection, mutation, migration nor any other systematic evolutionary force influences the evolution of the population. Mating will be at random and according to the following scheme: each generation every adult produces an identical and very large number of gametes, from the totality of which N pairs are sampled at random to form the N individuals of the succeeding generation; we posit that inheritance is tetrasomic. Genetic variation in the population will be characterized in generation t (t = 0, 1, 2, . . .) by two probabilities of identity, denoted f_t and g_t. We refer to f_t as the *gametic homozygosity* (DEMARLY 1963), defined to be the probability that a randomly chosen gamete from a randomly selected individual of generation t is homozygous at the given locus; since adults are autotetraploid, every gamete has two copies of each locus. The probability g_t represents the chance that a pair of homologous chromosomes, each randomly selected from randomly chosen gametes produced by distinct individuals in generation t, are *identical in state* at the given locus.

Two special cases will be distinguished according to whether or not selfing is allowed to occur.

No selfing: In order to analyze the behavior of f_t and g_t, we will derive the recursion relations that relate them between successive generations. To that end, we note that it is with probability 1/3 that the pair of chromosomes containing a given locus in a randomly selected gamete from an individual were paired in a gamete in the previous generation. With chance 2/3, these two chromosomes in our selected gamete were derived from chromosomes contained in separate gametes of the previous generation; since selfing is not allowed, these two gametes must have been derived from distinct individuals of the previous generation. This observation leads immediately to our first equation (KEMPTHORNE 1957)

$$f_{t+1} = \frac{1}{3} f_t + \frac{2}{3} g_t. \tag{1}$$

If we now consider two homologous genes drawn at random in generation $t + 1$ from gametes produced by distinct individuals, we readily see that it is with probability $1/N$ that these two genes were independently derived from the same individual of generation t, and with the complementary chance $1 - 1/N$ they were derived from distinct individuals. In the former instance the probability of identity of the two genes is easily shown to be (KEMPTHORNE 1957):

$$\frac{1}{4} + \frac{3}{4} f_t,$$

and in the latter situation the probability of identity is simply g_t. Consequently,

$$g_{t+1} = \frac{1}{4N} (1 + 3f_t) + \left(1 - \frac{1}{N}\right) g_t; \tag{2}$$

combined with (1) this gives the desired recursions.

The system will be analyzed in terms of the variables $h_t = 1 - f_t$ (the *gametic heterozygosity*) and $k_t = 1 - g_t$. This results in the homogeneous system

$$h_{t+1} = \frac{1}{3} h_t + \frac{2}{3} k_t,$$

$$k_{t+1} = \frac{3}{4N} h_t + \left(1 - \frac{1}{N}\right) k_t,$$

which is conveniently represented in the matrix form

$$\mathbf{v}_{t+1}=\begin{pmatrix}\frac{1}{3} & \frac{2}{3}\\ \frac{3}{4N} & 1-\frac{1}{N}\end{pmatrix}\mathbf{v}_t \quad \text{where} \quad \mathbf{v}_t=\begin{pmatrix}h_t\\ k_t\end{pmatrix}. \tag{3}$$

A straightforward calculation reveals that the largest eigenvalue λ_o of the matrix appearing in (3) is

$$\lambda_o=\frac{2}{3}-\frac{1}{2N}+\frac{1}{3}\left(1+\frac{3}{2N}+\frac{9}{4N^2}\right)^{1/2}. \tag{4}$$

A simple expansion elucidates the asymptotic behavior of λ_o for large population size:

$$\lambda_o=1-\frac{1}{4N}+O(N^{-2}) \quad \text{as} \quad N\to\infty.$$

Thus the gametic heterozygosity decays to 0 as generations advance, and the asymptotic rate of decay of gametic heterozygosity is approximately $1-1/(4N)$ for large N. This is slower than the corresponding rate of decay of heterozygosity in diploid populations, $1-1/(2N)$. As we expect, heterozygosity is lost less rapidly in autotetraploid populations, although the population will still eventually become monomorphic.

Random selfing allowed: If selfing occurs the recursion equations must be modified to account for the possibility that homologous chromosomes from an individual of generation $t+1$, originating from distinct gametes of the previous generation, are in fact descended from the same individual of generation t. Using similar reasoning as before we now obtain

$$f_{t+1}=\frac{1}{6N}+\left(\frac{1}{3}+\frac{1}{2N}\right)f_t+\frac{2}{3}\left(1-\frac{1}{N}\right)g_t, \tag{5}$$

$$g_{t+1}=\frac{1}{4N}(1+3f_t)+\left(1-\frac{1}{N}\right)g_t \tag{6}$$

as the pertinent recursion equations.

Changing variables as before now gives

$$\mathbf{v}_{t+1}=\begin{pmatrix}\frac{1}{3}+\frac{1}{2N} & \frac{2}{3}\left(1-\frac{1}{N}\right)\\ \frac{3}{4N} & 1-\frac{1}{N}\end{pmatrix}\mathbf{v}_t,$$

from which we determine the asymptotic rate of decay λ_s to be

$$\lambda_s=\frac{2}{3}-\frac{1}{4N}+\frac{1}{3}\left(1+\frac{9}{16N^2}\right)^{1/2}. \tag{7}$$

Because the leading eigenvalue of an irreducible nonnegative matrix is bounded below by the least and above by the greatest row sum (GANTMACHER 1960), we infer from (3) and the matrix following (6) that

$$1-\frac{1}{N}+\frac{1}{2N}<\lambda_s<1-\frac{1}{N}+\frac{3}{4N}<\lambda_o<1.$$

Thus, $\lambda_s<\lambda_o$ for all $N\geq 2$: as expected, genetic variation ultimately decays more slowly if the population does not self.

Expanding λ_s to first order in N^{-1} leads to the same approximation for large population sizes as previously obtained for λ_o:

$$\lambda_s\sim 1-\frac{1}{4N} \quad \text{as} \quad N\to\infty.$$

The difference between the rates of decay of gametic heterozygosity for selfing and outcrossing populations should therefore be negligible in large populations. This is not surprising because the probability of a random selfing declines to 0 as $N\to\infty$. For small populations though, this difference can be numerically significant: with $N=2$ we obtain $\lambda_s\approx 0.903$ and $\lambda_o\approx 0.926$.

Mutation and random drift: We will hereafter suppose that mutation occurs each generation at the rate u per gene, and that all mutants are novel allelic forms as per the standard infinite-alleles assumption. It is straightforward to see that the equilibrium probabilities of identity $\hat{f}$ and $\hat{g}$ satisfy the linear equations

$$\hat{f}=(1-u)^2\left[\frac{1}{3}\hat{f}+\frac{2}{3}\hat{g}\right],$$

$$\hat{g}=(1-u)^2\left[\frac{1}{4N}(1+3\hat{f})+\left(1-\frac{1}{N}\right)\hat{g}\right],$$

if selfing is not permitted. With selfing allowed, the corresponding equations are

$$\hat{f}=(1-u)\left[\frac{1}{6N}+\left(\frac{1}{3}+\frac{1}{2N}\right)\hat{f}+\frac{2}{3}\left(-\frac{1}{N}\right)\hat{g}\right],$$

$$\hat{g}=(1-u)^2\left[\frac{1}{4N}(1+3\hat{f})+\left(1-\frac{1}{N}\right)\hat{g}\right].$$

Solving these equations yields the equilibrium gametic homozygosities $\hat{f}_o$ and $\hat{f}_s$ for the nonselfing and selfing models, respectively:

$$\hat{f}_o=\frac{\nu^2}{\nu(6-5\nu)+2(1-\nu)(3-\nu)N}, \tag{8}$$

$$\hat{f}_s=\frac{\nu}{\nu(3-2\nu)+2(1-\nu)(3-\nu)N}, \tag{9}$$

where we have defined $\nu=(1-u)^2$. The ν in the numerator of (9) is not squared.

A simple calculation now shows that $\hat{f}_o<\hat{f}_s$ for $N\geq 2$ and $0<\nu<1$; thus there is more homozygosity at equilibrium if selfing occurs than if it does not, as we

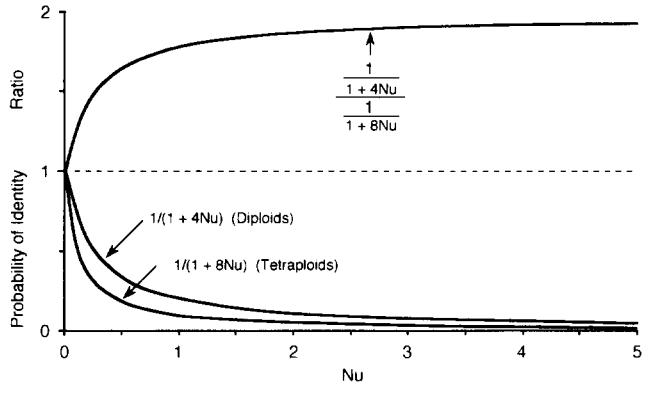

FIGURE 1.—Comparison of the equilibrium homozygosity in a diploid population to the gametic homozygosity in an autotetraploid population as a function of Nu (bottom two curves). The curve above the dashed line is the ratio of the diploid homozygosity to the gametic homozygosity in the tetraploid population. As is clear from the figure, there is less homozygosity at equilibrium in the tetraploid population. The upper curve approaches 2 asymptotically for large Nu.

expect. Furthermore, if $N \gg 1$ and $Nu^2 \ll 1$ we determine from (8) and (9) that $\hat{f}_o$ and $\hat{f}_s$ are asymptotically equal:

$$\hat{f}_o \quad \text{and} \quad \hat{f}_s \approx \frac{1}{1 + 8Nu} \equiv \hat{f}. \tag{10}$$

Since the conditions of the approximation are often likely to hold, the difference in equilibrium gametic homozygosity between selfing and non-selfing populations would not generally be apparent.

It is important to observe that the equilibrium gametic homozygosity is smaller for an autotetraploid population than the equilibrium homozygosity in a diploid population of the same size (which equals $[1 + 4Nu]^{-1}$). This difference can be substantial; in fact if Nu is large the equilibrium gametic homozygosity in an autotetraploid population is approximately one-half that of the homozygosity in a correspondingly large diploid population. This is illustrated in Figure 1, where the ratio of the tetraploid to the diploid homozygosity at equilibrium is plotted as a function of Nu. As the figure reveals, if Nu is relatively small compared to 1 then the tetraploid and diploid populations have comparable amounts of variation, whereas when Nu exceeds 1, there is considerably more variation in the tetraploid population.

Finally, it is useful to recast (10) for the equilibrium homozygosity $\hat{f}$ in the form

$$\hat{f} = \frac{1}{1 + 8Nu} = \frac{1}{1 + 2n_2},$$

where $n_2 = 4Nu$ is the expected number of new mutants arising each generation. For diploids, the equilibrium homozygosity is

$$\frac{1}{1 + 4Nu} = \frac{1}{1 + 2n_1},$$

where $n_1 = 2Nu$ is the expected number of new mutants per generation for diploids. From this perspective, we see that the equilibrium homozygosity takes the same form for both tetraploids and diploids. This invites us to conjecture that the equilibrium gametic homozygosity for an auto $2k$-ploid population will be (approximately)

$$\hat{f}^{(2k)} = \frac{1}{1 + 4kNu}, \quad k = 1, 2, 3 \ldots,$$

for $N \gg 1$, $Nu^2 \ll 1$; this generalization is justified in APPENDIX B.

It is well known for diploids (EWENS and GILLESPIE 1974) that the "homozygosity" estimator obtained by solving the previous equation for Nu is biased. Better estimates are obtained from more sophisticated sampling theory (Ewens 1979). We have not extended this theory to polysomics, and will rely on (12) in the application that follows below.

Gametic homozygosity and genotypic frequencies: As is clear from its definition, the gametic homozygosity is not equal to the homozygosity of the adult population but is instead a function of the frequencies of the several possible tetraploid genotypes. To understand the connection between f and the genotypic frequencies in a given generation, we will suppose that there are alleles A_1, A_2, ... at the given locus with respective frequencies p_1, p_2, ... in the population. Assume that the population is at equilibrium and not too small; from this assumption together with our stipulation that double reduction does not occur we infer that genotypes are approximately in Hardy-Weinberg proportions. We then define the following quantities:

$$\begin{aligned}
\pi_{1,1,1,1} &= \sum_{i<j<k<l} 24 p_i p_j p_k p_l, && (\oplus, \varnothing, \ominus, \otimes) \\
\tau_{2,1,1} &= \sum_i \sum_{j \neq i} \sum_{k>j, k=i} 12 p_i^2 p_j p_k, && (\oplus, \oplus, \ominus, \otimes) \\
\pi_{2,2} &= \sum_{i<j} 6 p_i^2 p_j^2, && (\oplus, \oplus, \otimes, \otimes) \\
\pi_{3,1} &= \sum_i \sum_{j \neq i} 4 p_i^3 p_j, && (\oplus, \oplus, \oplus, \otimes) \\
\pi_4 &= \sum_i p_i^4, && (\oplus, \oplus, \oplus, \oplus).
\end{aligned} \tag{11}$$

(The parenthetical lists of symbols represent generic genotypes involved in the corresponding summation: for example, $\pi_{2,1,1}$ $(\oplus, \oplus, \ominus, \otimes)$ is the total frequency of heterozygotes that have exactly three distinct alleles, π_4 $(\oplus, \oplus, \oplus, \oplus)$ is the total frequency of homozygotes, etc.

Using these definitions and assumptions, it is easy

to show that the gametic homozygosity is given by

$$f = 1 \cdot \pi_4 + \frac{1}{2} \cdot \pi_{3,1} + \frac{1}{3} \cdot \pi_{2,2} + \frac{1}{6} \cdot \pi_{2,1,1} + 0 \cdot \pi_{1,1,1,1} = \pi_4 + \frac{1}{2}\pi_{3,1} + \frac{1}{3}\pi_{2,2} + \frac{1}{6}\pi_{2,1,1}. \quad (12)$$

Thus, if the distribution of genotypic frequencies is known or can be inferred, (12) can be used to estimate the gametic homozygosity of the population. As has been previously noted, the gametic heterozygosity is insufficient to completely characterize the genetic structure of a panmictic tetraploid population (GUY 1972; GLENDINNING 1989). We remark, however, that our aim is to determine the effect of polyploidy on observed genetic variation, whence our focus on identity in state. Thus, we have not developed our theory using the more complicated identity by descent formulae introduced by GALLAIS (1967) for tetraploids.

We suggest that f is an appropriate statistic for comparing genetic variation in autotetraploid populations with diploid populations of comparable size. This notion is reinforced when we realize that the gametic homozygosity of an autotetraploid population and the homozygosity of a diploid population each define the probability that a pair of homologous genes, randomly drawn without replacement from a randomly selected individual, are identical in state at the given locus. The simple distinctions between homozygote and heterozygote that prevail with diploids are not adequate for autotetraploids: not all heterozygotes are created equal. Some are "more heterozygous" than others; alternately, some are more homozygous than others. The gametic homozygosity accounts for this natural heterogeneity of autotetraploid genotypes, as expressed in the above weighted average.

Our calculations suggest that we may order genotypes within a population according to the probability that a randomly chosen homologous pair of genes are identical in state. For a given genotype G let this probability be $\phi(G)$. Define the symbol "$\prec$" to mean "is less homozygous than" and stipulate that $G_1 \prec G_2$ if, and only if, $\phi(G_1) < \phi(G_2)$, for two genotypes G_1 and G_2. With this terminology, we decompose the collection of genotypes into equivalence classes $\mathcal{G}_1$, $\mathcal{G}_2$, . . ., according to the value of ϕ. In other words, two distinct genotypes G_1 and G_2 are in the same equivalence class if, and only if, $\phi(G_1) = \phi(G_2)$. Hence the collection of genotypes G_1, G_2, . . . is partially ordered under $(\phi, \prec)$ and the set of equivalence classes of genotypes $\{\mathcal{G}_1, \mathcal{G}_2, \ldots\}$ is totally ordered under $(\phi, \prec)$, where we define $\phi(\mathcal{G}_i) = \phi(G_j)$ for any $\mathcal{G}_j \in \mathcal{G}_i$. With this characterization the gametic homozygosity for the population becomes

$$f = \sum_{\mathcal{G}_i \in \mathcal{G}} \phi(\mathcal{G}_i)\pi(\mathcal{G}_i), \quad (13)$$

where $\mathcal{G} = \cup_i \mathcal{G}_i$ is the set of all equivalence classes and $\pi(\mathcal{G}_i)$ is the total frequency of genotypes in class $\mathcal{G}_i$. It should be noted that our treatment will apply to any auto-$2k$-ploid population in the absence of double reduction, $k = 1, 2, 3, \ldots$, although higher ploidy levels have diminished biological relevance.

In particular, a diploid population has at most two equivalence classes of genotypes under $(\phi, \prec)$: the homozygote genotypes $[\mathcal{G}_1, \phi(\mathcal{G}_1) = 1]$ and the heterozygotes $[\mathcal{G}_2, \phi(\mathcal{G}_2) = 0]$. Clearly the gametic homozygosity satisfies $f = \pi(\mathcal{G}_1)$ and equals the usual homozygosity of the population. For an autotetraploid population there are at most five equivalence classes $\mathcal{G}_1$, $\mathcal{G}_2$, $\mathcal{G}_3$, $\mathcal{G}_4$, $\mathcal{G}_5$ corresponding respectively to the five classes of genotypes depicted from top-to-bottom in (11). Equation 12 is now seen to differ from (13) merely in notation; the procedure to describe populations with even higher levels of autopolyploidy is completely analogous. It should be noted that, for diploids and autotetraploids, the "natural" ordering of genotypes according to the number of distinct alleles present partitions the genotypes exactly as does $(\phi, \prec)$. This intuitive approach does not, however, agree with the ordering based on gametic homozygosity for higher ploidy levels and, in any event, does not suggest a means of quantifying the relative contributions of different genotypes to the overall genetic variation. Our treatment suggests that the gametic homozygosity offers a unified approach to comparing genotypes within a population as well as comparing genetic variation between populations with different degrees of ploidy.

Effective number of alleles: In analogy with the diploid situation, define the effective number of alleles n_e for an autotetraploid population with gametic homozygosity f as that number of equally frequent alleles giving rise to this same value of f. Consider an autotetraploid population at equilibrium with $n \geq 1$ equally frequent alleles. Put $p_i = 1/n$, $i = 1, \ldots, n$ in (11) to infer, after some elementary combinatorics,

$$\pi_{1,1,1,1} = \frac{24\binom{n}{4}}{n^4}, \quad \pi_{2,1,1} = \frac{12n\binom{n-1}{2}}{n^4},$$

$$\pi_{2,2} = \frac{6\binom{n}{2}}{n^4}, \quad \pi_{3,1} = \frac{4n(n-1)}{n^4}, \quad \pi_4 = \frac{n}{n^4}.$$

When these expressions are inserted into (12), we

determine that $f = 1/n$ and hence from (10) identify the effective number of alleles as

$$n_e = 1 + 8Nu \text{ i.e., } \hat{f} = \frac{1}{n_e}$$

exactly as we would have wished. We conjecture that the effective number of alleles for 2k-ploid populations is $n_e^{(2k)} = 1 + 4kNu$, so that $\hat{f}^{(2k)} = [n_e^{(2k)}]^{-1}$.

APPLICATION

Recent studies have shown that the diploid angiosperm *Tolmiea menziesii* has an average heterozygosity of 7% while individuals in related autotetraploid populations are heterozygous at 24% of their loci (SOLTIS and SOLTIS 1989a) It is natural to inquire whether the genetic system alone accounts for the increased levels of genetic variation observed in these autotetraploid populations. The theory developed above provides a means to address this question.

We proceed as follows. First, Nu can be estimated from the observed heterozygosity in the diploid populations and the theoretical expectation that this heterozygosity will approximately equal $4Nu/(1 + 4Nu)$. This estimate of Nu is then used in (10) to predict the expected gametic homozygosity in the autotetraploid population. We have of course assumed that the diploid and autotetraploid populations both experience the same forces of random drift and mutation.

Using data from 15 loci and 15 local diploid populations this procedure yields a theoretical gametic homozygosity of 0.87 with a 95% confidence interval of [0.78, 0.98] for the autotetraploid populations. Applying (12) to the tetraploid data, we have directly estimated the average gametic homozygosity (see (A1)) for 7 local autotetraploid populations to be 0.88 (±0.09, 95% C.I.). Clearly the genetic system alone can satisfactorily account for the large increase in genetic variation of these autotetraploid populations. The details of our statistical analysis is presented in the APPENDIX A.

We note that our calculations do not necessarily support the contention that the observed genetic variation is neutral. Indeed models of natural selection can predict heterozygosity in diploid populations comparable to neutral models (GILLESPIE 1979). It is likely that an autotetraploid version of GILLESPIE's model would predict an increase in standing genetic variation similar to our neutral model.

DISCUSSION

Recent studies have indicated that tetrasomic inheritance may be much more common in plants than previously realized. As such it is important to understand how differences in the genetic systems of plants can influence the amount of genetic variation maintained in populations. We have shown that the so called "gametic homozygosity" is a natural measure of genetic variation for populations of any degree of ploidy. The gametic homozygosity appropriately weighs the contributions of the various kinds of genotypes to genetic variation within populations, and provides a unified approach for comparing genetic variation between populations of different ploidy.

We have computed the equilibrium gametic homozygosity under mutation and random drift for plants that randomly self and for those that do not self. The theoretical results reveal that the equilibrium gametic homozygosity is lower in autotetraploid populations than in diploid populations of equivalent size. This reduction in homozygosity is entirely due to the difference in the genetic systems between diploid and autotetraploids. It was thus possible to demonstrate that the apparent increase in heterozygosity observed in autotetraploid varieties of the diploid angiosperm *T. menziesii* is entirely consistent with our expectations based on tetrasomic inheritance.

This work has been supported by National Science Foundation (NSF) grants BSR-8700680 and INT-9014242 to M.E.M., National Institute of Health grant GM34303 to L.D.M and NSF grant BSR-8516721 to D.E.S. We thank J. N. THOMPSON, THOMAS NAGYLAKI and three anonymous reviewers for their helpful comments. We especially thank A. GALLAIS for his comments on the manuscript.

LITERATURE CITED

BARBER, N. H., 1970 Hybridization and the evolution of plants. Taxon **19:** 154–160.

BENNETT, J. H., 1953 Linkage in hexasomic inheritance. Heredity **7:** 265–284.

BENNETT, J. H., 1954 Panmixia with tetrasomic and hexasomic inheritance. Genetics **39:** 150–158.

BENNETT, J. H., 1968 Mixed self- and cross-fertilization in a tetrasomic species. Biometrics **24:** 485–500.

BENNETT, J. H., 1976 Expectations for inbreeding depression on self-fertilization of tetraploids. Biometrics **32:** 449–452.

BUSBICE, T. H., and C. P. WILSIE, 1966 Inbreeding depression and heterosis in autotetraploids with application to *Medicago sativa* L. Euphytica **15:** 52–67.

CLAUSEN, J., D. D. KECK and W. H. HIESEY, 1945 Experimental studies on the nature of plant species. II. Plant evolution through amphidiploidy and autopolyploidy with examples from the Madiinae. Carnegie Inst. Wash. Publ. No. **564:** 1–174.

CRAWFORD, D. J., 1985 Electrophoretic data and plant speciation. Syst. Bot. **10:** 405–416.

CROW, J. F., 1954 Random mating with linkage in polysomics. Am. Nat. **88:** 431–434.

CROW, J. F., and M. KIMURA, 1970 *An Introduction to Population Genetics Theory.* Harper & Row, New York.

DEMARLY, Y., 1963 The genetics of tetraploids and plant breeding. Ann. Amelior. Plant. **13:** 307–400.

EHRENDORFER, F., 1980 Polyploidy and distribution, pp. 45–60 in *Polyploidy–Biological Relevance*, edited by W. H. LEWIS. Plenum Press, New York.

ELANDT-JOHNSON, R. C., 1967 Equilibrium conditions in polysomic inheritance for a panmictic population. Bull. Math. Biophy. **29:** 437–449.

EWENS, W. J., 1979 Mathematical Population Genetics. Springer-Verlag, Berlin.

EWENS, W. J., and J. GILLESPIE, 1974 Some simulation results for the neutral alleles model, with interpretations. Theor. Popul. Biol. **6:** 35–57.

FISHER, R. A., 1943 Allowance for double reduction in the calculation of genotypic frequencies with polysomic inheritance. Ann. Eugen. **12:** 169–171.

FISHER, R. A., 1947 The theory of linkage in polysomic inheritance. Phil. Trans. R. Soc. B **233:** 55–87.

FISHER, R. A., 1962 Enumeration and classification in polysomic inheritance. J. Theor. Biol. **2:** 309–311.

FOWLER, N. L., and D. A. LEVIN, 1984 Ecological constraints on the establishment of a novel polyploid in competition with its diploid progenitor. Am. Nat. **124:** 703–711.

GALLAIS, A., 1967 Moyennes des populations tétraploïdes. Ann. Amelior. Plant. **17:** 215–228.

GALLAIS, A., 1969 Interactions between alleles and their variability in autotetraploid cross-fertilized plants. Genet. Agrar. **23:** 312–323.

GANTMACHER, F. R., 1960 *Matrix Theory*, Vol. II. Chelsea, New York.

GEIRINGER, H., 1949a Contribution to the linkage theory of autopolyploids: I. Bull. Math. Biophys. **11:** 59–82.

GEIRINGER, H., 1949b Contribution to the linkage theory of autopolyploids: II. Bull. Math. Biophys. **11:** 197–219.

GEIRINGER, H., 1949c Chromatid segregation of tetraploids and hexaploids. Genetics **34:** 665–684.

GILLESPIE, J. H., 1979 Molecular evolution and polymorphism in a random environment. Genetics **93:** 737–754.

GLENDINNING, D. R., 1989 Some aspects of autotetraploid population dynamics. Theor. Appl. Genet. **78:** 233–242.

GRANT, V., 1981 *Plant Speciation.* Columbia University Press, New York.

GRIFFING, B., 1957 Linkage in trisomic inheritance. Heredity **11:** 67–92.

GUY, P., 1970 Genotype structure of an autotetraploid hybrid multiplied panmictically. I. Theoretical expression of the principal parameters and genotypic structures of a hybrid plant and a hybrid population. Ann. Amelior. Plant. **20:** 263–295.

GUY, P., 1972 Genotype structure of an autotetraploid hybrid multiplied panmictically II. Elements for generalization. Annales Amelior. Plantes **22:** 239–261.

GUY, P., and G. GRENIER, 1967 Structure and evolution in the course of generations of multiplication under panmixia in an autotetraploid hybrid. Ann. Amelior. Plant. **17:** 277–290.

HALDANE, J. B. S., 1930 Theoretical genetics of autopolyploids. J. Genet. **22:** 359–372.

KEMPTHORNE, O., 1955 The correlation between relatives in a simple autotetraploid population. Genetics **40:** 168–174.

KEMPTHORNE, O., 1957 *An Introduction to Genetic Statistics.* John Wiley & Sons, New York.

KILLICK, R., 1971 The biochemical genetics of autotetraploids I. Generations derived from a cross between two pure lines. Heredity **27:** 331–346.

LEVIN, D. A., 1983 Polyploidy and novelty in flowering plants. Amer. Natur. **122:** 1–25.

LEWIS, W. H., 1980 Polyploidy in species populations, pp. 103–144 in *Polyploidy–Biological Relevance*, edited by W. H. LEWIS. Plenum Press, New York.

LI, C. C., 1955 *Population Genetics.* The University of Chicago Press, Chicago.

LI, C. C., 1957 The genetic variance of autotetraploids with two alleles. Genetics **42:** 583–592.

MALÉCOT, G., 1966 *Probabilités et Hérédité.* Presses Universitaires de France.

MATHER, K., 1935 Reductional and equational separation of the chromosomes in bivalents and multivalents. J. Genet. **30:** 53–78.

MATHER, K., 1936 Segregation and linkage in autotetraploids. J. Genet. **32:** 287–314.

MCCONNELL, G., and J. L. FYFE, 1975 Mixed selfing and random mating with polysomic inheritance. Heredity **34:** 271–272.

MUELLER, L. D., and F. J. AYALA, 1982 Estimation and interpretation of genetic distance in empirical studies. Genet. Res. Camb. **40:** 127–137.

MULLER, H. J., 1914 A new mode of segregation in Gregory's tetraploid primulas. Am. Nat. **48:** 508–512.

NEI, M., and A. K. ROYCHOUDHURY, 1974 Sampling variances of heterozygosity and genetic distances. Biometrics **76:** 379–390.

PARSONS, P. A., 1959 Equilibria in auto-tetraploids under natural selection for a simplified model of viabilities. Biometrics **15:** 20–29.

RICE, A., and J. W. DUDLEY, 1974 Gene effects responsible for inbreeding depression in autotetraploid Maize. Crop Sci. **14:** 390–393.

RIESEBERG, L. H., and M. F. DOYLE, 1989 Tetrasomic segregation in the naturally occurring autotetraploid *Allium nevii* (Alliaceae). Hereditas **111:** 31–36.

ROWE, D. E., 1986 Effects and control of genetic drift in the autotetraploid population. Crop Sci. **26:** 89–92.

SAMUEL, R., W. PINSKER and F. EHRENDORFER, 1990 Allozyme polymorphism in diploid and polyploid populations of *Galium*. Heredity **65:** 369–378.

SEYFFERT, W., 1959 Theoretische Untersuchungen über die Zusammensetzung tetrasomer Populationen II. Selbstbefruchtung. Z. Vererbungsl. **90:** 356–374.

SEYFFERT, W., 1960 Theoretische Untersuchungen über die Zusammensetzung tetrasomer Populationen I. Panmixie. Biomet. Z. **2:** 1–44.

SOLTIS, D. E., and L. H. RIESEBERG, 1986 Autopolyploidy in Tolmeia menziesii (Saxifragaceae): genetic insights from enzyme electrophoresis. Am. J. Bot. **73:** 310–318.

SOLTIS, D. E., and P. S. SOLTIS, 1989a Genetic consequences of autopolyploidy in *Tolmeia menziesii* (Saxifragaceae). Evolution **43:** 586–594.

SOLTIS, D. E., and P. S. SOLTIS, 1989b Tetrasomic inheritance in *Heuchera microantha* (Saxifragaceae). J. Hered. **80:** 123–126.

STEBBINS, G. L., 1947 Types of polyploids: their classification and significance. Adv. Genet. **1:** 403–429.

STEBBINS, G. L., 1950 *Variation and Evolution in Plants.* Columbia University Press, New York.

STEBBINS, G. L., 1980 Polyploidy in plants: unresolved problems and prospects, pp. 495–520 in *Polyploidy*, edited by W. H. Lewis. Plenum Press, New York.

WRIGHT, S., 1938 The distribution of gene frequencies in populations of polyploids. Proc. Nat. Acad. Sci. USA **24:** 372–377.

Communicating editor: B. S. WEIR

APPENDIX A

Estimation of *f*: Let the gametic nomozygosity at locus l in population i be $f_{l,i}$. A sample will consist of an enumeration of the number of genotypes in the five exhaustive classes according to (11). If we let #($\oplus$, $\varnothing$, $\ominus$, $\otimes$) represent the number of ($\oplus$, $\varnothing$, $\ominus$, $\otimes$) genotypes in the sample, with a similar notation for the other categories of genotypes, then we can estimate the frequencies in (11) using the sample genotypic frequencies $\pi_{1,1,1,1} = (\oplus, \varnothing, \ominus, \otimes)/N$, $\pi_{2,1,1} = (\oplus, \oplus, \ominus, \otimes)/N$, etc, where N is the total number of genotypes observed at locus l in population i. The estimate $\hat{f}_{l,i}$ is

then obtained by replacing the subscripted "π" frequencies of (12) with the sample frequencies. If there are a total of n_L loci sampled in n_K populations, then the mean gametic homozygosity for population i is estimated by,

$$\hat{f}_{\cdot,i} = \sum_{l=1}^{n_L} \frac{\hat{f}_{l,i}}{n_L}, \qquad \text{(A1)}$$

and hence the estimate of the mean gametic homozygosity over all populations is

$$\hat{f}_{\cdot,\cdot} = \sum_{i=1}^{n_K} \frac{\hat{f}_{\cdot,i}}{n_K}. \qquad \text{(A2)}$$

Variance of f: The variance of $\hat{f}_{\cdot,i}$ may be partitioned into between- and within-locus contributions as follows:

$$\mathrm{Var}(\hat{f}_{\cdot,i} = (V_{\cdot,i}^{(b)} + V_{\cdot,i}^{(w)})/n_L.$$

The estimated variance for population i is simply

$$\mathrm{Var}(\hat{f}_{\cdot,i}) = \sum_{l=1}^{n_L} (\hat{f}_{l,i} \geq \hat{f}_{\cdot,i})^2/n_L;$$

$(n_L - 1)$ should replace n_L in the denominator if an unbiased estimate is preferred. We now estimate the within-locus variance as the mean over loci of the variance due to multinomial sampling at each locus. In our notation this variance for a single locus is

$$\begin{aligned}\hat{\mathrm{Var}}(\hat{\pi}_4) &+ \frac{\hat{\mathrm{Var}}(\hat{\pi}_{3,1})}{4} + \frac{\hat{\mathrm{Var}}(\hat{\pi}_{2,2})}{9} + \frac{\hat{\mathrm{Var}}(\hat{\pi}_{2,1,1})}{36} \\ &+ 2\left[\frac{\hat{\mathrm{Cov}}(\hat{\pi}_4, \hat{\pi}_{3,1})}{2} + \frac{\hat{\mathrm{Cov}}(\hat{\pi}_4, \hat{\pi}_{2,2})}{3}\right. \\ &+ \frac{\hat{\mathrm{Cov}}(\hat{\pi}_4, \hat{\pi}_{2,1,1})}{6} \\ &+ \frac{\hat{\mathrm{Cov}}(\hat{\pi}_{3,1}, \hat{\pi}_{2,1,1})}{12} + \frac{\hat{\mathrm{Cov}}(\hat{\pi}_{3,1}, \hat{\pi}_{2,2})}{6} \\ &\left. + \frac{\hat{\mathrm{Cov}}(\hat{\pi}_{2,2}, \hat{\pi}_{2,1,1})}{18}\right],\end{aligned}$$

$$\hat{\mathrm{Var}}(\hat{\pi}_{1,1,1,1}) = (\hat{\pi}_{1,1,1,1})(1 - \hat{\pi}_{1,1,1,1})/N,$$

$$\hat{\mathrm{Cov}}(\hat{\pi}_{2,2}, \hat{\pi}_4) = -\hat{\pi}_{2,2}\hat{\pi}_4/N,$$

The calculation of $\mathrm{Var}(\hat{f}_{\cdot,\cdot})$ is complicated by the fact that the $\hat{f}_{\cdot,i}$ are not independent. This lack of independence arises from the common practice of sampling the same loci in all populations examined. Thus, loci are not sampled at random which implies that the $\hat{f}_{\cdot,i}$ are likely to be positively correlated. A similar effect has been described previously for genetic distance statistics (MUELLER and AYALA 1982) and applies also to heterozygosites averaged over populations. Thus

$$\hat{\mathrm{Var}}\hat{f}_{\cdot,\cdot} = \left[\sum_{i=1}^{n_K} (\hat{f}_{\cdot,i} - \hat{f}_{\cdot,\cdot})^2 + 2\sum_{i=1}^{n_K-1}\sum_{j=i+1}^{n_K} \hat{\mathrm{Cov}}(\hat{f}_{\cdot,i}, \hat{f}o_{\cdot,j})\right] \Big/ n_K^2,$$

where

$$\hat{\mathrm{Cov}}(\hat{f}_{\cdot,i}, \hat{f}_{\cdot,j}) = \sum_{l=1}^{n_L} (\hat{f}_{l,i} - \hat{f}_{\cdot,i})(\hat{f}_{l,j} - \hat{f}_{\cdot,j})/n_L^2$$

From the data analyzed in this study we note the following. For samples of 15 loci and about 30 individuals the intralocus contribution to variance is one or two orders of magnitude less than the between-locus contribution. This is a consequence of the loci generally falling in two classes: monomorphic or highly polymorphic. The best strategy for obtaining accurate estimates of gametic homozygosity is, therefore, to sample as many loci as possible. Similar recommendations apply to estimating heterozygosity in diploid populations (NEI and ROYCHOUDHURY 1974).

The correlation of heterozygosity between local diploid populations and of gametic homozygosity in autotetraploid populations is very high. Thus, about 63% of the variance of the mean heterozygosity of diploid Tolmiea is due to the correlation between local populations. Nearly 97% of the variance in mean gametic homozygosity in tetraploid Tolmeia is due to these correlations.

APPENDIX B

In this appendix we supply a justification for the generalization to higher ploidy levels of the equilibrium approximations given in (10). Consider a $2k$-ploid population, $k = 1, 2, \ldots$ with mutation at the constant rate u under the usual assumptions of the infinite alleles model. In the absence of selfing, we adapt the argument of KEMPTHORNE (1957, p. 91) to obtain the recursion relations for the probabilities of identity $f^{(2k)}$ and $g^{(2k)}$:

$$f_{t+1}^{(2k)} = (1-u)^2\left[\frac{k-1}{2k-1}f_t^{(2k)} + \frac{k}{2k-1}g_t^{(2k)}\right]$$

$$g_{t+1}^{(2k)} = (1-u)^2\left\{\frac{1}{N}\left[\frac{1}{2k} + \frac{2k-1}{2k}f_t^{(2k)}\right] + \left(1 - \frac{1}{N}\right)g_t^{(2k)}\right\}$$

With $\nu \equiv (1-u)^2$ we identify the equilibrium probability of identity as

$$\hat{f}^{(2k)} = \frac{\nu^2}{2N(1-\nu)[2k-1-(k-1)\nu] + \nu[2(2k-1)+(3-4k)\nu]}$$

which, when $N \gg 1$, $kNu^2 \ll 1$, gives the approximation

$$\hat{f}^{(2k)} \approx \frac{1}{1 + 2(2kNu)}.$$

This approximation generalizes (10). In the absence of mutation, we deduce from the recursion equations that the asymptotic rate of convergence to homozygosity is, to first order in N^{-1},

$$\lambda_0^{(2k)} \approx 1 - \frac{1}{2kN}.$$

This generalizes the approximations appearing in (4) and (7).

The approximations for $f^{(2k)}$ and $\lambda_0^{(2k)}$ agree with the corresponding quantities for random selfing, although the recursion equations and exact expressions must be suitably modified.