Allele Frequency Changes in Germplasm Collections:
Theoretical Considerations of Mutation and Fitness

J.E. Staub
USDA-ARS, Vegetable Crops Research Unit, Horticulture Department, University of Wisconsin, Madison, WI 53706

ADDITIONAL INDEX WORDS. germplasm management, genetic drift, computer simulation, dominance, partial dominance, genetic diversity

ABSTRACT. Often managerial resources (labor, space and budget) in germplasm banks are such that increases of accessions are made using relatively few plants (5 to 50). This type of management of genetic resources can render accessions susceptible to allelic frequency changes which in turn can change the relative fitness of individuals in the population. Mutation and environmental fluctuations (e.g., water stress) are two factors that can cause allelic frequency changes. Simplistic computer simulations (one locus, 2 alleles, p and q) were conducted to investigate the possible role of these factors on changes in allelic frequency and fitness in germplasm collections over 20 generations. Simulations evaluated the role of forward and reverse mutation, relative fitness with and without dominance in changing the population structure of small populations (n = 1000). A predictable decrease in the frequency of p was observed with increasing forward mutation rates of p. However, even with recurrent mutation the frequency of p was not dramatically altered by high mutation rates (10⁻⁷), suggesting that population size must be extremely low and/or mutation rate must be extremely high to have a demonstrable effect on allele frequency over short time periods (in this case 20 generations). Decreases in the frequency of p are slowed by reverse mutation. Where reverse mutation rate exceeds the forward mutation rate, allele frequency can increase. Allele frequency shifts were evident after 10 generations of selection for p, and did not change as dramatically between the tenth and twentieth generation when mutation rates are relatively high (0.07 to 0.1). Frequency changes in p are most dramatic in the first 10 generations when fitness values range between 0.4 to 1.0. However, such changes are not striking after the tenth generation. Partial dominance affects the rate of change in allele frequency over 20 generations, and dependent on the relative fitness of p which itself changes each generation according to the increase or decrease in p. Moreover, dramatic changes in allele frequency can occur regardless of dominance level between generations 1 to 10, especially if fitness values are high (>0.4).

For gene frequencies to change; mutation must first introduce the innovation that leads to genetic differences. The mere appearance of new genes is, however, no guarantee that they will either persist or prevail over other genes. Newly mutated genes in random mating natural populations ordinarily have a small chance of survival (Hartl, 1988). However, if a mutation has a positive survival value (fitness) in the population, then the gene and the frequency of the allele possessing increased fitness value will increase in the population.

In small artificial populations, such as germplasm collections, genetic drift in allelic frequencies from a given equilibrium can occur with mutation (Hartl, 1988). Allelic fixation in such populations depends in large part on its relative fitness. Allelic fitness is characterized by various genetic and environmental factors. For example, tight linkage of a neutral allele to an allele conferring high fitness may ensure its relative abundance in a small population. Likewise, if selection intensity for a particular trait conditioned by a single locus is great (i.e., short day length), then allelic fixation may be rapid.

Recently, I have thought more about the potential effects of mutation and the role of allelic fitness in small populations such as germplasm collections. Often managerial resources are such that increases of accessions are made using relatively few plants (5 to 50): Such increases are more affected by environmental fluctuations (e.g., differ-
ferential water stress) than are relatively large populations (>1000). The size of the increase population and the environment can change allelic frequencies, and thus can alter the accession's population structure (i.e., genetic drift). To investigate the possible role of mutation and allele fitness in germplasm collections, computer simulations were conducted using simplistic assumptions (i.e., diploid genome, no linkage, immigration, no crossing over). These simulations were designed to evaluate the effect of mutation (rate and frequency) and allele fitness on allelic frequencies produced in a single generation (mutation rate from A to a = u) can be expressed as (after Hartl, 1988) p = u + vq for A and q + u - vq for a.

The change in frequency of a from one generation to the next due to mutation can be expressed as a = uA0 (for generation 0), such that q1 = q0 + u0. When all other forces are ignored the mutation rate at equilibrium can be expressed as 0 = q = u - vq0 or u = vq0 or q = u / (u + v). Thus, an unstable equilibrium is one which drifts further from the original equilibrium with each successive generation. Such drifts in allelic frequencies are more likely in small populations (e.g., germplasm collections) under rigorous selection intensity than in large populations.

**THE ROLE OF RELATIVE FITNESS IN POPULATION STRUCTURE DETERMINATION:** Selection pressure results in an increase in the frequency of certain alleles and a decrease in the frequency of others. The proportionate reduction in contribution of a given genotype to the subsequent generation can be measured. Such a change in a population is defined by a selection coefficient which is a measure of the fitness of a genotype relative to a standard (e.g., mean) (Hartl, 1988). Selection operates successfully because 1) individual differences are characteristic of natural populations; 2) these differences are determined by heritable (genetic) factors, and 3) traits observed in more fit individuals will occur in subsequent generations in greater frequency.

The word fitness requires special consideration. There is often confusion between the words fittest and fitness. The word fittest assumes a comparative relationship (better or fitter), while fitness is an absolute term (i.e., the relative success in leaving offspring surviving to reproductive age). Whenever genotypes differ in fitness, natural selection can cause changes in population structure (variation and genetic diversity) (Crow and Nagylaki, 1976). There are various factors which...
can contribute to average fitness \( \bar{w} \) in a population. Over time these factors have led to changes in population dynamics (Demetrius, 1977).

The genetic relationship between fitness and selection can be understood through observing changes in allele frequency. Consider for a moment two alleles \( (A, a; A \) is dominant over \( a) \) at one locus, such that one of the homozygotes \( (AA) \) has the highest fitness. Further, assume that this allele will be fixed in a population given no interference from other factors. Let \( p \) equal the frequency of \( A \) and \( q \) the frequency of \( a \), and let the fitness of \( AA, Aa, \) and \( aa \) be expressed as \( 1, 1, \) and \( 1 - w \), respectively: The change in allele frequency after one generation \( (\Delta) \) can be expressed in the following manner (modified from Futuyma, 1986):

\[
\Delta p = \frac{wpq^2}{1 - wq^2}
\]

\[\text{[Eq. 1]}\]

\[
\Delta q = -\frac{wpq^2}{1 - wq^2}
\]

\[\text{[Eq. 2]}\]

where \( w \) represents the fitness selection strength.

Clearly when \( p \) is positive, the frequency of \( A \) (advantageous allele) increases. It increases proportional to \( w \) and the frequency in which \( q \) is being replaced. It turns out that \( p \) will be greatest when \( p = 1/3 \), but this rise will be relatively slow (compared to a common allele in the population) if \( p \) has just arisen from a mutational event. On the other hand, if allele \( A \) is common in the population and if the environment changes such that the genotype \( AA \) is better adapted than \( aa \), then the rate of change in the population will be most rapid if \( A \) is at an intermediate frequency rather than a low frequency (i.e., rare allele).

The question then arises: “When is equilibrium \( (\Delta = 0) \) reached and what is the rate of approach to equilibrium?” The rate of approach to equilibrium depends upon the frequency of the advantageous allele and whether it is dominant. In general, for a dominant allele the predicted equilibrium frequency is \( \hat{p} = 1 \). In this case the allele is fixed and the recessive allele is eliminated. Dominant advantageous alleles increase more rapidly than advantageous recessive alleles. As a deleterious recessive allele becomes infrequent in a population, \( q \) is proportionately smaller and therefore the rate of approach to equilibrium is slower. Moreover, because allele \( a \) is masked in the heterozygous condition \( (Aa) \), elimination of \( a \) from the population almost never occurs. This means that any population will have deleterious alleles in the gene pool. When the deleterious condition is dominant, dominant alleles are quickly eliminated from the population because they are not protected by a contrasting allele \( (a) \) at the same locus.

Comprehensive empirical evaluations of the effects of even simple genetic variables (e.g., gene action) on population structure are complex and resource demanding. Simple, independent tests of the effect of dominance on allelic frequency and its potential affect on frequency changes over time can be performed by computer simulation. Such simulations provide the vehicle for the assessments described herein. The simulations presented are not meant to be comprehensive, but rather provide a first approximation for clarifying the significance of mutation and fitness in germplasm collections.

**Materials and methods**

**Allelic frequency change by mutation.** In order to estimate changes in allele frequency with mutation pressure consider the case of two alleles \( A \) and \( a \) at a locus. The allele frequencies of \( A \) and \( a \) in generation \( t \) are designated as \( p_t \) and \( q_t \), respectively (where \( q_t = 1 - p_t \)). There are two possibilities for the origin of allele \( A \) in generation \( t \):

1) Allele \( A \) could have existed in generation \( t - 1 \) and could not have mutated to \( a \) (probability \( 1 - u \)), or
2) Allele \( a \) could have mutated to \( A \) (probability \( v \)).

This relationship at time \( t \) can be expressed as follows (after Hartl, 1988):

\[
p_t = p_{t-1} (1 - u) + (1 - p_{t-1})u
\]

\[\text{[Eq. 3]}\]

such that by substituting a series of mathematical identities, \( p_t \) can be solved in terms of \( p_0 \):

\[
p_t = \frac{v}{u+v} + \left( p_0 - \frac{v}{u+v} \right) (1 - u - v)^t
\]

\[\text{[Eq. 4]}\]

The meaning of \( p_t \) can be expressed either in short or long term generation times. In the case of a small number of generations (<100), such as that found in currently operating germplasm collections created in the last 100 years with an average
of one cycle of increase per year (1 x 100), the term (1 - u - v) is about equal to 1 - t(u + v) (Staub, 1994). This refinement was used in conjunction with Eq. 4 to develop the algorithms used to create the simulated effects of mutation over 20 generations.

Initial allelic frequency of the base population (generation 0) was set at Hardy-Weinberg equilibrium (p and q = 0.5). Reverse mutation was considered in the model at the rate of 0.01 and 0.001 in each generation to determine allele frequency change for comparative analysis.

Effects of Relative Fitness on Allele Frequency.
The average fitness \( \bar{w} \) of an individual in a population can be expressed mathematically as follows (modified from Futuyma, 1986):

Given: Genotypes AA, Aa, and aa
Fitness \( w \) of AA = 1, Aa = 1 - \( w \), aa = 1 - \( w \).
Frequency of AA = \( p^2 \), Aa = 2pq, aa = \( q^2 \).

Then

\[
\bar{w} = p^2 + 2pq (1 - w) + q^2 (1 - w) \quad [\text{Eq. 5}]
\]

\[
= 1 - 2w, q + 2w, q^2 - w, q^2 \quad [\text{Eq. 6}]
\]

The allele (a) frequency change \( q \) can be expressed as

\[
\Delta q = q' - q = pq(2w, q - w, - w, q^2) \quad [\text{Eq. 7}]
\]

where \( q' \) is the frequency of \( q \) in the next generation. This expression is adequate to describe the general genetic change (1 locus, 2 alleles) in a single generation by natural selection when it acts to affect viability where fitness is constant. Equations 5 and 6 were used in the algorithm for calculation of allele frequency and frequency with concomitant fitness change over 20 generations simulating response of a diploid species without the effects of mutation, immigration, or linkage (Staub, 1994). The change in allele frequency \( p \) was calculated as \( 1 - q \) change in \( q \) (Eq. [7]), and relative fitness was evaluated at 0.1 to 1.0 in tenth of degree increments. Initial allelic frequency of the base population (generation 0) was set at Hardy-Weinberg equilibrium (\( p \) and \( q = 0.5 \)).

Fitness with Dominance. The differential af-
fects of fitness with dominance (values of 0.2, 0.5, and 0.8) and partial dominance (Aa/aa = 0.2/0.1, 0.2/0.05, 0.2/0.025, 0.1/0.05, 0.1/0.025), and the calculation of average dominance over 20 generations were simulated (Staub, 1994) using Eq. 8 in an iterative algorithm (after Hartl, 1988).

$$\Delta q = \frac{-wq[p + d(p - q)]}{1 - wq(2dp + q)}$$  \[Eq. 8\]

where, d denotes the degree of dominance with respect to fitness, and fitness of aa and Aa is 1-w and 1-dw, respectively. Thus, if allele A is dominant, then d = 0, if allele a is dominant, then d = 1, and if neither allele A or a is dominant, then d = 1/2. The degree of dominance (d) affects the rapidity of allelic change, because q increases with increases in d. Moreover, in the case of a rare deleterious allele (q < 1/2), the more dominant it is, the more rapidly that allele will decrease in frequency. Initial allelic frequency of the base population (generation 0) was set at Hardy-Weinberg equilibrium (p and q = 0.5).

**Results and discussion**

Natural selection may act to alter the population in two major ways: 1) it may eliminate all but the least fit allele(s), or 2) it may maintain stable genetic polymorphisms by either balancing alternatively fit alleles or by operating in conjunction with mutation and/or gene flow. Selection under semi-restricted conditions such as cage increase under cultivation can operate to change allele frequency by eliminating least fit alleles (e.g., photoperiod dependent gene expression) or affecting allele frequency through mutation and/or fitness. Changes in allele frequencies are affected (limited) by the amount of inherent genetic variability available, the generation time of the species, the strength of the selection coefficient, the degree of dominance of the trait(s) being selected for, the initial frequency of the advantageous allele in the population, and the intragenomic interactions present in the species. The effects of mutation and fitness on population structure using computer simulation are summarized below.

**Allelic frequency change by mutation.** The case of mutation without allelic reversion is depicted in Figure 1 (panels A-1 and A-2). A predictable decrease in the frequency of p was observed.
with increasing mutation rate. The mutation rate of 0.0001 was the lowest rate evaluated, but it is still considered extremely high in natural populations. Nevertheless, even with recurrent mutation at a rate equal to $10^{-3}$, the frequency of $p$ was not dramatically altered (panel A-I; change from 0.50 to 0.43 at a mutation rate = 0.0001), suggesting that population size must be extremely low and/or the mutation rate must be extremely high to have a demonstrable effect on allele frequency over short time periods (in this case 20 generations). This is

Figure 3. Allele frequency, its change and relative fitness over time ($t$) under different degrees of dominance given original Hardy-Weinberg equilibrium ($p = 0.5$).
important since germplasm collections are relatively new (created in the last 100 years) and increase population sizes can vary depending of the species.

The affect of reverse mutation on allele frequency change is considered in Figure 1, panels B-1 through C-2. Predictably, decreases in the frequency of \( p \) are slowed by reverse mutation (compare A-1 vs. B-1 and C-1, and A-2 vs. B-2 and C-2). Where reverse mutation rate exceeds the forward mutation rate allele frequency can increase (panels B-1 and B-2). Allele frequency shifts are detectable after 10 generations and do not change as dramatically between the tenth and twentieth generation when mutation rates are relatively high (0.07 to 0.1) (panels A-2, B-2, C-2). These simulations suggest that stable equilibriums are reached where heterozygous individuals buffer the effects of mutation over time. Thus, even at high mutation rates not usually encountered in natural populations or germplasm collections, an allele may be reduced to a frequency where it is designated as rare (frequency <0.1) but still continue to be present a population even with continuing mutation pressure.

**Effects of relative fitness on allele frequency.**

Allele frequency can be altered if allelic fitness is dominant in a random mating population (Figure 2, panels A and B). Allele frequency change is dependent on allelic fitness value; in the case of these simulations 0.1 to 1.0. Frequency changes in \( p \) are most dramatic in the first 10 generations when fitness values range between 0.4 to 1.0 (panel A). Relative frequency changes are not striking between the tenth and twentieth generations (panel B).

Partial dominance affects the rate of change in allele frequency over time (Figure 2, panels C and D). This rate of change is dependent on the relative fitness of \( p \) which itself changes each generation according to the increase or decrease in \( p \) (panel E). Ingermplasm collections where population size is small, these changes in fitness can be dramatic. With relatively low fitness, dramatic changes in allele frequency may not be observable during seed increase in early generations (1 to 10 generations), but may be significant in later generations (panels C and D).

**Fitness with dominance.**

Allelic values for dominance can range between 0 (no dominance) to 1.0 (complete dominance). Dominant gene action can interact with allelic fitness values to change allele frequency and average fitness over generations. Simulation suggests that dramatic changes in allele frequency can occur regardless of dominance level between generations 1 to 10, especially if fitness values are high (>0.4) (Figure 3). Change in allele frequency is less dramatic after the tenth generation. Likewise, relative fitness is increased dramatically over 10 generations of selection pressure, but is less affected thereafter, regardless of dominance value.

Factors affecting the fitness of an allele are complex and can have a dramatic affect. However, the computer simulation data presented herein suggests that although allelic fitness can be altered by selection, allelic frequency changes may not be dramatic over short time periods (1 to 10 generations) if genetic factors such as partial dominance are operating or where selected alleles have low relative fitness values. In order to maintain a collection’s genetic diversity, however, the increase of accessions requires the management of important factors (e.g., water and heat stress) that can affect fitness.

**Literature cited**


