The Maintenance of Single-locus Polymorphism by Maternal Selection

Hamish G. Spencer and Kai X. Chiew

Allan Wilson Centre, Department of Zoology, University of Otago
Running Title: Maternal Selection and Polymorphism

Keywords: Gavrilets model, Multiple alleles, Natural selection, Mathematical model

Corresponding Author: Hamish G. Spencer
Allan Wilson Centre
Department of Zoology
University of Otago
P.O. Box 56
Dunedin 9054
New Zealand

Email: hamish.spencer@otago.ac.nz
Phone: +64-3-479 7981
Fax: +64-3-479 7584
Abstract

Population geneticists have long been interested in the ability of natural selection to maintain the levels of standing variation observed in natural populations. Here we study the polymorphism-maintaining properties of maternal selection, in which the fitness of an individual is a function of its own and its mother’s genotype. Using a model proposed by Gavrilets, we first estimate the proportion of parameter/state space that preserves allelic variation, before investigating the construction of polymorphism over time through the joint action of mutation and selection. These two methods, the “parameter-space” and “constructionist” approaches, respectively, enable us to draw some general conclusions. We argue that, even though the proportion of parameter-state space allowing multi-allele polymorphism is greater under maternal selection than under the standard model of constant viability selection, the former is, in fact, less likely to maintain large numbers of alleles. Nevertheless, variation that is balanced by maternal selection is likely to show elements of heterozygous advantage and be resistant to depletion by genetic drift. We observe that the population mean fitness frequently decreases after the successful invasion of a new mutation, but such declines are usually temporary.
One of the long-standing goals of evolutionary genetics is to understand the relative importance of the various processes that shape genetic variation in natural populations (Lewontin 1974; Kimura 1984; Mitchell-Olds et al. 2007; Hahn 2008; Leffler et al. 2012; Delph and Kelly 2014; Gao et al. 2015). Especially after the advent of gel electrophoresis in the 1970s had revealed large levels of variation in many natural populations, scientists have been interested in the part played by natural selection. Does it actively shape standing variation (as in the balance school; Lewontin 1974), or is its role confined to removing deleterious mutations (as in the neutral hypothesis; Kimura 1984)?

The ascendency of the neutralist view in the 1980s was, in part, due to theoretical problems underlying the balance school. Most critically, Monte-Carlo simulation studies showed that the simplest form of selection, constant viability differences among genotypes, were only able to maintain more than 3-4 alleles for an extremely restricted set of fitness values (Lewontin et al. 1978; Gillespie 1977). These results have been extended to models of several other types of viability selection, such as sexually differential selection (Marks and Ptak 2001), frequency-dependent selection (Trotter and Spencer 2007) and spatially heterogeneous selection (Star et al. 2007a), as well as constant fertility selection (Clark and Feldman 1986).

Although such results were often interpreted as showing that selection could not maintain large numbers of alleles (e.g., Árnason 1982; Kimura 1984; Keith et al. 1985; Nevo 2001; Charlesworth 2006; Gloria-Soria et al. 2012; Ejsmond et al. 2014), such a conclusion is logically flawed. Quite generally, the size of some region of parameter space affords little or no information about the likelihood that real parameters lie in that region. Indeed, organic evolution is well recognized to produce systems with exceptionally unusual outcomes (and hence also parameters). In a response to this realization, Spencer and Marks (1988, 1992; Marks and Spencer 1991) showed that a simple system combining recurrent mutation and
natural selection iterated towards those parts of viability-parameter space in which numerous alleles could be maintained. Polymorphism was naturally “constructed.” This result, using the so-called “constructionist approach,” holds for numerous models of selection and mutation (Marks and Ptak 2001; Star et al. 2007b, 2008; Trotter and Spencer 2008, 2013). We note that this approach has significant commonalities with the later approach of adaptive dynamics (Waxman and Gavrilets 2005; but see also Spencer and Feldman 2005).

In this paper, we investigate the ability of maternal selection to maintain genetic variation. Many organisms’ phenotypes are profoundly affected by their maternal environment (Mousseau and Fox 1998; Mousseau 2006; Wolf and Cheverud 2012), but population-genetic models of the selective consequences of these effects have been developed only in the last twenty years (Gavrilets 1998; Wade 1998; Spencer 2003; Spencer et al. 2006). In our investigations we first determine the size of parameter/state space that leads to polymorphism under Gavrilets’s (1998) model of maternal selection, before using the constructionist approach to assess the likelihood of polymorphism under this model.

Model

We start by generalizing the 2-allele model of maternal selection of Gavrilets (1998; see also Spencer 2003). Consider a single locus with $n$ alleles, $A_1, A_2, \ldots, A_n$, at respective frequencies $p_1, p_2, \ldots, p_n$ (with $\sum_{i=1}^{n} p_i = 1$), in a randomly mating, dioecious population, in which the effects of mutation and genetic drift are negligible.
In order to model maternal selection we must parameterize each combination of offspring-mother genotypes (Cheverud and Wolf 2009; Wolf and Wade 2009; Wolf and Cheverud 2012), say an $A_iA_j$ offspring with an $A_kA_l$ mother, where $i = k$ or $l$, or $j = k$ or $l$, or both) by a fitness parameter, $w_{ijkl}$. As in most models of selection, only the relative size of these fitnesses matters, and so we assume that $0 \leq w_{ijkl} \leq 1$. It is critical to ensure the various combinations are correctly enumerated. For $n$ alleles, there are $n$ possible homozygotes and $\binom{n}{2}$ different heterozygotes. Each homozygote, say $A_iA_i$, can have $n$ possible different mothers, $A_iA_1, A_iA_2, \ldots, A_iA_n$, which makes for $n^2$ combinations. Each heterozygote, say $A_iA_j$ ($i \neq j$), can have 2 sorts of homozygous mothers, $A_iA_i$ and $A_jA_j$, 1 identical (heterozygous) mother, $A_iA_j$, and $2(n-2)$ non-identical heterozygous mothers, $A_iA_k$ and $A_jA_k$ (where $k \neq i, j$ can take on $n-2$ different values). Hence the number of heterozygote-offspring and mother combinations is

$$\binom{n}{2}\left(2+1+2(n-2)\right)=\frac{1}{2}n(n-1)(2n-1)$$

and so, the number of distinct fitness parameters is $C(n) = n^2 + \frac{1}{2}n(n-1)(2n-1)$. These numbers rapidly become large: for $n = 1, 2, 3, 4$ and $5$, we require, respectively, 1, 7, 24, 58 and 115 distinct fitness parameters.

Accordingly, there are five distinct classes of offspring-mother combinations: Homozygotes with homozygous mothers, homozygotes with heterozygous mothers, heterozygotes with homozygous mothers, heterozygotes with identical heterozygous mothers and heterozygotes with non-identical heterozygous mothers. The relative sizes of the fitnesses of these different classes might be expected to be differentially important in maintaining variation. For a given fitness set, we denote the mean of the fitnesses in each class by,
respectively, \( C_{iii}, C_{iiij}, C_{ijii}, C_{ijij} \) and \( C_{ijik} \). Note that for smaller values of \( n \) some of these classes are empty.

Maternal selection requires that we keep track of genotype frequencies. Let \( x_{ij} (= x_{ji}) \) be the frequency of genotype \( A_iA_j \) (so that \( \sum_{i=1}^{n} \left( x_{ij} + \sum_{j=i+1}^{n} x_{ij} \right) = 1 \)) and so \( p_i = x_{ii} + \frac{1}{2} \sum_{j=i+1}^{n} x_{ij} \). \( A_iA_i \) homozygotes always receive an \( A_i \) allele from their father with frequency \( p_i \). They also inherit an \( A_i \) allele from their mother, either an \( A_iA_i \) homozygote, with frequency \( x_{ii} \), or an \( A_iA_k \) heterozygote (for any \( k \neq i \)), with frequency \( \frac{1}{2} x_{ik} \). These two types of mothers confer different fitnesses on their offspring, \( w_{iii} \) and \( w_{iiik} \), respectively. Hence, the iteration for the homozygote frequencies is given by

\[
\overline{w}x'_{ii} = p_i \left( w_{iii}x_{ii} + \frac{1}{2} \sum_{k=i+1}^{n} w_{iik}x_{ik} \right),
\]

(2)

An \( A_iA_j \) heterozygote can receive its \( A_i \) allele from its father and its \( A_j \) from its mother (who, as above, may be homozygous or heterozygous for the \( A_j \) allele), or vice versa, giving rise to the two terms on the right-hand side in the following equation. Using the same logic as above, we find that the iteration for heterozygote frequencies is given by

\[
\overline{w}x'_{ij} = p_i \left( w_{iji}x_{ij} + \frac{1}{2} \sum_{k=i+1}^{n} w_{ijk}x_{jk} \right) + p_j \left( w_{jii}x_{ji} + \frac{1}{2} \sum_{k=i+1}^{n} w_{jik}x_{ik} \right),
\]

(3)

for \( i \neq j \). In these equations, \( \overline{w} \) is the mean fitness, the sum of the right-hand sides of Equations (2) and (3) for \( i = 1, 2, 3, \ldots, n \) and \( j > i \). This formulation ensures that the genotype frequencies in the following generation also sum to one:

\[
\sum_{i=1}^{n} \left( x'_{ii} + \sum_{j=i+1}^{n} x'_{ij} \right) = 1.
\]
Estimating Potential for Polymorphism

For a fixed number of alleles, \( n = 2, 3, \ldots, 6 \), we estimated the proportion of viability-parameter space and initial genotype-frequency state-space that maintained all \( n \) alleles. This proportion we call the “potential” for polymorphism and the method the “parameter-space approach,” following Trotter and Spencer (2007). We did so by first generating \( 10^5 \) random sets of \( C(n) \) fitnesses. As in Lewontin et al. (1978), these fitnesses were drawn from the uniform distribution between zero and one, \( U[0, 1] \). We do not wish to imply that natural fitnesses are uniformly distributed; rather, using \( U[0, 1] \) allows us to model parameter space as a \( C(n) \)-dimensional cube, which provides for a robust method for estimating the potential for polymorphism (see also Trotter and Spencer 2007).

Under constant viability selection, a given fitness set has at most one fully polymorphic equilibrium (Bürger 2000), but such a property does not hold under maternal selection (Gavrilets 1998; Spencer 2003), and, indeed, allele-frequency cycling is possible but rare (Spencer 2003). In other words, iteration of Equations (2) and (3) from different (non-trivial) initial genotype frequencies does not necessarily lead to a unique genotype-frequency vector even if all \( n \) alleles are retained. Rather, for each fitness set we determined the fraction of initial genotype frequencies that lead to an \( n \)-allele polymorphism (either an equilibrium or the cycling of \( n \) alleles). Thus, for each fitness set, we iterated from 100 randomly selected initial genotype frequencies and recorded the proportion that retained \( n \) alleles. If all 100 initial genotype frequencies for a particular fitness set retained all \( n \) alleles, we called this fitness set an “always” fitness set; if only some iterations retained all \( n \) alleles, the fitness set was described as a “sometimes” fitness set. The initial genotype frequencies
were derived from randomly generated allele-frequencies found using the broken-stick method (Holst 1980) and assuming Hardy-Weinberg genotype proportions.

The potential for polymorphism for a given $n$ is thus given by the proportion of initial genotype-frequency vectors over all $10^5$ fitness sets that iterated to an $n$-allele equilibrium or cycled with $n$ alleles. This value is equal to the proportion of “always” fitness sets plus the proportion of iterations over all “sometimes” fitness sets that maintained $n$ alleles.

### Constructing Polymorphisms

Following the approach outlined in Spencer and Marks (1988), we began our constructionist simulations with a monomorphism ($n = 1$) and assumed $w_{1111} = 0.5$. Every generation, a new mutant allele, $A_{n+1}$ was added to the existing $n$-allele system at a frequency of $p_{n+1} = 10^{-5}$ and the frequency of one randomly chosen allele already present in the population was reduced by the same amount. The addition of $(n + 1)^{th}$ allele results in $n + 1$ new genotypes and hence $C(n+1) - C(n) = 3n^2 + 2n + 1$ new random fitnesses must be generated. These parameter values were drawn from U[0, 1]. Again, we do not wish to imply that fitnesses in natural populations have such a distribution, but doing so provides a more straightforward comparison with the results of other forms of selection.

We then iterated genotype (and hence allele) frequencies according to Equations (2) and (3). Any allele whose frequency fell below $10^{-5}$ was considered to be extinct and was removed from the system. Each generation we recorded the number and frequencies of all alleles, as well as the mean fitness. After $10^4$ generations each simulation was stopped and the fitness set of the extant alleles recorded.
We ran $10^3$ replicate simulations, differing only in the seed for the pseudo-random number generator. In all our simulations we used the pseudo-random number generator of Marsaglia et al. (1990).

**Results**

*Potential for Polymorphism*

The proportion of parameter/state space that affords full ($n$-allele) polymorphism is shown in Figure 1. Also plotted is the proportion of “always” fitness sets, and the corresponding results for constant viability selection from Lewontin et al. (1978) are shown for comparison. (Under constant viability selection, when an $n$-allele equilibrium is possible, all (non-trivial) initial genotype frequencies iterate to this equilibrium, and so there is just a single line.) Table 1, also shows the numbers of “always” and “sometimes” fitness sets for each value of $n$. The proportion of initial genotype frequencies that maintained all $n$ alleles for the latter type of fitness set declined with $n$ (Table 1).

Clearly, the potential for polymorphism declines with $n$, a result that holds for all forms of selection investigated to date. The potential for polymorphism under maternal selection is, however, greater than that under constant viability selection, whereas the proportion of fitness sets that always maintain all alleles is very similar for the two forms of selection.
Allele-frequency cycling was very rare, with 0, 65, 44, 12 and one cases for \( n = 2, 3, 4, 5 \) and 6, respectively, out of the \( 10^5 \) fitness sets for each value of \( n \).

**Constructing Polymorphisms**

The results of some representative constructionist simulations are shown in Figure 2. In each simulation, the number of alleles rapidly increases, but soon drops to \( \leq 5 \). Long periods of no change in this number mean that no new mutations have successfully invaded. In several cases, successful invasion is quickly followed by the extinction of one or more alleles; the successful mutation is driving out some of the standing variation. The mean fitness generally increases, but this increase is not monotonic. Significant changes (especially increases) in the mean fitness coincide with changes the numbers of alleles (i.e., successful invasions and extinctions).

The distribution of the number of alleles in each of the 1000 simulations after \( 10^4 \) generations is shown in Figure 3. Also shown is the distribution under constant viability selection (data from Spencer and Mark 1992). The modal number of alleles under maternal selection is just 2 (with a mean of 2.04), significantly lower than the modal number of 5 (and mean of 5.4) under viability selection. Thus, the larger potential for polymorphism under maternal selection (as compared to constant viability selection) is not realized in a larger number of alleles in constructed polymorphisms.

Constructed polymorphisms with larger numbers of alleles tended to have lower mean fitnesses (Fig. 4). Successful invasion by a mutant allele is more likely, everything else being equal, if the mean fitness is lower, and so polymorphisms with lower mean fitnesses are
likely to be invaded more easily, leading to greater numbers of alleles. Such dynamics could also explain why polymorphisms with a large number of alleles are rare: easy invasion may soon be followed by one or more extinctions, lowering the numbers of alleles.

Fig. 5 shows the mean fitnesses in each simulation for the different classes of offspring-mother combinations for each value of \( n \) for which there were at least 10 simulations (i.e., \( 1 \leq n \leq 4 \)). The pattern is consistent: heterozygous offspring tended to have higher fitnesses than homozygous offspring, differences that were accentuated if their mothers were also heterozygous (i.e., \( C_{i\,i\,i} < C_{i\,i\,j} < C_{i\,j\,i} < C_{i\,j\,j} \) and \( C_{i\,j\,i} \)).

It is also of interest to know about the values of the allele frequencies at Generation \( 10^4 \): Are these even or skewed? We can summarize information about the distribution using the sum of the squared distances from the centroid of the state space of allele frequencies,

\[
I = \sum_{i=1}^{n} \left( p_i - \frac{1}{n} \right)^2 .
\]

Large values of \( I \) indicate uneven allele frequencies; equal allele frequencies give \( I = 0 \). The distribution of the values of \( I \) for \( n = 2, 3 \) and \( 4 \) is shown in Fig. 6, along with the values expected in allele frequencies drawn at random under the broken-stick model (Holst 1980). For \( n = 2 \) we also plot the values of \( p_i \). It is clear that for \( n = 2 \) and \( 3 \) (but not \( 4 \)), constructed polymorphisms have allele frequencies that are more even than our random expectation. Also shown in Fig. 6 are histograms of the values of \( I \) under the parameter-space approach. For \( n = 2 \) and \( 3 \), these values are intermediate; for \( n = 4 \), they are more even than both the random and constructed frequencies.

**Discussion**
Our investigation of fitness-parameter space under maternal selection suggests that this important form of selection has a greater potential to maintain large numbers of alleles than does standard viability selection. Somewhat surprisingly, though, we also show that this potential is not realized when selection and mutation construct polymorphisms over time: maternal-selection simulations maintain fewer alleles than do those of constant viability selection. This pair of results reinforces the logical point that the size of some portion of parameter space is not informative about the likelihood of a system to lie in that portion.

The difference between viability and maternal selection in their potential to maintain variation is almost entirely due to the existence of offspring-mother fitness sets that cannot maintain all \( n \) alleles starting from all (non-trivial) allele frequencies (the “sometimes” fitness sets; Fig. 1; Table 1). Indeed, for \( n = 4 \) and 5, the majority of fitness sets that maintained all \( n \) alleles did so from only some initial allele frequencies. For such fitness sets there must exist one (or more) allele(s) that, when rare, will be eliminated but, when common, will iterate to a polymorphic equilibrium. If such a fitness set were generated in a constructionist simulation and that same allele were the new mutant, there would be no increase in the number of alleles. Thus this fitness set would count towards the ability of maternal selection to maintain greater levels of variation under the parameter-space approach than under the constructionist approach. This difference may explain the lower numbers of alleles found in our constructionist simulations.

Under both viability and maternal selection, the conditions for successful invasion of a new mutation become very restrictive as mean fitness increases and fewer alleles can invade (Fig. 2; see also Fig 2. in Spencer and Marks 1988). Nevertheless, this constraint is greater in the parameter-rich model of maternal selection because in this model, on average, more fitnesses must be large. Thus under maternal selection, fewer successful invasions occur
and the consequent number of alleles is lower. The maximum number of alleles we observed in our constructionist investigation was 5, which happened only twice in 1000 simulations. A negative correlation between the relative potential for polymorphism (compared to standard viability selection) and its realization have previously been found in investigations of sex-dependent viabilities (Marks and Ptak 2001).

Gavrilets (1998) and Spencer (2003) showed that, under maternal selection, the mean fitness of a population may sometimes decline. Most of our selected simulations in Fig. 2 exhibit such decreases, although they are short-lived and, consequently, over long periods of time mean fitness generally increases. Thus maternal selection, like constant viability selection, is likely to lead to improved adaptation.

Perhaps not surprisingly, mean fitness changes most often after the invasion of a new mutation, but this change need not be an increase. Nevertheless, a successful invasion is often followed by the extinction of one or more alleles from the standing variation, accompanied by significant increases in mean fitness. In only a minority of our Fig. 2 simulations does a decrease in fitness persist for more than a few generations.

The sorts of fitness sets that maintain polymorphism under maternal selection tend to show a generalized heterozygote advantage (Fig. 5). The class of offspring-mother combinations that had the highest mean fitness in our constructionist simulations was that of heterozygous offspring with identically heterozygous mothers ($C_{ijj}$). Interestingly, this mean was higher than that of heterozygotes with non-identical heterozygous mothers, with whom they shared just one allele. The lowest class fitness was that of homozygotes with (identical) homozygous mothers ($C_{iii}$).

The relatively even allele frequencies found using both parameter-space and constructionist approaches (Fig. 6) is in contrast to that under constant viability selection
(Spencer and Marks 1988). In finite populations, genetic drift is likely to reduce the number of alleles in the standing variation (Star and Spencer 2013). Everything else being equal, even equilibrium allele frequencies will be more resistant to extinction due to genetic drift and so maternal selection may play a more important role in maintaining variation in smaller populations than our results might, at first, suggest.

Acknowledgements

This research was funded by the Allan Wilson Centre and K.X.C. received a summer bursary from the Allan Wilson Centre. Two anonymous reviewers made several helpful suggestions.
Literature Cited


Table 1: Results from “Parameter-Space” Simulations

<table>
<thead>
<tr>
<th>n</th>
<th>No. of “Always” Fitness Sets (out of 100,000)</th>
<th>No. of “Sometimes” Fitness Sets (out of 100,000)</th>
<th>Proportion of Iterations Maintaining ( n ) Alleles for “Sometimes” Fitness Sets</th>
<th>Potential for Polymorphism</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>31,154</td>
<td>7,842</td>
<td>0.7068</td>
<td>0.36697</td>
</tr>
<tr>
<td>3</td>
<td>3,894</td>
<td>3,825</td>
<td>0.6814</td>
<td>0.06500</td>
</tr>
<tr>
<td>4</td>
<td>273</td>
<td>891</td>
<td>0.5769</td>
<td>0.00787</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>134</td>
<td>0.4663</td>
<td>0.00070</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>13</td>
<td>0.3192</td>
<td>0.00004</td>
</tr>
</tbody>
</table>
Figure Legends

Figure 1. Parameter/state space maintaining all $n$ alleles. The blue line with open circles shows the potential for polymorphism under maternal selection, i.e., proportion of parameter/state space for which all $n$ alleles are maintained for random fitness sets and initial genotype frequencies. The red line with filled circles shows the proportion of maternal fitness sets that always maintained all $n$ alleles, i.e., the proportion of “always” fitness sets. The corresponding values for constant viability selection (Lewontin et al. 1978) are shown in dashed black.

Figure 2. Representative simulations of the construction of polymorphism maintained by maternal selection. Upper (dashed red line) shows the mean fitness of the population, $\bar{w}$, and the lower (solid blue) line shows the numbers of alleles over Generations 0 to 10,000.

Figure 3. Bar chart of red bars showing the number of alleles at Generation $10^4$ in 1000 replicate simulations of the constructionist simulations of maternal selection. Also shown, for the sake of comparison, is the distribution for constant viability selection using data from Spencer and Marks (1992) (blue bars).

Figure 4. Relationship between the number of alleles ($n$) and population mean fitness ($\bar{w}$) at Generation $10^4$ in 1000 replicate simulations of the constructionist simulations of maternal selection. Mean values plotted in yellow circles.

Figure 5. Mean fitnesses for different classes of offspring-mother combinations at Generation $10^4$ in 1000 replicate constructionist simulations of maternal selection, plotted for different final numbers of alleles ($n$). Combination $C_{iii}$ is the mean of homozygous offspring with homozygous mothers, for example, whereas $C_{iij}$ is the mean of heterozygous offspring with
identically heterozygous mothers. For a full explanation see text. Means for each class are shown in yellow circles.

Figure 6. Histograms of $I = \sum_{i=1}^{n} (p_i - \frac{1}{n})^2$, a measure of centrality of allele frequencies at equilibria in the parameter-space simulations (blue dotted lines) and at Generation $10^4$ in the constructionist approach (red dashed lines). Also shown (black solid lines) is the distribution expected if allele frequencies were random, values generated using the broken-stick approach. For $n = 2$, the same information is shown as a small histogram of the respective values of $p_1$. 
Maternal Mean = 2.02  Viability Mean = 5.40