The Maintenance of Single-locus Polymorphism by Maternal Selection

Kai X. Chiew

Allan Wilson Centre, Department of Zoology, University of Otago

Introduction

Mechanisms to maintain polymorphism have been a puzzle for evolutionary scientists for a long time, especially after the development of gel electrophoresis in the 1970s (Lewontin 1974; Kimura 1984; Mitchell-Olds et al. 2007; Hahn 2008; Leffler et al. 2012; Delph and Kelly 2014;). Natural selection appears to maintain genetic variation in natural populations (Gao et al. 2015). Two different polymorphism theories, the balancing school (Lewontin 1974) and the neutral hypothesis (Kimura 1984), arose from the 70s and 80s, respectively.

The balancing school hypothesis argues that genetic polymorphism is maintained at the frequencies higher than mutation rate (Lewontin 1974). The neutral hypothesis assumes most mutations are neutral and deleterious mutations can be removed by natural selection (Kimura 1984). Kimura (1984), who postulated the neutral hypothesis, has pointed out some theoretical problems underlying the balance school theory. Following balancing-school theory, Lewontin et al. (1978) have shown the proportion of parameter space in models of constant viability selection maintaining 4 or 5 alleles is very small. Similar results are shown in several other types of viability selection including 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).

Many scientists misinterpret these results as meaning that selection could not maintain a large number 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). However, the Lewontin et al. (1978) result does not necessarily mean the likelihood of real parameters lying in the region is small because there is no iteration of allele frequencies over many generations. Instead, it is just that the proportion of parameter/state space that preserves allelic variation is small. Spencer and Marks (1988, 1992; Marks and Spencer 1991) have realized these core problems underlying this “parameter-space” approach and proposed a “constructionist” approach.

**Constructed polymorphism** is started from a monomorphism, then shaped actively by the joint action of mutation and selection (Spencer and Marks 1988, 1992; Marks and Spencer 1991). For each generation, there are mutation, selection and extinction checks. Mutation produces new mutants that have either higher fitness to get into the system to become part of the standing variation, or lower fitness and are removed. Selection iterates the allele frequencies based on their fitness. Finally the extinction check acts to remove the alleles that have fallen below a threshold frequency from the system. Numerous models of selection have been investigated under this “constructionist” approach (Marks and Ptak 2001; Star et al. 2007b, 2008; Trotter and Spencer 2008, 2013).

Maternal selection is a very important form of selection. Many mammals’ genetic and phenotypic expression are largely compounded by the maternal effect (Mousseau and Fox 1998; Mousseau 2006; Wolf and Cheverud 2012). Following the one-locus two-allele model proposed by Gavrilets (1998), we first use the “parameter-space” approach to determine the size of parameter/state space that leads to polymorphism, then the “constructionist” approach to access the likelihood of polymorphism that lies under the region. We investigate the level of polymorphisms and asking questions like how many alleles are maintained in a polymorphism? Does maternal selection increase the level of polymorphism and lead to higher number of alleles when compared to constant viability selection?
Models of Maternal Selection

Gavrilets (1998) has proposed one-locus two-allele model of maternal selection and Spencer (2003) has derived further properties. We extended the model to multiple alleles for further investigation of polymorphism under maternal selection. 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.

Maternal selection means we must record the offspring’s own allele and the mother’s alleles (Cheverud and Wolf 2009; Wolf and Wade 2009; Wolf and Cheverud 2012). Suppose \( w_{ijkl} \) is the relative fitness \((0 \leq w_{ijkl} \leq 1)\) of offspring of genotype \( AA_i \) with genotype \( A_kA_l \) mothers. One or both of offspring’s alleles in a genotype \( AA_i \) must be inherited from mother’s alleles of genotype \( A_kA_l \), therefore \( i = k \) or \( l \), or \( j = k \) or \( l \), or both.

Under these conditions, we have enumerated the possible number of offspring-mother combinations under the formula \( C(n) = n^2 + \frac{1}{2} n(n-1)(2n-1) \) where \( n^2 \) is the possible homozygote offspring combination and \( \frac{1}{2} n(n-1)(2n-1) \) is the possible heterozygote offspring combinations. We only include the fitness parameters that fit the formula into our consideration when writing the programme. For example, when \( n=2 \), the number of fitness parameters, \( C(n) \) is only 7 and we will exclude fitness parameters such as \( w_{1122} \) and \( w_{2211} \) for which offspring-mother combinations do not exist. The collection of 7 different fitness parameters is considered a fitness set.

For those possible fitness parameters, we have classified them into five distinct classes: homozygotes with homozygous mother \( (C_{ii}) \), homozygotes with heterozygous mothers \( (C_{ij}) \), heterozygotes with homozygous mothers \( (C_{ij}) \), heterozygotes with identical heterozygous mothers \( (C_{ij}) \) and heterozygotes with non-identical heterozygous mothers \( (C_{ij}) \). We examine heterozygote advantage between classes by comparing their mean fitness.
Following Gavrilets (1998) and Spencer (2003), we do iterations of genotype frequencies for homozygotes through \( \bar{w}x'_{ij} = p_i \left( w_{iii}x_{ii} + \frac{1}{2} \sum_{k \neq i} w_{iik}x_{ik} \right) \) and heterozygotes offspring through \( \bar{w}x'_{ij} = p_j \left( w_{jjj}x_{jj} + \frac{1}{2} \sum_{k \neq j} w_{jik}x_{jk} \right) + p_j \left( w_{jii}x_{ji} + \frac{1}{2} \sum_{k \neq i} w_{jik}x_{ik} \right) \) equations. These iterations mean the genotype frequencies are changing every generation starting from initial genotype frequencies. We obtain the mean fitness through the sum of the right-hand sides of both equations.

**Estimating Potential for Polymorphism:** Following Trotter and Spencer’s (2007) method using “parameter-space” approach, we first iterate 100 randomly selected initial genotype frequencies for every fitness set, the values of which are drawn from \( U[0, 1] \). The “potential” for polymorphism is calculated by adding the proportion initial genotype-frequency vectors that “Always” and “Sometimes” maintain all \( n \) alleles. Spencer (2003) has shown that allele-frequency “Cycling” is possible but rare. Initial allele frequencies are drawn using the broken-stick method (Holst 1980). The proportion of “Always” fitness sets in \( 10^5 \) fitness sets is also recorded.

**Constructing Polymorphism:** Spencer and Marks (1988) developed the constructionist approach to investigate the level of polymorphism across the generations. The model begins with monomorphism \( (n=1) \) and assumes \( w_{1111}=0.5 \) from first generation. Every generation, a new mutant allele, \( A_{n+1} \), is added to the existing \( n \)-allele system at low frequency of \( \rho_{n+1} = 10^{-5} \). The genotype frequencies are iterated every run. Extinction check is carried out as those allele frequencies fell below \( 10^{-5} \) to be removed from the system. We recorded the number, frequency and mean fitness of all alleles for every run until \( 10^4 \) generations. This whole simulation process is considered one run. We repeat the simulations, differing only in the seed for the pseudo-random number generator (Marsaglia et al 1990), for a thousand times and capture the fitness set of extant alleles at the \( 10^4 \)th generation.
Results

Potential for Polymorphism: The proportion of full polymorphism parameter space is shown in Figure 1. Full polymorphism is achieved when the initial genotype frequencies of all the alleles at equilibrium. “Always” fitness set achieve full polymorphism for all 100 initial genotype frequencies. We also included potential for polymorphism because sometimes only a few out of 100 initial genotype frequencies achieve full polymorphism. The constant viability selection result from Lewontin et al. (1978) is plotted to make the comparison. From Figure 1, the proportion of full polymorphism drops sharply with $n$ for both maternal and constant viability selection. This result is consistent with other forms of selections investigated to date. However, the potential for polymorphism under maternal selection is larger 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.
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.

*Constructing Polymorphism:* Eight constructionist simulations with significant pattern were chosen and are shown in Figure 2. There are a number of successful new mutant invasions into the system in the early generations, which are quickly followed by extinction of one or more alleles. Almost all the simulations followed this pattern as the number of allele rises quickly, then soon drops to ≤ 5. However, most invasions are unsuccessful as the number of alleles remain unchanged for most of the generations. Variation or polymorphism needs a long period of time before it changes again. Mean fitness is changing accordingly to the number of alleles due to the different finesses for successfully invading or extinct alleles.
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.
The distribution of the number of alleles in each of the 1000 simulations after $10^4$ generations is shown in Figure 3. We also included the data from Spencer and Mark (1998) of constant viability selection as a comparison for our maternal selection. The distribution of number of alleles for maternal selection is narrow, ranges from 1 to 5 with a peak at 2. Viability selection has a wider distribution of number of alleles ranges from 1 to 20 with a peak at 5. This data generated using the constructionist approach contrast with Figure 1 which is parameter-space approach. The larger potential for polymorphism in maternal selection is not reflected in the constructed polymorphisms.

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).
The mean fitnesses of 1000 simulations for different number of alleles are shown in Figure 4. Mean fitness is lower for the simulations with higher number of alleles. Invasion takes place more easily when the mean fitness is lower.

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.

From Figure 5, mean fitnesses are different for different classes of offspring-mother combination with the heterozygotes enjoyed the higher fitnesses (i.e. $C_{iii} < C_{ijj} < C_{ijj} < C_{ijj}$ and $C_{ijk}$).
Interestingly, the heterozygote advantage is even more noticeable when their mothers are also heterozygous.

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 $C_{ii}$ is the mean of homozygous offspring with homozygous mothers, for example, whereas $C_{ij}$ 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.
Our investigation of maintenance of polymorphism by maternal selection using two different approaches is showing us two different kind of results. Using the parameter-space approach without iteration over thousands of generations, we see that the “potential” for polymorphism to maintain large number of alleles under maternal selection is much bigger than viability selection. However, the approach using constructed polymorphisms with selection and mutation over time is showing a lesser number of alleles maintained in polymorphism. This pair of results is different because they are referring to two different aspects. The size of parameter space provides us useful information of the “potential” of polymorphism. The combined effects of selection and mutation under the “constructionist” approach shows us how easy it is to reach the portion that maintains polymorphism via a simple process.

The higher potential of polymorphism in maternal selection is due to inclusion of “Sometimes” and “Cycling” fitness sets that are not always maintaining all \( n \) alleles at equilibrium starting from all allele frequencies (Fig. 1). For these “Sometimes” and “Cycling” fitness sets, they do not always maintain all \( n \) alleles and can lose one or more alleles depending on their frequencies. However, if using the “constructionist” approach, there is a constant invasion of new mutants that make up for the lost alleles. Thus these “Sometimes” and “Cycling” fitness sets would only affect and increase the proportion of parameter/state space under parameter-space approach. This difference may explain the lower numbers of alleles found in our “constructionist” approach.

The maternal selection model produces a large number of fitness parameters when the number of alleles, \( n \), increases and causes a higher mean fitness. It is harder for a new mutant to invade under this parameter-rich model of maternal selection because the appropriate fitness for a successful invasion is higher. The maximum number of alleles observed in the constructionist simulations was 5, which only happened twice in 1000 simulations (Fig. 3). In Fig. 2 in our selected
simulations, the mean fitness of a population can sometime decline, however over many
generations the mean fitness generally increases.

Heterozygote advantage is reflected in the heterozygous offspring-mother combinations
with higher mean fitness. The fitness sets under these classes tend to be higher than the other
homozygous offspring-mother combination. The class with the highest mean fitness in our
collectionist simulations is that of heterozygotes with identically heterozygous mothers $C_{ij}$. It is
interesting to point out that their mean fitness is higher than that of heterozygotes with non-
identical heterozygotes $C_{iii}$, even though their genotype is identical to their mothers.

In this project, we have contributed new mathematical models and equations to discover
new maternal polymorphic maintenance mechanisms which can be built upon in the future to
improve our understanding of maternal selection.

**Acknowledgements:**

I would like to thank Prof. Hamish G. Spencer for the opportunity to work on this project over
summer, and his invaluable guidance and expertise. Many thanks go to the Allan Wilson Centre for
providing the funding for this summer studentship.

**Literature Cited**

Árnason, E., 1982 An experimental study of neutrality at the malic dehydrogenase and esterase-5

*PLoS Genet.* 2: e64.


Wolf, J. B., and M. J. Wade, 2009 What are maternal effects (and what are they not)? Philos. Trans. R. Soc. Lond. B Biol. Sci. 364