The approach to equilibrium of multilocus genotype diversity under clonal selection and cyclical parthenogenesis

Charles E. King & Justin Schonfeld

1Department of Zoology, Oregon State University, Corvallis, OR 97331, U.S.A.
E-mail: kingc@ava.bcc.orst.edu
2Department of Computer Sciences, Oregon State University, Corvallis, OR 97331, U.S.A.
3Current address: Computational Biology & Bioinformatics, Iowa State University, Ames, IA 50011, U.S.A.
E-mail: schonfju@iastate.edu

Key words: adaptive population structure, cyclical parthenogenesis, clonal selection, rotifers

Abstract

Sexual generations in cyclical parthenogens are typically separated by multiple generations of clonal reproduction. In contrast to sexual reproduction, during parthenogenesis the genome of the parent is passed on to the offspring as a unit. The absence of recombination during parthenogenesis leads to differences in the action of natural selection in the two reproductive phases. In addition, since recombination is a sampling process, random genetic drift is potentially more important in sexual reproduction than in parthenogenesis. A recent development in the study of rotifer population genetics is the use of microsatellites to characterize natural populations. Microsatellites are selectively neutral, show patterns of Mendelian inheritance and tend to be much more variable than allozymes. An advantage over allozymes is that microsatellite DNA can be cloned with PCR and thus multiple loci can be assayed from a single individual. We use a new computer model in this paper to investigate the response of selectively active and selectively neutral genes to evolutionary forces during cyclical parthenogenesis. Selectively active alleles may respond differently to selection in the parthenogenetic and sexual phases of cyclical parthenogenesis. Even when strong clonal selection is acting on loci associated with adaptation, the view that emerges with microsatellites may be one of Hardy-Weinberg and linkage equilibrium. Thus studies using selectively neutral loci may fail to detect clonal selection even when it is an important feature of the rotifer population’s adaptive structure.

Introduction

In cyclical parthenogens such as monogonont rotifers, the life cycle is divided into two distinct phases. The sexual phase includes the classic genetic processes of segregation and recombination. Natural selection in this phase also conforms to the classical model; that is, alleles that make positive contributions to fitness at independent loci are expected to increase in frequency, and those which are deleterious are expected to become less common.

In contrast, during the parthenogenetic phase of the life cycle, the genome is effectively congealed so that the offspring of a single female are genetically identical to each other and their mother. Different parthenogenetic clones may have different relative fitnesses thereby creating a process of clonal selection.

This clonal selection is based on a composite measure of fitness taken over all loci — those clones having high fitness increase in frequency while less well-adapted clones become relatively less common. Clonal selection may also be a major determinant of population structure in cyclical parthenogens.

Two methods have been used to characterize the genetic structure of rotifers. Allozyme studies have revealed the presence of significant amounts of genetic variation in a number of populations (reviewed by King & Serra, 1998). However, the identification of clones from field samples requires that each individual be genotyped at a number of loci so that multilocus genotypes (MLGs) can be compared. This creates a major problem since rotifers are small, and even under ideal conditions single individuals can provide enough material to assay for only one or two allozyme loci.
The alternative is to isolate and clone field-captured individuals in the laboratory, a labor-intensive and expensive procedure. Consequently, most allozyme analyses of rotifer populations have both smaller numbers of individuals and fewer assayed loci than is desirable.

Accordingly, it is fair to ask whether the pattern of clonal selection inferred from allozyme studies is an accurate picture of the genetic structure of rotifer populations. That is, does our view of rotifer population structure reflect biological reality or simply the limitations of our analytical techniques? Might new analytical methods produce a different view?

A promising new method is the use of PCR-amplified DNA markers that require little material and generally have much higher levels of genetic variation than allozymes. Gómez et al. (1998), studying Brachionus plicatilis, scored seven polymorphic microsatellite loci having 34 alleles. (Microsatellites are tandemly repeated, short sequences of DNA found abundantly within the genomes of eukaryotes. They are generally less than six nucleotides in length and show a pattern of Mendelian inheritance.)

This system has recently been used by Gómez & Carvalho (2000) to characterize microsatellite variation in a population of B. plicatilis living in Poza Sur of Torreblanca Marsh on the Mediterranean coast of Spain.

Relative to allozyme studies of rotifers, but not to most microsatellite studies of other groups, an enormous amount of genetic variation was found in this population: of the 390 individuals assayed, 349 had unique MLGs. One interpretation of the high clonal diversities observed by Gómez & Carvalho (2000) is that clonal selection may be weak. In support of this, interpretation statistical analysis was unable to reject the null hypotheses of Hardy-Weinberg and linkage equilibrium in most of the samples, including a collection of resting eggs from the surface sediment of Poza Sur. However, the last sample from the parthenogenetic phase showed evidence of clonal selection and, as expected, it also displayed linkage disequilibrium.

A number of interesting questions arise from Gómez & Carvalho's (2000) study. First, how does the genetic structure of rotifers reflect differences in the selective regime between the sexual and parthenogenetic phases? That is, selection may produce quite different effects when it is accompanied by segregation and recombination than when it is directed at the entire genome as a unit. Second, how does the perceived genetic structure reflect the types of loci being investigated? Although microsatellites are thought to be selectively neutral (Jarne & Lagoda, 1996), selection may act on linked genes and thereby influence both the quantity of genetic variation and its distribution across the genome as a whole. Hitchhiking phenomena are particularly intense during the parthenogenetic phase of rotifer reproduction when all loci are effectively congealed. Finally, how does the genetic structure of cyclical parthenogens reflect levels of variation in the system being investigated? Microsatellite loci may be highly polymorphic and contain as many as 30–60 alleles per locus. While this high variability may seem ideal for studying the genetic structure of a population, it is accompanied by a price: as levels of variation increase, larger sample sizes are needed to accurately determine allele and genotype frequencies (Ruzzante, 1998). Gómez & Carvalho (2000) found much less variation in their study than the rates of polymorphism cited above; they detected 27 alleles in the seven loci they investigated for an average of about 4 alleles per locus. Still, this level of variation far exceeds that found in any study of rotifer populations using allozyme markers.

To explore these questions, and in particular to examine the role of cyclical parthenogenesis, we have constructed a computer model that compares the dynamics of alleles at loci subject to natural selection with alleles at selectively neutral loci. In this paper, we evaluate how the intensity of clonal selection affects both the genotype diversity and clonal diversity (number of MLGs) of cyclical parthenogens. Then we use our model to study the interactions of clonal selection, mutation, and random genetic drift as determinants of population structure. Finally, we explore potential effects of sexual recombination on clonal diversity in cyclical parthenogens.

A model for natural selection, mutation and random genetic drift in cyclical parthenogens

Program structure

Our model is written in Visual C++ and contains two loops, one for sexual reproduction and a second for parthenogenesis (Fig. 1). Input to the model includes information on the number of loci and the allele frequencies, selective values and mutation rates at each locus. A run is initiated by forming a pool of the input genotypes in Hardy-Weinberg equilibrium at each locus. Then one genotype is randomly selected from
each pool to constitute the first MLG. This process of random selection is repeated until the desired number of MLGs is obtained.

For both loops, first mutation and then selection is applied to the MLG pool. Mutation is handled by calculating the total number of mutations expected to occur at each locus and then randomly assigning the mutations to specific alleles at the locus. After all mutational events have been completed, MLG frequencies are adjusted to account for the losses and gains derived from mutation. At this point, the modeled population can either undergo sexual reproduction or go into the parthenogenetic loop for any desired number of generations.

Random numbers are used in two parts of the model, to form the MLG pool and to distribute mutations. The procedure used for both is nearly identical. In each sexual generation, a random number generator is seeded using the current time. To form the MLG pool, the Hardy-Weinberg frequencies of each of the genotypes present at the locus are placed in an array. A single genotype is randomly selected from the array for inclusion in the MLG being formed. Rare genotypes in the array will seldom be selected thus raising the possibility of loss of variation through random genetic drift. The process is then iterated for other loci until the complete MLG has been constructed. The same procedure is used for selecting the MLG alleles that will mutate except that MLG frequencies are used to create the array.

Assumptions

Loci are diploid, assort independently during sexual reproduction, and are inherited as a unit during parthenogenesis.

Alleles subject to natural selection are assumed to make additive contributions to the fitness of each genotype, so in this system there is no heterozygote advantage, frequency-dependent or fluctuating selection.

All alleles at a given locus are assumed to have the same mutation rate, but mutation rates may vary between any two loci in our model. Mutation rates of microsatellites ($10^{-4}$ to $10^{-2}$) appear to be two to three orders of magnitude higher than those of allozymes (Weber & Wong, 1993; Jarne & Lagoda, 1996; Ruzzante, 1998). In the present paper, mutation rates of $10^{-6}$ have been assumed for loci under selection and $10^{-4}$ for selectively neutral loci.

In order to follow the effects of selection, the population is initiated with a single, MLG that is homozygous for the lowest fitness allele at each locus. This construction means that the first generation of individuals will produce only beneficial mutations. In subsequent generations, mutations can be deleterious,
neutral or beneficial depending on the genotype that is mutating and the allele that is introduced.

All runs in the present paper were made by random sampling from the genotype pool for each locus to construct a population of 1,000,000 MLGs (individuals). We restrict our view in this paper to six loci. Loci 1–3 are under natural selection and each has three alleles (A–C). Loci 4–6 are neutral and each has six alleles (A–F). Fitnesses are assumed to be determined by additive effects. The fitness contributions of alleles A, B and C at locus 1 are 0.01, 0.5 and 1.0; for locus 2 they are 0.25, 0.625 and 1.0, and for locus 3 they are 0.5, 0.75 and 1.0. Thus, selection at each of the three loci will always favor allele B over A, and C over either A or B. Furthermore, selection will be more intense on 1 than on 2 or 3 and more intense on 2 than on 3. The response of a clonally reproducing population to selection at these levels over a wide range of mutation rates is presented in Figure 2. Note that the three curves do not overlap, and at a given level of mutation, selection intensity acts to scale the population response.

Output

For each generation, output includes total genotype diversity (GDT), number and frequency of MLGs, and genotype and allele frequencies at each locus. Genotype diversity as given by Hoffman (1986) is:

\[
GDT = \frac{1}{KT} \sum_{i=1}^{KT} p_i^2
\]

in which KT is the total number of unique multilocus genotypes in the simulated population and \( p_i \) is the frequency of the \( i \)-th MLG. Stoddart and Taylor (1988) explored some of the statistical properties of GDT distributions.

Because we are interested in the role of selection on genotype diversity, we partition GDT into one component, GDN, based only on selectively neutral loci and a second component, GDS, based on loci that respond directly to natural selection. Similarly, KT, the total number of unique MLGs, can be partitioned into neutral and selected components, KN and KS, based on the numbers of unique MLGs considering the two types of loci separately.

Results

A series of three replicate simulations of a model population with 100 generations of parthenogenesis is presented in Figure 3. These runs are included to illustrate population behavior under clonal selection alone; subsequent runs consider effects of sexual reproduction and variation in the relative intensity of clonal selection. Differences in the output of the three runs are solely determined by the stochastic timing and identity of the new variation produced by mutation.

All runs in this paper were initiated with a single MLG that was homozygous for the A alleles at each of its six loci. As mutation occurs, there is an increase in the number of unique MLGs at both selected and neutral loci as well as the total number of MLGs. Note that KT is not the sum of KN + KS because of the way the two types of loci are defined. For instance, if KS = 1 and KN = 4, KT = 4. If a single new mutation occurs at a selected locus in one of the four MLG classes, KS = 2 and KN = 4 but KT = 5. If the same new mutation occurs in all four MLG classes, KS = 2, KN = 4 but KT = 8. In Figure 3 KN is larger than KS primarily because of the higher mutation rates at neutral loci and the larger number of neutral than selectively active alleles.

Genotype diversity is sensitive to variation in both MLG frequency and number. Accordingly, GDT patterns primarily reflect the effects of selection acting on new MLGs. For example, the first peak in the center panel of population ‘a’ is caused by the appearance and rapid increase of a new allele at locus 1 that produces MLGs having higher fitness that the original AA homozygote. A second peak is produced as allele 1C moves towards fixation. The more complex pattern in population ‘b’ is attributable to the added contribu-
tions of a second locus that is also under selection. In general, the neutral alleles have a small initial impact on genotype diversity since their frequencies early in a run are determined by mutation and rare alleles make small contributions to GDT. However, under clonal selection rare neutral alleles may be associated with high fitness genotypes in MLGs and their frequencies can increase rapidly because of the hitchhiking effect (population 'c' in Fig. 3).

Genotype diversity under strong clonal selection.

For a given set of allelic contributions to fitness, the strength of clonal selection acting on a population is proportional to the number of parthenogenetic generations intervening between instances of sexual reproduction. Figure 4 shows a 100-generation simulation conducted with the 6-locus case described above. Again, the population was started with a single MLG that was homozygous for the 'A' alleles at each locus. Sexual reproduction occurs at intervals of 25 generations.

The most striking difference between the examples in Figures 2 and 3 relates to the change in population structure after sexual reproduction. Whenever sex occurs, there is a dramatic decline in the number of MLGs. Immediately prior to sexual reproduction most of the neutral alleles are present in low frequency. When the population is reconstituted by random sampling from the alleles at each locus, rare alleles are likely to be lost. This loss is due to random genetic drift, specifically a founder effect.

In the upper panel of Figure 4 there is a rapid increase in the number of MLGs following sexual reproduction. This increase is due to both new mutation and the replacement of low fitness alleles at the three loci subject to natural selection. The fitness of an MLG is taken over all loci that are subject to direct natural selection and expressed relative to that of other MLGs. In this system, the most fit MLG has a relative fitness of 1 and a genotype containing only C alleles at the
Figure 4. Strong clonal selection: Effects of mutation and clonal selection on a life history that has 24 generations of parthenogenesis between successive sexual generations. Other input parameters are the same as in Figure 3.

Figure 5. Weak clonal selection. Effects of mutation and clonal selection on a life history that has 4 generations of parthenogenesis between successive sexual generations. Other input parameters are the same as in Figure 3.

selection loci. As $B$ and $C$ alleles are generated by mutation they rapidly replace the $A$ alleles and ultimately the $C$ alleles at loci 1–3 will become fixed in the population. Although new $A$ and $B$ alleles continue to be generated by mutation, because of their low mutation rates and negative influence on fitness, they have little impact on the population's genetic structure. The frequencies of neutral alleles associated with high fitness MLGs increase as the selection occurs. A succession of selective changes is seen in the bottom panel of Figure 4. First, the $A$ allele of locus 1 is replaced by the $C$ allele and this is followed by the same change at the second locus and finally again at the third locus. The order of this change is determined by the relative contributions to fitness of the alleles at the three loci. Selection is strongest at selection locus 1 and weakest at selection locus 3.

In the center and bottom panels of Figure 4, GDT fluctuates as mutation creates and selection eliminates genetic variation. Not surprisingly, GDS mirrors the GDT variation. Clonal selection acting on the allele frequencies shown in the bottom panel is the major determinant of the patterns of genotype diversity.

**Genotype diversity of selected and neutral loci under weak clonal selection**

All else being equal, the intensity of clonal selection is directly proportional to the number of parthenogenetic generations intervening between sexual generations. In Figure 5, all input parameters are the same as in Figure 4 except that instead of 24 there are now 4 generations of parthenogenesis followed by a single generation of sexual reproduction. Again, a brief drop in the number of MLGs and the genotype diversity after each occurrence of sex is attributable to random genetic drift. Genotype diversity is buffered by the fact that the lost MLGs were present in low frequencies so
the major changes in GDT reflect the effects of clonal selection displayed in the bottom panel. Notice that the frequency of allele 5A decreases as other alleles at this neutral locus are generated by mutation. Chance include mutations of neutral alleles may have important effects on genotype diversity even when sexual reproduction is frequent.

Although the qualitative outcomes of clonal selection are identical for the 4 and 24 parthenogenetic generation cases, there are fewer MLGs produced when sex is more frequent (Figures 4 vs 5). As explained above, this is because frequent sexual reproduction maximizes the effect of random genetic drift as the population is reformed.

Discussion

Equilibrium genotype and clonal diversities

Expected equilibrium levels of both genotype diversity and clonal diversity in cyclical parthenogens are different for neutral and selected loci. In this paper we have used a model population with 3 selected and 3 neutral loci. Since there are 6 possible genotypes at each selected locus and 21 at each neutral locus, maximum clonal diversity at the selected loci (KS) is $6^3 = 216$ MLGs. At the neutral loci, maximum clonal diversity (KN) is $(21)^3 = 9261$ MLGs. Ignoring minor fluctuations due to recurrent mutation, under clonal selection the population in our model will always move towards KS = 1 at equilibrium. This MLG will be homozygous for the allele having highest fitness at each locus. Genotype diversity equilibrium at the selected loci will also occur at GDS = 1. Note, however, that attainment of these equilibrium values may take much longer than the 100 generations us in our simulations.

In striking contrast to loci under selection, mutation drives the population toward an equilibrium in which all possible neutral genotypes are present at each locus and in which all alleles at each locus have the same relative frequency. This state follows from our assumption of equivalence of mutation rates at a single locus. Genotype diversity of the neutral loci under these circumstances also moves towards a value of GDN=1 which requires that all neutral MLGs have the same frequencies. This condition would be possible only if the population had no sexual reproduction.
the average locus, recombination during sexual reproduction increases the frequencies of MLGs containing neutral heterozygotes relative to those containing individual classes of homozygotes. In sexual populations when the neutral genotypes at each locus are in Hardy-Weinberg equilibrium, an equilibrium GDN > 1 will occur. Total genotype diversity (GDT) as the system approaches equilibrium under the combined action of mutation and selection reflects both of these forces. Maximum GDT is expected to occur when GDS = 1 and the single selected MLG is present on all possible neutral MLGs. Under these conditions, GDN equals GDT.

**Detection of clonal selection in natural population**

A frequency of 10% non-unique multilocus genotypes (41 of the 390 individuals scored by Gómez & Carvalho, 2000) might imply weak clonal selection unless the number of potential MLGs is considered. One of the loci studied by Gómez & Carvalho (2000) had 2 alleles, a second 3 alleles, four loci each had 4 alleles and the remaining locus had 5 alleles. None of the alleles in their samples was near fixation. This system potentially has $4.05 \times 10^{6}$ unique MLGs. In the absence of clonal selection, the probability of randomly drawing two individuals having identical genotypes at the seven loci with a sample size of 390 is very small. From this perspective, clonal selection in the Poza Sur population would appear to be quite strong.

What is the significance of the relative frequencies of unique and non-unique microsatellite genotypes in a population? Consider a highly simplified hypothetical population in which there is one locus, S, with three alleles (A-C) subject to selection favoring a homozygote, and one microsatellite locus, N, having 4 neutral alleles (A-F). As discussed above, when the genotype with highest fitness is a homozygote, clonal selection tends to eliminate genotypic variation in the locus under selection, while mutation tends to increase the number and uniformity of distribution of microsatellite alleles. Equilibrium is reached at KS=1 for locus S when all individuals in the population have the same homozygous genotype (ignoring new, unfavorable mutations). At this point clonal selection stops. New alleles continue to be produced by mutation at locus N until it reaches its maximum clonal diversity at KN=10. Presuming equality of mutation rates, mutational equilibrium is attained when the four alleles have the identical frequencies of 0.25. At this point, any sample of $10 + n$ individuals will contain at least n non-unique MLGs. However, since all selectively inferior genotypes have been eliminated, the relative frequency of non-unique genotypes in the sample does not provide information on the presence or absence of clonal selection.

Now, let us assume that our hypothetical population has a seasonal structure in which homozygous AA individuals at locus S have a selective advantage in the spring, BB in the summer, and CC in the fall. Clonal selection operating on locus S over a period of time would tend to produce a seasonal succession of the three genotypes. Furthermore, if all sexual reproduction occurs within seasonal subpopulations, the relative frequencies of the three S homozygotes in the resting egg pool will reflect variates in genotype-specific reproductive success and frequencies of the three environmental states. Under these circumstances, loci under selection are likely to have large deviations from Hardy-Weinberg equilibrium show extensive multilocus linkage disequilibrium, and display large Wahlund effects in the resting egg pool.

However, since the alleles at locus N are not subject to selection, they are driven toward identical states of maximum diversity in each of the seasonal subpopulations. Therefore, when we examine allele frequencies at locus N in the resting egg pool we expect to see Hardy-Weinberg equilibrium, linkage equilibrium and no Wahlund effects. These are the observations that were made by Gómez & Carvalho (2000). Based on this reasoning, we do not think that the relative frequency of non-unique genotypes or, by extension, whether the microsatellite allele frequencies are or are not in Hardy-Weinberg and linkage equilibrium, constitute informative indicators of clonal selection. This mechanism may also explain the failure of allozyme analyses to detect seasonal changes in population structure even though they have been quite successful in detecting frequency changes of sibling species (King & Serra, 1998).

In summary, selectively neutral loci are inappropriate genetic markers to use in the study of the adaptive dynamics of cyclical parthenogens. In the absence of genes that can respond directly and differentially to clonal selection as the environment changes, it is not possible to study such adaptive processes as seasonal selection. This requirement suggests that neither allozymes nor microsatellites are likely to shed much light on the adaptive structure of rotifer populations.
Acknowledgements

We are deeply indebted to Dr Africa Gómez for valuable discussion and comments on the manuscript that have helped to resolve some of the differences of interpretation encountered by comparing natural and modeled populations.

References