

Sexual conflict and protein polymorphism

Ralph Haygood

Department of Zoology, University of Wisconsin, Madison, WI 53706

rhaygood@wisc.edu

Evolution **58**:1414–1423 (2004)

Abstract

Sexual conflict, where male and female reproductive interests differ, is probably widespread and often mediated by male or sperm proteins and female or egg proteins that bind to each other during mating or fertilization. One potential consequence is maintenance of polymorphism in these proteins, which might result in reproductive isolation between sympatric subpopulations. Here, I investigate the conditions for polymorphism maintenance in a family of mathematical models of sexual conflict over mating or fertilization frequency. The models represent a male or sperm ligand and a female or egg receptor, and they differ in whether expression of either protein is haploid or diploid. For diploid expression, the conditions imply that patterns of dominance, which involve neither overdominance nor underdominance, can determine whether polymorphism is maintained. For example, suppose ligand expression is diploid, and consider ligand alleles L_1 and L_2 in interactions with a given receptor genotype; if, say, L_1/L_1 -males are fitter than L_2/L_2 -males in these interactions, then polymorphism is more likely to be maintained when L_1/L_2 -males more closely resemble L_1/L_1 -males in these interactions. Such fitter-allele dominance might be typical of a ligand or its receptor due to their biochemistry, in which case polymorphism might be typical of the pair.

Introduction

Sexual conflict exists when an increase in a component of the fitness of one sex is accompanied by a decrease in a component of the fitness of the other (Chapman et al., 2003). The basis of such conflict is anisogamy and the differences between the sexes that follow from it, and the focus may be mating or fertilization frequency, litter or clutch size, paternal versus maternal investment in offspring, or any other aspect of reproduction in which male and female interests need not wholly agree. For example, in water striders, it is common for males to attempt mating and a female to resist it (Rowe et al., 1994); for a male, higher mating frequency is desirable, because his fitness is limited by access to eggs, and there is last-male fertilization advantage, whereas for a female, lower mating frequency is preferable, because she stores enough sperms from one mating to fertilize all her eggs for many days, so her fitness is limited by resource consumption, and mating interrupts foraging. There is evidence for sexual conflict in several taxa, and it seems likely to be widespread.

The evolutionary consequences of sexual conflict may be significant in several ways. Assuming the traits mediating it are sex limited and affected by different genetic loci, sexual conflict may lead to an arms race, that is, a succession of adaptations in the male trait and counteradaptations in the female trait. Mathematical models suggest that sexual conflict readily engenders “endless coevolutionary chase between the sexes” (Gavrilets, 2000; Gavrilets et al., 2001), and there is evidence for escalation of, for example, male clasping and female anticlasping anatomy in water striders (Arnqvist and Rowe, 2002). If escalation occurred simultaneously in separate populations, then the traits would probably diverge between the populations, especially if the traits were susceptible to modification in several ways. Such divergence in reproductive traits might result in reproductive isolation on secondary contact (Parker and Partridge, 1998).

Alternatively, sexual conflict may lead to polymorphism maintenance. For example, one type of male may do better with one type of female, another type of male may do better with another type of female, and the frequencies of the types of female may be such that both types of male do equally well on the average. Such polymorphism might result in reproductive isolation be-

tween subpopulations in the same location (Gavrilets and Waxman, 2002). Mathematical models affirm this possibility (Frank, 2000; Gavrilets and Waxman, 2002), but further investigation of the conditions for it is needed. To appreciate the evolutionary consequences of sexual conflict, it is necessary to delineate which ecological and genetic circumstances are conducive to polymorphism, and which are not. The primary goal of the present study is to investigate the conditions for polymorphism maintenance in a family of mathematical models of sexual conflict over mating or fertilization frequency.

The models represent two proteins, one expressed only in males or sperms and the other only in females or eggs. Variation in the anatomical and behavioral traits that have traditionally motivated studies of sexual selection (e.g., tail length in male birds) is presumably polygenic, but in many reproductive interactions, especially those between male seminal fluid and female reproductive tract or between sperm and egg, molecules of male or sperm proteins and female or egg proteins bind to each other, with variation in binding strength due to variation in molecular shape, size, etc. (Swanson and Vacquier, 2002). Several pairs of proteins are suspected of mediating sexual conflict. For example, in abalones and sea urchins, the strength of binding between certain sperm and egg proteins affects the frequency of fertilization when a sperm and egg meet (Metz et al., 1994; Palumbi, 1999; Kresge et al., 2001); mutations in the sperm proteins may be favored if they raise the frequency of fertilization and hence of victory in sperm competition, whereas mutations in the egg proteins may be favored if they lower the frequency of fertilization and hence of death by polyspermy (Rice and Holland, 1997).

When the gene for a protein is sequenced, something may be learned about the form and intensity of selection on the protein, even if the mechanism of selection is unclear, by comparing non-synonymous (protein-altering) with synonymous (protein-preserving) variation in the gene. The genes for several proteins suspected of mediating sexual conflict have been sequenced and indicate that the rapid evolution of these proteins is driven by selection (Swanson and Vacquier, 2002). For example, the abalone sperm protein lysin is one of the fastest-evolving proteins known, and a large

excess of nonsynonymous substitutions per nucleotide site over synonymous ones between species indicates that lysin evolution is driven by strong selection (Kresge et al., 2001). The abalone egg protein VERL (Vitelline Envelope Receptor of Lysin) also evolves rapidly, and the evolution of at least some parts of VERL is driven by selection (Galindo et al., 2003).

Moreover, some proteins suspected of mediating sexual conflict are nearly monomorphic within species, but others are highly polymorphic, and the polymorphism is maintained by selection. For example, lysin is nearly monomorphic, but within several species, the sea urchin sperm protein bindin is one of the most polymorphic proteins known, and an excess of nonsynonymous polymorphisms per nucleotide site over synonymous ones indicates that the polymorphism in at least some parts of bindin is maintained by selection (Metz and Palumbi, 1996; Biermann, 1998; Palumbi, 1999). The bindin receptor has not been isolated yet, but it is almost certainly distinct from bindin (Vacquier et al., 1995) and polymorphic within several species (Metz et al., 1994; Palumbi, 1999). Supposing these proteins indeed mediate sexual conflict, a secondary goal of the present study is to suggest reasons why some are nearly monomorphic, but others are highly polymorphic.

This article is organized as follows. First, I introduce ideas relevant to all models in a family of mathematical models of sexual conflict over mating or fertilization frequency. Next, I present these models and formulate algebraic conditions for maintenance of diallelic (two alleles per locus) polymorphism in a male or sperm protein and a female or egg protein. The models differ in whether expression of either protein is haploid or diploid. For diploid expression, the conditions imply that patterns of dominance, which involve neither overdominance (heterozygote advantage) nor underdominance (heterozygote disadvantage), can determine whether polymorphism is maintained. Finally, I discuss the significance of the models and conditions, including their implications for lysin, bindin, and other proteins suspected of mediating sexual conflict.

Ligand–receptor conflict

Assume one protein, the ligand, is expressed only in males or sperms, and another protein, the receptor, is expressed only in females or eggs. Assume there is no selection on these proteins

except through their interactions with each other, which is particularly plausible if they are gamete proteins. Assume gametes are haploid, and the life cycle includes a diploid phase; if, as in most animals, meiosis yields gametes, which do not undergo multicellular development, then adults are diploid, but if, as in many algae, fungi, and plants, meiosis yields spores, which do undergo multicellular development, then adults (gametophytes) are haploid. If there are sex chromosomes, then assume the ligand locus L and receptor locus R are autosomal. Let r be the frequency of recombination between L and R ($0 < r \leq 1/2$), and neglect recombination within L or R . Let L_1, L_2, \dots, L_{n_L} be the n_L alleles at L and R_1, R_2, \dots, R_{n_R} the n_R alleles at R . (Table 1 is a glossary of notation.)

If adults are haploid, then expression of both adult and gamete proteins is haploid. Even if adults are diploid, gamete proteins may be expressed from gamete genomes; many examples are known (e.g., protamines, Hecht, 1998). My first model supposes haploid expression of both ligand and receptor; their evolution is relatively simple, because they cannot exhibit dominance. If adults are diploid, then expression of adult proteins is diploid, and that of gamete proteins may be too. Even if gamete proteins are expressed from gamete genomes, if they are expressed in gamete precursors, then cytoplasmic bridges between these cells may result in effectively diploid expression. My second model supposes diploid expression of both ligand and receptor; their evolution may be complicated by dominance. My third model supposes haploid expression of the ligand and diploid expression of the receptor, as for many pollen and pistil proteins in flowering plants.

For brevity, I will refer to males or sperms as males, to females or eggs as females, and to mating or fertilization as mating. When a male and female have an opportunity to mate, the probability that they do mate depends on the biochemical fit between them. Define the compatibility of a receptor genotype with a ligand genotype as the frequency of mating when a female expressing the receptor genotype and a male expressing the ligand genotype meet. Specifically, for haploid expression, let $c_{i,j}$ be the frequency of mating when an R_i -female and an L_j -male meet ($0 < c_{i,j} \leq 1$); for diploid expression, let $c_{i/k,j/\ell}$ be the frequency of mating when an R_i/R_k -

female and an L_j/L_ℓ -male meet ($0 < c_{i/k,j/\ell} \leq 1$); and for haplodiploid expression, let $c_{i/k,j}$ be the frequency of mating when an R_i/R_k -female and an L_j -male meet ($0 < c_{i/k,j} \leq 1$). Implicitly, a compatibility pertains to a particular environment, and it averages over many kinds of individual variation. For simplicity, assume the compatibilities are distinct and do not change over time.

Under many different assumptions about the ecology of the organism, the ligand and receptor mediate sexual conflict over mating frequency. For simplicity and definiteness, my models make four ecological assumptions. The first is that the fitness of a male is never decreased by mating. If the “male” is actually a sperm, which meets only a single egg, then this assumption is trivial. If the “male” is actually a male, which may meet multiple females, then this assumption is appropriate if, as in many animals, costs of mating are low for males, and males invest nothing in mates or offspring.

In contrast, the fitness of a female may be decreased by mating too often. Accordingly, if several males expressing a given ligand genotype meet the same female, then the fitness of this genotype may be decreased if too many of these males mate with this female. If the males a female meets expressed correlated ligand genotypes, then ligand genotypes with higher compatibilities might have lower fitnesses. For simplicity, the second ecological assumption of my models is that females meet males at random. In particular, the probability that any one male a female meets expresses a given ligand genotype is the frequency of this genotype in the population, independent of any other male the female meets. This assumption is appropriate if the population is well mixed. Under the first and second assumptions, any ligand mutation whose only effects are to raise one or more compatibilities is favored.

Under the second assumption, the fitness of a receptor genotype is a function of the average compatibility of this genotype with the ligand genotypes in the population. Let $f[\bar{c}]$ be the fitness of a receptor genotype with average compatibility \bar{c} . No mating yields no fitness, so $f[0] = 0$, and some mating yields some fitness, so $f[\bar{c}]$ initially increases with \bar{c} . If more mating always yielded more fitness, then f would increase indefinitely. The third ecological assumption of my models is

that f eventually decreases, because costs of additional matings eventually outweigh benefits for females. Let \hat{c} be the intermediate compatibility that is optimal for receptor genotypes ($0 < \hat{c} < 1$). Figure 1 depicts a possible f , which I will use in numerical examples. This f might be appropriate for a broadcast spawner (e.g., abalones or sea urchins); it follows from a gamete-kinetic model similar to that of Styan (1998). In this context, \hat{c} strikes the best balance between the risk of going unfertilized, which dominates to the left of the peak, and that of becoming polyspermic, which dominates to the right. Implicitly, f pertains to a particular environment. For simplicity, assume f does not change over time.

For simplicity, the fourth ecological assumption of my models is that all compatibilities exceed \hat{c} , so as \bar{c} increases, $f[\bar{c}]$ decreases. This may or may not be evolutionarily probable. On one hand, if a compatibility does not exceed \hat{c} , then any mutation whose only effects are to raise this compatibility and perhaps others toward \hat{c} is favored. On the other hand, such mutations may not occur. The highest compatibility attainable may not exceed \hat{c} . More likely, mutations may raise some compatibilities but lower others. Under the third and fourth assumptions, any receptor mutation whose only effects are to lower one or more compatibilities toward \hat{c} is favored.

Haploid expression

Model

Let x_i and y_j be the frequencies of alleles R_i and L_j . The average compatibility of R_i is $\bar{c}_i = \sum_j y_j c_{i,j}$, and the fitness of R_i is $f[\bar{c}_i]$.

Evolution equations are conveniently formulated in terms of haplotype frequencies. Let z_{ij} be the frequency of haplotype $R_i L_j$. An $R_i L_j$ -female is expected to produce a number of offspring proportional to $f[\bar{c}_i]$, of which $R_k L_\ell$ -males are expected to sire a fraction $z_{k\ell} c_{i,\ell} / \bar{c}_i$, of which a fraction $(1 - r)/2$ are expected to be $R_i L_j$ -individuals. Neglecting genetic drift, the number of $R_i L_j$ -individuals produced in this way is proportional to $(1 - r) z_{ij} z_{k\ell} W_{i,\ell} / 2$, where $W_{i,\ell} = f[\bar{c}_i] c_{i,\ell} / \bar{c}_i$. Assuming nonoverlapping generations, similar reasoning about other ways of

producing $R_i L_j$ -individuals yields the one-generation evolution equations

$$z'_{ij} = \frac{1}{2\bar{W}} \sum_{k,\ell} ((1-r)z_{ij}z_{k\ell}(W_{i,\ell} + W_{k,j}) + rz_{i\ell}z_{kj}(W_{i,j} + W_{k,\ell})), \quad (1)$$

where $\bar{W} = \sum z_{ij}z_{k\ell}W_{i,\ell}$. In the $W_{i,j}$'s, $\bar{c}_i = \sum_{j,k} z_{kj}c_{i,j}$, so (1) is a closed system. Essentially this model was developed independently by Gavrillets and Waxman (2002).

For $n_R = n_L = 2$, it is instructive to change variables from the z_{ij} 's to $x_1 = z_{11} + z_{12}$, $y_1 = z_{11} + z_{21}$, and the linkage disequilibrium $D = z_{11} - x_1 y_1$. Routine algebra gives

$$x'_1 = \frac{1}{2}x_1 + \frac{1}{2\bar{W}}(x_1 U_1 + D(V_1 - V_2)) \quad (2a)$$

and

$$y'_1 = \frac{1}{2}y_1 + \frac{1}{2\bar{W}}(y_1 V_1 + D(U_1 - U_2)), \quad (2b)$$

where $U_i = f[\bar{c}_i]$ is the fitness of R_i , $V_j = \sum_i x_i f[\bar{c}_i]c_{i,j}/\bar{c}_i$ is the fitness of L_j , and $\bar{W} = \sum x_i U_i = \sum y_j V_j$ is the mean fitness. The structure of (2a) and (2b) is typical of diallelic haploid models of sexual selection (e.g., Kirkpatrick, 1982; Bulmer, 1989; Barton and Turelli, 1991); for example, in (2a), the first term represents the absence of direct selection on R through males, the first term in parentheses represents direct selection on R through females, and the second term in parentheses represents indirect selection on R through males due to nonrandom association of R_1 with L_1 when $D > 0$ or with L_2 when $D < 0$. The effect of D on the evolution of x_1 and y_1 is usually small.

Analysis

In general, (1) is resistant to algebraic analysis, but the cases $n_R = 1$, $n_L = 1$, and $n_R = n_L = 2$ are algebraically tractable and biologically instructive. For $n_R = 1$, if $c_{1,j} = \max\{c_{1,\ell}\}$, then L_j is unconditionally favored, and polymorphism at L is not maintained. Similarly, for $n_L = 1$, polymorphism at R is not maintained. However, for $n_R = n_L = 2$, polymorphism at both L and R may be maintained, as follows.

Polymorphism maintenance obviously requires fitness trade-offs between L_1 and L_2 and between R_1 and R_2 ; for example, if L_1 is favored in interactions with R_1 ($c_{1,1} > c_{1,2}$), then L_2 must be favored in interactions with R_2 ($c_{2,2} > c_{2,1}$), else L_1 would be unconditionally favored. However, it is not obvious that mere existence of these trade-offs is enough.

Figure 2 depicts an example of polymorphism maintenance, with trajectories spiraling toward an interior equilibrium. In this plot, three-dimensional (x_1 , y_1 , and D) trajectories are projected into the (x_1 , y_1)-plane. Polymorphism maintenance means the boundary is repelling. Specifically, the boundary must have two properties. First, all equilibria in the boundary must be unstable. As argued above, there are no equilibria in the edges, where one locus is polymorphic, so the only equilibria in the boundary are the corners, where neither locus is polymorphic. The edges are trajectories joining the corners to each other, so all four corners are unstable if and only if the edges form a loop (formally, a heteroclinic cycle). Second, the loop must be repelling.

(At the interior equilibrium in Figure 2, $U_1 = U_2$ and $V_1 = V_2$, so there is no sexual selection. However, there is assortative mating. There are two partially reproductively isolated subpopulations, one comprising R_1 -females and L_1 -males and the other R_2 -females and L_2 -males.)

To formulate polymorphism conditions, label alleles so that $c_{1,1} = \max\{c_{i,j}\}$. In particular, $c_{1,1} > c_{1,2}$; in a plot such as Figure 2, this implies the right edge points upward. Assuming $\hat{c} < \min\{c_{i,j}\}$, $f[c_{1,1}] < f[c_{2,1}]$; this implies the top edge points to the left. Given these assumptions, the edges form a loop if and only if

$$c_{2,2} > c_{2,1} \tag{3a}$$

and

$$f[c_{2,2}] < f[c_{1,2}]; \tag{3b}$$

condition (3a) implies the left edge points downward, and (3b) implies the bottom edge points to the right. Condition (3a) means there is a fitness trade-off between L_1 and L_2 , as described above, and (3b) means there is a fitness trade-off between R_1 and R_2 . These trade-offs are necessary for polymorphism maintenance.

Moreover, they are sufficient, because the boundary loop they entail is always repelling. (See the appendix for a sketch of a proof. The proof supposes selection is weak enough that the continuous-time limit of (1) is a good approximation. Simulations of the discrete-time system have revealed no exceptions.) Thus, for haploid expression, the polymorphism conditions can be summarized by saying that there must be fitness trade-offs between the alleles at each locus. For diploid expression, essentially these conditions are still necessary but no longer sufficient.

Diploid expression

Model

Let $x_{i/i}$, $2x_{i/k}$, $y_{j/j}$, and $2y_{j/\ell}$ be the frequencies of genotypes R_i/R_i , R_i/R_k for $i \neq k$, L_j/L_j , and L_j/L_ℓ for $j \neq \ell$. The average compatibility of R_i/R_k is $\bar{c}_{i/k} = \sum_{j,\ell} y_{j/\ell} c_{i/k,j/\ell}$, and the fitness of R_i/R_k is $f[\bar{c}_{i/k}]$.

Evolution equations are conveniently formulated in terms of diplotype frequencies. Let $z_{ij/ij}$ and $2z_{ij/k\ell}$ be the frequencies of diplotypes $R_i L_j/R_i L_j$ and $R_i L_j/R_k L_\ell$ for $i \neq k$ or $j \neq \ell$. Reasoning as before yields the one-generation evolution equations

$$z'_{ij/k\ell} = \frac{1}{2\bar{W}} \sum_{m,n,o,p} ((1-r)z_{ij/mn} + rz_{in/mj}) ((1-r)z_{k\ell/op} + rz_{kp/o\ell}) (W_{i/m,\ell/p} + W_{k/o,j/n}), \quad (4)$$

where $W_{i/m,\ell/p} = f[\bar{c}_{i/m}]c_{i/m,\ell/p}/\bar{c}_{i/m}$ and $\bar{W} = \sum z_{ij/mn}z_{k\ell/op}W_{i/m,\ell/p}$.

For $n_R = n_L = 2$, it is instructive to change variables from the $z_{ij/k\ell}$'s to x_1 , y_1 , and the disequilibria D_R , D_L , D_{RL} , $D_{R/L}$, D_{RLL} , D_{RRL} , and D_{RLRL} . Weir and Cockerham (1989) define these disequilibria in terms of the $z_{ij/k\ell}$'s and discuss their interpretation. The most important are the intralocus disequilibria D_R and D_L , which measure nonrandom associations of alleles within loci (i.e., deviations from Hardy–Weinberg genotype frequencies), the gametic (or linkage) disequilibrium D_{RL} , which measures nonrandom associations of alleles across loci within gametes, and the nongametic disequilibrium $D_{R/L}$, which measures nonrandom associations of

alleles across loci between gametes. Routine algebra gives

$$x'_1 = \frac{1}{2}x_1 + \frac{1}{2\bar{W}} \left((x_1^2 + D_R)U_{1/1} + (x_1x_2 - D_R)U_{1/2} + \right. \\ \left. (D_{RL} + D_{R/L})((y_1V_{1/1} + y_2V_{1/2}) - (y_2V_{2/2} + y_1V_{1/2})) + \right. \\ \left. D_{RLL}(V_{1/1} - V_{1/2} + V_{2/2} - V_{1/2}) \right) \quad (5a)$$

and

$$y'_1 = \frac{1}{2}y_1 + \frac{1}{2\bar{W}} \left((y_1^2 + D_L)V_{1/1} + (y_1y_2 - D_L)V_{1/2} + \right. \\ \left. (D_{RL} + D_{R/L})((x_1U_{1/1} + x_2U_{1/2}) - (x_2U_{2/2} + x_1U_{1/2})) + \right. \\ \left. D_{RRL}(U_{1/1} - U_{1/2} + U_{2/2} - U_{1/2}) \right), \quad (5b)$$

where $U_{i/k} = f[\bar{c}_{i/k}]$ is the fitness of R_i/R_k , $V_{j/\ell} = \sum_{i,k} x_{i/k} f[\bar{c}_{i/k}] c_{i/k,j/\ell} / \bar{c}_{i/k}$ is the fitness of L_j/L_ℓ , and $\bar{W} = \sum x_{i/k} U_{i/k} = \sum y_{j/\ell} V_{j/\ell}$ is the mean fitness. Like (2a) and (2b), (5a) and (5b) represent both direct and indirect selection. The disequilibria would decay under random mating, but they can persist under nonrandom mating. The primary effect is on D_{RL} and $D_{R/L}$; the others tend to be negligible. The effects of the disequilibria on the evolution of x_1 and y_1 are usually small.

Analysis

Like (1), (4) is amenable to algebraic analysis when $n_R = 1, n_L = 1$, or $n_R = n_L = 2$. For $n_R = 1$, polymorphism at L may be maintained only if there is overdominance in the $c_{1/1,j/\ell}$'s; for example, if $c_{1/1,1/2} > c_{1/1,1/1}$ and $c_{1/1,1/2} > c_{1/1,2/2}$, then both L_1 and L_2 may be maintained. Henceforth, assume there is no overdominance in the $c_{i/i,j/\ell}$'s for any i , for two reasons. First, this is more interesting from a population-genetic perspective—polymorphism maintenance through overdominance is already familiar. Second, it is conservative from a biochemical perspective—plausible mechanisms for overdominance are scarce. Likewise, assume there is no underdominance in the $c_{i/i,j/\ell}$'s for any i .

For $n_L = 1$, polymorphism at R may be maintained only if there is overdominance in the $f[c_{i/k,1/1}]$'s. Such overdominance does not require overdominance in the $c_{i/k,1/1}$'s themselves, provided these compatibilities straddle \hat{c} ; for example, if $c_{1/1,1/1} < c_{1/2,1/1} \approx \hat{c} < c_{2/2,1/1}$, then both R_1 and R_2 may be maintained. Henceforth, assume there is no overdominance in the $c_{i/k,j/j}$'s for any j . Assuming $\hat{c} < \min\{c_{i/k,j/j}\}$, it follows that there is no overdominance in the $f[c_{i/k,j/j}]$'s either. Likewise, assume there is no underdominance in the $c_{i/k,j/j}$'s for any j . It follows that there is no underdominance in the $f[c_{i/k,j/j}]$'s either.

For $n_R = n_L = 2$, Figure 3a depicts an example of polymorphism maintenance, with trajectories spiraling toward an interior equilibrium. In this plot, nine-dimensional (x_1, y_1 , and seven disequilibria) trajectories are projected into the (x_1, y_1) -plane. Polymorphism maintenance means the boundary is repelling. As before, the edges must form a loop, and the loop must be repelling.

(At the attracting equilibrium in Figure 3a, $U_{1,1} = U_{1,2} = U_{2,2}$ and $V_{1,1} = V_{1,2} = V_{2,2}$, because the $c_{i/k,j/\ell}$'s exhibit additivity (e.g., $c_{1/1,1/2}$ is halfway between $c_{1/1,1/1}$ and $c_{1/1,2/2}$). In general, the $U_{i/k}$'s and $V_{j/\ell}$'s can exhibit intermediate dominance (e.g., $U_{1,1} > U_{1,2} > U_{2,2}$), overdominance (e.g., $U_{1,2} > U_{1,1}$ and $U_{1,2} > U_{2,2}$), or underdominance (e.g., $U_{1,2} < U_{1,1}$ and $U_{1,2} < U_{2,2}$) at an attracting equilibrium. Moreover, trajectories can approach a limit cycle instead of an equilibrium.)

Label alleles so that $c_{1/1,1/1} = \max\{c_{i/i,j/\ell}\} = \max\{c_{i/k,j/j}\}$. Reasoning as before, the edges in a plot such as Figure 3a form a loop if and only if

$$c_{2/2,2/2} > c_{2/2,1/1} \tag{6a}$$

and

$$f[c_{2/2,2/2}] < f[c_{1/1,2/2}]. \tag{6b}$$

Like (3a) and (3b), (6a) and (6b) mean there are fitness trade-offs between L_1 and L_2 and between R_1 and R_2 . These trade-offs are still necessary but no longer sufficient for polymorphism maintenance.

Figure 3b depicts an example of polymorphism decay, with trajectories spiraling toward the boundary. The edges form a loop, but the loop is attracting. Figures 3a and 3b represent the same parameter values, except for ligand heterozygotes. In Figure 3a, L_1 and L_2 are additive (i.e., $c_{i/i,1/2}$ is halfway between $c_{i/i,1/1}$ and $c_{i/i,2/2}$), whereas in Figure 3b, L_1 is partially recessive in interactions with R_1 (i.e., $c_{1/1,1/2}$ is closer to $c_{1/1,2/2}$), and L_2 is partially recessive in interactions with R_2 (i.e., $c_{2/2,1/2}$ is closer to $c_{2/2,1/1}$). To understand the effect of dominance, consider trajectories from the lower left of Figure 3a or 3b. Here L_2 is common, so R_1 is favored (because $f[c_{1/1,2/2}] > f[c_{1/2,2/2}] > f[c_{2/2,2/2}]$), and trajectories proceed to the right. As R_1 becomes common, eventually L_1 becomes favored (because $c_{1/1,1/1} > c_{1/1,1/2} > c_{1/1,2/2}$), and trajectories turn upward. They turn upward sooner in Figure 3a, where L_1 and L_2 are additive, than in Figure 3b, where L_1 is partially recessive in interactions with R_1 . The delay is destabilizing, as delays in dynamical systems often are (Murray, 1989, p. 13).

To formulate the additional polymorphism condition, define χ_1 , χ_2 , ϕ_1 , and ϕ_2 by

$$c_{1/1,1/2} = \frac{1 + \chi_1}{2} c_{1/1,1/1} + \frac{1 - \chi_1}{2} c_{1/1,2/2}, \quad (7a)$$

$$c_{2/2,1/2} = \frac{1 + \chi_2}{2} c_{2/2,2/2} + \frac{1 - \chi_2}{2} c_{2/2,1/1}, \quad (7b)$$

$$f[c_{1/2,1/1}] = \frac{1 + \phi_1}{2} f[c_{2/2,1/1}] + \frac{1 - \phi_1}{2} f[c_{1/1,1/1}], \quad (7c)$$

and

$$f[c_{1/2,2/2}] = \frac{1 + \phi_2}{2} f[c_{1/1,2/2}] + \frac{1 - \phi_2}{2} f[c_{2/2,2/2}]. \quad (7d)$$

These dominance parameters are between -1 and 1 . Zero signifies additivity, a positive value signifies that a heterozygote resembles the more fit of two homozygotes in interactions with a homozygote at the other locus, and a negative value signifies the opposite. Given (6a) and (6b) and the assumptions preceding them, the edges in a plot such as Figure 3a or 3b form a loop, and to a

good approximation, the loop is repelling if and only if

$$\frac{c_{1/1,2/2}}{c_{1/1,1/1}} \frac{c_{2/2,1/1}}{c_{2/2,2/2}} \frac{f[c_{1/1,1/1}]}{f[c_{2/2,1/1}]} \frac{f[c_{2/2,2/2}]}{f[c_{1/1,2/2}]} \frac{1 - \chi_1}{1 + \chi_1} \frac{1 - \chi_2}{1 + \chi_2} \frac{1 - \phi_1}{1 + \phi_1} \frac{1 - \phi_2}{1 + \phi_2} < 1. \quad (8)$$

(See the appendix for a sketch of a proof. The proof supposes selection is weak enough that the continuous-time limit of (4) is a good approximation. Simulations of the discrete-time system with strong selection have revealed that occasionally, the loop is attracting even if (8) is satisfied.)

Given (6a) and (6b) and the assumptions preceding them, the first four factors in (8) are less than 1. If all dominance parameters are nonnegative, then the last four factors do not exceed 1, and (8) is satisfied, but if any dominance parameter is negative, then the corresponding factor exceeds 1, and (8) may be violated. A dominance parameter is negative when a heterozygote resembles the less fit of two homozygotes in interactions with a homozygote at the other locus; for example, in interactions with R_1/R_1 , L_2/L_2 is less fit than L_1/L_1 (because $c_{1/1,2/2} < c_{1/1,1/1}$), and the more negative χ_1 , the more closely L_1/L_2 resembles L_2/L_2 in this context. Call this resemblance fitter-allele recessiveness and the opposite fitter-allele dominance. Which allele is fitter depends on the genotype at the other locus; for example, in interactions with R_2/R_2 , L_2/L_2 is more fit than L_1/L_1 (because $c_{2/2,2/2} > c_{2/2,1/1}$), so resemblance of L_1/L_2 to L_2/L_2 in this context is fitter-allele dominance. If there is net fitter-allele recessiveness (i.e., the product of the last four factors exceeds 1), then (8) is more likely to be satisfied when the first four factors are smaller, that is, when fitness trade-offs are stronger. Thus, the additional polymorphism condition can be summarized by saying that fitter-allele dominance and stronger fitness trade-offs promote polymorphism.

Haplodiploid expression

The analysis for haplodiploid expression is a straightforward interpolation of those for haploid and diploid expression. For $n_R = 1$, polymorphism at L is not maintained. For $n_L = 1$, polymorphism at R may be maintained only if there is overdominance in the $f[c_{i/k,1}]$'s. Henceforth, assume there is no overdominance or underdominance in the $f[c_{i/k,j}]$'s for any j . For $n_R = n_L = 2$, assuming

$c_{1/1,1} = \max\{c_{i/k,j}\}$ and $\hat{c} < \min\{c_{i/k,j}\}$, the polymorphism conditions are

$$c_{2/2,2} > c_{2/2,1}, \quad (9a)$$

$$f[c_{2/2,2}] < f[c_{1/1,2}], \quad (9b)$$

and

$$\frac{c_{1/1,2}}{c_{1/1,1}} \frac{c_{2/2,1}}{c_{2/2,2}} \frac{f[c_{1/1,1}]}{f[c_{2/2,1}]} \frac{f[c_{2/2,2}]}{f[c_{1/1,2}]} \frac{1 - \phi_1}{1 + \phi_1} \frac{1 - \phi_2}{1 + \phi_2} < 1, \quad (10)$$

where

$$f[c_{1/2,1}] = \frac{1 + \phi_1}{2} f[c_{2/2,1}] + \frac{1 - \phi_1}{2} f[c_{1/1,1}] \quad (11a)$$

and

$$f[c_{1/2,2}] = \frac{1 + \phi_2}{2} f[c_{1/1,2}] + \frac{1 - \phi_2}{2} f[c_{2/2,2}]. \quad (11b)$$

There must be fitness trade-offs between the alleles at each locus, and fitter-allele dominance in the receptor and stronger fitness trade-offs promote polymorphism.

Discussion

Given ligand–receptor conflict, maintenance of diallelic polymorphism requires fitness trade-offs and is promoted by fitter-allele dominance and stronger fitness trade-offs. These findings are intriguing from a theoretical perspective and challenging from an empirical one.

Fitness trade-offs

As remarked above, polymorphism maintenance obviously requires fitness trade-offs. However, for diploid or haplodiploid expression, mere existence of these trade-offs may not be enough. If there is net fitter-allele recessiveness, then they must be strong enough to overcome it. Moreover, even for haploid expression, stronger fitness trade-offs promote polymorphism, because they make the boundary loop more repulsive. The stronger the repulsion, the shorter the periods alleles spend at low frequencies, where they are vulnerable to genetic drift in finite populations.

Ligand–receptor binding seems likely to involve several amino acid residues per protein. Some residues may be mutable to several alternatives that change the local shape but not the global structure of the protein, leading to many functional but distinguishable alleles. Variation in local shape seems likely to result in trade-offs in binding strength; for example, a ligand allele that fits one receptor allele well may fit most other receptor alleles poorly. If these intuitions are correct, then fitness trade-offs should be common.

Fitter-allele dominance

The effect of dominance on polymorphism maintenance is less obvious. For simplicity, studies of sexual selection have emphasized haploid expression (e.g., Kirkpatrick, 1982; Bulmer, 1989; Barton and Turelli, 1991; but see Gomulkiewicz and Hastings, 1990). Diploid expression might be expected to be qualitatively similar, unless there were overdominance or underdominance. The present study demonstrates that diploid expression can be qualitatively different, even if there is neither overdominance nor underdominance.

As defined above, fitter-allele dominance supposes intermediate dominance, but it naturally extends to encompass overdominance, in that a superior heterozygote more closely resembles the fitter of the corresponding homozygotes. For $n_R = n_L = 2$, it can be shown that, as might be expected, overdominance in the $c_{i/i,j/\ell}$'s or $f[c_{i/k,j/j}]$'s promotes polymorphism. Extended to encompass overdominance, fitter-allele dominance is a generalization of overdominance relevant to sexual conflict and perhaps other kinds of frequency-dependent selection.

Fitter-allele dominance or recessiveness might be typical of a ligand or its receptor due to their biochemistry. For example, suppose the ligand is a dimer (e.g., lysin, Kresge et al., 2001). For diploid expression, half the ligand molecules a heterozygous male produces are presumably heterodimers. If heterodimers function well, because only one subunit of a dimer must bind strongly to a receptor molecule, then fitter-allele dominance seems likely, but if heterodimers function poorly, because both subunits must bind strongly, then fitter-allele recessiveness seems likely.

More alleles

Conditions for maintenance of polyallelic (more than two alleles per locus) polymorphism constitute a challenge for future research, but it is already clear that an arbitrarily large number of alleles can be maintained. For example, Gavrilets and Waxman (2002) considered haploid expression with $c_{i,j}$ a decreasing function of $|i - j|$. In some simulations, more than 50 alleles were maintained. Under certain conditions, the population split into several substantially reproductively isolated subpopulations.

Fitness trade-offs are obviously necessary for polymorphism maintenance, but they can be more complicated than those built into Gavrilets and Waxman's model. For example, for haploid expression, simulations indicate that three alleles per locus can be maintained with $c_{1,1} > c_{1,3} > c_{1,2}$, $c_{2,2} > c_{2,3} > c_{2,1}$, and $c_{3,1} \approx c_{3,2} \approx c_{3,3}$. In particular, L_3 can be maintained even if it is not the most compatible ligand allele with any receptor allele. Such possibilities make characterizing fitness trade-offs among more than two alleles challenging.

For diploid expression, the effect of dominance is not wholly clear. For two alleles per locus, there is one doubly heterozygous compatibility $c_{1/2,1/2}$, which does not enter the polymorphism conditions, but for more than two alleles per locus, there are more such compatibilities, which presumably do enter the polymorphism conditions; for example, the dynamics of L_3 near an equilibrium of L_1 , L_2 , R_1 , and R_2 presumably depend on $c_{1/2,1/3}$ and $c_{1/2,2/3}$. Doubly heterozygous compatibilities can be characterized by both χ -like and ϕ -like parameters, but it is not obvious that this characterization is helpful.

In simulations with randomly generated parameter values (but without overdominance or underdominance), maintenance of polyallelic polymorphism is correlated with rough-and-ready measures of fitness trade-offs and fitter-allele dominance. For example, let $\chi_{i,j/\ell}$ be the resemblance of L_j/L_ℓ to the more fit of L_j/L_j and L_ℓ/L_ℓ in interactions with R_i/R_i (as defined above, $\chi_1 = \chi_{1,1/2}$ and $\chi_2 = \chi_{2,1/2}$), and as a measure of fitter-allele dominance among ligand alleles, let $\bar{\chi}$ be the average of $\chi_{i,j/\ell}$. In simulations with two, three, or four alleles per locus, $\bar{\chi}$ is correlated with maintenance of all alleles.

More loci

Mating frequency might depend in a complicated way on several proteins at once, but one simple scheme is probably common, particularly for gamete proteins: a sequence of pairs of proteins mediating a sequence of steps toward mating, such that if binding between the members of a pair is too weak, then the step the pair mediates is not taken, and subsequent steps are not attempted. For example, in abalones, a meeting of a sperm and egg begins with exocytosis of the sperm acrosome (releasing lysin) in the egg jelly coat, continues with penetration of the sperm through the egg vitelline envelope (mediated by lysin and VERL), and ends with fusion of the sperm and egg membranes (Kresge et al., 2001); progress toward fertilization may halt at any of these steps.

The simplest scheme is a sequence of two pairs of proteins. Call the first pair L' and R' and the second pair L'' and R'' . For haploid expression of all four proteins, let $c'_{i,j}$ be the frequency of success in the first step toward mating when an R'_i -female and an L'_j -male meet, and let $c''_{k,\ell}$ be the frequency of mating when an R''_k -female and an L''_ℓ -male meet, and the first step toward mating succeeds. It follows that the frequency of mating when an $R'_i R''_k$ -female and an $L'_j L''_\ell$ -male meet is $c'_{i,j} c''_{k,\ell}$. Complete analysis of this model is beyond the scope of the present study, but simulations indicate that if there are fitness trade-offs between the alleles at each locus, then diallelic polymorphism may be maintained at all four loci.

Empirical examples

It is plausible that sperm competition and polyspermy lead to selection on lysin, bindin, their receptors, and several other gamete proteins studied to date, including the sperm proteins sp18 in abalones (Kresge et al., 2001), lysin and TMAP in tegulas (Hellberg and Vacquier, 1999; Hellberg et al., 2000), and M7 lysin in mussels (Riginos and McDonald, 2002) and the egg proteins ZP2 and ZP3 in mammals (Swanson et al., 2001). These proteins exhibit rapid evolution driven by selection, and they are probably expressed only in gametes, so the selection is probably through gamete interactions. Sperm competition is probably common, and despite polyspermy blocks, polyspermy is not necessarily rare, even in broadcast spawners (Brawley, 1992; Pearson and Brawley, 1996;

Serrão et al., 1999; Yund, 2000; Franke et al., 2002). However, experiments are needed to assess whether, as the average compatibility of an egg (i.e., of the receptor genotype it expresses) with the sperms it meets (i.e., with the ligand genotypes they express) increases, the increasing risk of becoming polyspermic eventually outweighs the decreasing risk of going unfertilized.

Bindin exhibits rapid evolution and high polymorphism in some groups of species but slow evolution and low polymorphism in others (Metz et al., 1998; Zigler and Lessios, 2003). Often, in groups where bindin evolves rapidly, species are sympatric, and heterospecific gametes are incompatible, whereas in groups where bindin evolves slowly, species are allopatric, and heterospecific gametes are compatible. This pattern suggests reinforcement (Servedio and Noor, 2003), where the bindin receptor would evolve to avoid heterospecific fertilizations yielding dysfunctional hybrids, and bindin would evolve to follow its receptor (Geyer and Palumbi, 2003). However, sympatry may be not a cause but a consequence of bindin evolution. In groups where bindin evolves rapidly, the resulting gametic incompatibility may be what enables these species to coexist in sympatry (Zigler and Lessios, 2003). In groups where bindin evolves slowly, it may be because the ecology of the organisms does not engender ligand–receptor conflict; for example, sperm limitation may be more prevalent than sperm competition, or gametes may not be well mixed.

There is evidence for fitness trade-offs in bindin and lysin. In experiments with the sea urchin *Echinometra mathaei*, which of two males did better at fertilizing a given female often depended on the female, and males with similar bindin alleles usually had similar fertilization success (Palumbi, 1999). In abalones, the near monomorphism of lysin precludes such comparisons within species, but males do better at fertilizing conspecific females, partly because conspecific lysin–VERL binding is stronger (Kresge et al., 2001). There is no evidence regarding fitness trade-offs in the bindin receptor or VERL. Such evidence might emerge from experiments aimed at assessing whether these proteins indeed mediate sexual conflict.

The present study suggests that expression of lysin and/or VERL may be diploid with a tendency toward fitter-allele recessiveness, whereas expression of bindin and/or its receptor may be

either haploid or diploid with a tendency toward fitter-allele dominance. There is evidence for diploid expression and a tendency toward fitter-allele dominance in *bindin*. In experiments with *E. mathaei*, *bindin* alleles belonged to two major clades, A and B, and the fertilization success of an AB-male with a given female was usually similar to that of an AA-male or BB-male, whichever did better (Palumbi, 1999). There is no evidence regarding ploidy of expression or patterns of dominance in the *bindin* receptor, *lysin*, or *VERL*. Preliminary data suggest that *lysin* heterodimers function well (V. D. Vacquier, personal communication), which might result in fitter-allele dominance. However, *lysin* expression may be haploid. Experiments are needed to resolve these issues.

In *Drosophila*, there is evidence for sexual conflict mediated by reproductive proteins, in that females die sooner after mating with males that do better in sperm competition (Civetta and Clark, 2000). Exactly which proteins mediate this conflict is unclear, but they probably include male seminal fluid and female reproductive tract proteins (Chapman, 2001; Wolfner, 2002). This situation may be more complicated than those envisioned in the present study, in that a ligand may bind to other ligands and/or multiple receptors. As experiments reveal these mechanisms, modeling their evolutionary consequences will become appropriate. Polymorphism maintenance will presumably be found to require fitness trade-offs, and because delays in dynamical systems are often destabilizing, fitter-allele dominance and stronger fitness trade-offs may well be found to promote polymorphism. It will be interesting to see whether these expectations are fulfilled in fruitflies and other taxa where sexual conflict is detected as research on reproductive proteins expands.

Appendix

Haploid expression

Given (3a) and (3b) and the assumptions preceding them, the fixation equilibria are contained in a heteroclinic cycle. The goal of what follows is to formulate the condition for the cycle to be repelling. Immediately, an obstacle arises: equation (1) is a discrete-time system, but relatively little is known about heteroclinic cycles in discrete time. To make progress, consider the continuous-time system $\dot{z}_{ij} = z'_{ij} - z_{ij}$, which the discrete-time system approaches when selection is weak.

The desired condition involves the eigenvalues of the linearized system at each fixation equilibrium. At the equilibrium where R_i and L_j are fixed, the three eigenvalues are

$$\lambda_{i,j}^{(1)} = \frac{1}{2} \left(\frac{f[c_{k,j}]}{f[c_{i,j}]} - 1 \right), \quad (\text{A1a})$$

$$\lambda_{i,j}^{(2)} = \frac{1}{2} \left(\frac{c_{i,\ell}}{c_{i,j}} - 1 \right), \quad (\text{A1b})$$

and

$$\lambda_{i,j}^{(3)} = (1 - r)(\lambda_{i,j}^{(1)} + \lambda_{i,j}^{(2)}) - r, \quad (\text{A1c})$$

where $k = (i \bmod 2) + 1$ and $\ell = (j \bmod 2) + 1$. Either $\lambda_{i,j}^{(1)} > 0$ and $\lambda_{i,j}^{(2)} < 0$, or vice versa. At this point, another obstacle arises: if r is small, then $\lambda_{i,j}^{(3)}$ may be positive. The cycle is then actually a network, but relatively little is known about heteroclinic networks. To make progress, assume r is large enough that all $\lambda_{i,j}^{(3)}$'s are negative. Given weak selection, $\lambda_{i,j}^{(1)} + \lambda_{i,j}^{(2)}$ is small, so this assumption is not stringent.

The desired condition is the condition for the intersection of the cycle with a Poincaré section to be an unstable equilibrium of the corresponding Poincaré map. By standard reasoning (e.g., Hofbauer and Sigmund, 1998, secs. 17.1–17.3), the condition is

$$\prod_{i,j}^4 \frac{\alpha_{i,j}}{\rho_{i,j}} < 1, \quad (\text{A2})$$

where $\alpha_{i,j} = \min\{\lambda_{i,j}^{(1)}, \lambda_{i,j}^{(2)}\}$ and $\rho_{i,j} = \max\{\lambda_{i,j}^{(1)}, \lambda_{i,j}^{(2)}\}$. Routine algebra gives

$$\frac{c_{1,2}}{c_{1,1}} \frac{c_{2,1}}{c_{2,2}} \frac{f[c_{1,1}]}{f[c_{2,1}]} \frac{f[c_{2,2}]}{f[c_{1,2}]} < 1, \quad (\text{A3})$$

which is always satisfied. The cycle is presumably repelling even if some $\lambda_{i,j}^{(3)}$'s are positive; additional unstable eigenvalues should make the cycle more repulsive.

Diploid expression

The reasoning for diploid expression is similar. The difference is that at the equilibrium where R_i and L_j are fixed, six of the nine eigenvalues are -1 , and the other three are

$$\lambda_{i,j}^{(1)} = \frac{1}{2} \left(\frac{f[c_{i/k,j/j}]}{f[c_{i/i,j/j}]} - 1 \right), \quad (\text{A4a})$$

$$\lambda_{i,j}^{(2)} = \frac{1}{2} \left(\frac{c_{i/i,j/\ell}}{c_{i/i,j/j}} - 1 \right), \quad (\text{A4b})$$

and

$$\lambda_{i,j}^{(3)} = (1 - r)(\lambda_{i,j}^{(1)} + \lambda_{i,j}^{(2)}) - r. \quad (\text{A4c})$$

Reasoning as before yields (8).

Acknowledgments

I thank L. B. Geyer, J. H. Gillespie, R. K. Grosberg, C. H. Langley, D. L. Leighton, S. R. Palumbi, M. Turelli, and V. D. Vacquier for helpful conversations, the referees for helpful suggestions, and R. Bürger for an invitation to visit the Erwin Schrödinger International Institute for Mathematical Physics, where I did part of this research. This research was partly supported by funds from the Center for Population Biology at the University of California, Davis and by Nation Science Foundation grants DBI-9602226 and DEB-0089716.

References

- Arnqvist, G., and Rowe, L., 2002. Antagonistic coevolution between the sexes in a group of insects. *Nature* **415**:787–789.
- Barton, N. H., and Turelli, M., 1991. Natural and sexual selection on many loci. *Genetics* **127**:229–255.
- Biermann, C. H., 1998. The molecular evolution of sperm bindin in six species of sea urchins (Echinoidea: Strongylocentrotidae). *Molecular Biology and Evolution* **15**:1761–1771.

- Brawley, S. H., 1992. Fertilization in natural populations of the dioecious brown alga *Fucus ceranoides* and the importance of the polyspermy block. *Marine Biology* **113**:145–157.
- Bulmer, M., 1989. Structural instability of models of sexual selection. *Theoretical Population Biology* **35**:195–206.
- Chapman, T., 2001. Seminal fluid-mediated fitness traits in *Drosophila*. *Heredity* **87**:511–521.
- Chapman, T., Arnqvist, G., Bangham, J., and Rowe, L., 2003. Sexual conflict. *Trends in Ecology and Evolution* **18**:41–47.
- Civetta, A., and Clark, A. G., 2000. Correlated effects of sperm competition and postmating female mortality. *Proceedings of the National Academy of Sciences of the United States of America* **97**:13162–13165.
- Frank, S., 2000. Sperm competition and female avoidance of polyspermy mediated by sperm–egg biochemistry. *Evolutionary Ecology Research* **2**:613–625.
- Franke, E. S., Babcock, R. C., and Styan, C. A., 2002. Polyspermy under sperm-limited conditions. *American Naturalist* **160**:485–496.
- Galindo, B. E., Vacquier, V. D., and Swanson, W. J., 2003. Positive selection in the egg receptor for abalone sperm lysin. *Proceedings of the National Academy of Sciences of the United States of America* **100**:4639–4643.
- Gavrilets, S., 2000. Rapid evolution of reproductive barriers driven by sexual conflict. *Nature* **403**:886–889.
- Gavrilets, S., Arnqvist, G., and Friberg, U., 2001. The evolution of female mate choice by sexual conflict. *Proceedings of the Royal Society of London B, Biological Sciences* **268**:531–539.
- Gavrilets, S., and Waxman, D., 2002. Sympatric speciation by sexual conflict. *Proceedings of the National Academy of Sciences of the United States of America* **99**:10533–10538.

- Geyer, L. B., and Palumbi, S. R., 2003. Reproductive character displacement and the genetics of gamete recognition in tropical sea urchins. *Evolution* **57**:1049–1060.
- Gomulkiewicz, R. S., and Hastings, A., 1990. Ploidy and evolution by sexual selection: A comparison of haploid and diploid female choice models near fixation equilibria. *Evolution* **44**:757–770.
- Hecht, N. B., 1998. Molecular mechanisms of male germ cell differentiation. *BioEssays* **20**:555–561.
- Hellberg, M. E., Moy, G. W., and Vacquier, V. D., 2000. Positive selection and propeptide repeats promote rapid interspecific divergence of a gastropod sperm protein. *Molecular Biology and Evolution* **17**:458–466.
- Hellberg, M. E., and Vacquier, V. D., 1999. Rapid evolution of fertilization selectivity and lysin cDNA sequences in teguline gastropods. *Molecular Biology and Evolution* **16**:839–848.
- Hofbauer, J., and Sigmund, K., 1998. *Evolutionary Games and Population Dynamics*. Cambridge University Press, Cambridge, UK.
- Kirkpatrick, M., 1982. Sexual selection and the evolution of female choice. *Evolution* **36**:1–12.
- Kresge, N., Vacquier, V. D., and Stout, C. D., 2001. Abalone lysin: The dissolving and evolving sperm protein. *BioEssays* **23**:95–103.
- Metz, E. C., Gomez-Gutierrez, G., and Vacquier, V. D., 1998. Mitochondrial DNA and binding gene sequence evolution among allopatric species of the sea urchin genus *Arbacia*. *Molecular Biology and Evolution* **15**:185–195.
- Metz, E. C., Kane, R. E., Yanagimachi, H., and Palumbi, S. R., 1994. Fertilization between closely related sea urchins is blocked by incompatibilities during sperm-egg attachment and early stages of fusion. *Biological Bulletin* **187**:23–34.

- Metz, E. C., and Palumbi, S. R., 1996. Positive selection and sequence rearrangements generate extensive polymorphism in the gamete recognition protein bindin. *Molecular Biology and Evolution* **13**:397–406.
- Murray, J. D., 1989. *Mathematical Biology*, 2nd ed. Springer-Verlag, Berlin.
- Palumbi, S. R., 1999. All males are not created equal: Fertility differences depend on gamete recognition polymorphisms in sea urchins. *Proceedings of the National Academy of Sciences of the United States of America* **96**:12632–12637.
- Parker, G. A., and Partridge, L., 1998. Sexual conflict and speciation. *Philosophical Transactions of the Royal Society of London B, Biological Sciences* **353**:261–274.
- Pearson, G. A., and Brawley, S. H., 1996. Reproductive ecology of *Fucus distichus* (Phaeophyceae): An intertidal algae with successful external fertilization. *Marine Ecology Progress Series* **143**:211–223.
- Rice, W. R., and Holland, B., 1997. The enemies within: Intergenomic conflict, interlocus contest evolution (ICE), and the intraspecific Red Queen. *Behavioral Ecology and Sociobiology* **41**:1–10.
- Riginos, C., and McDonald, J. H., 2002. Positive selection on an acrosomal sperm protein, M7 lysin, in three species of the mussel genus *Mytilus*. *Molecular Biology and Evolution* **20**:200–207.
- Rowe, L., Arnqvist, G., Sih, A., and Krupa, J. J., 1994. Sexual conflict and the evolutionary ecology of mating patterns: Water striders as a model system. *Trends in Ecology and Evolution* **9**:289–293.
- Serrão, E. A., Brawley, S. H., Hedman, J., Kautsky, L., and Samuelsson, G., 1999. Reproductive success of *Fucus vesiculosus* (Phaeophyceae) in the Baltic Sea. *Journal of Phycology* **35**:254–269.

- Servedio, M. R., and Noor, M. A. F., 2003. The role of reinforcement in speciation: Theory and data. *Annual Review of Ecology and Systematics* **34**:339–364.
- Styan, C. A., 1998. Polyspermy, egg size, and the fertilization kinetics of free-spawning marine invertebrates. *American Naturalist* **152**:290–297.
- Swanson, W. J., and Vacquier, V. D., 2002. The rapid evolution of reproductive proteins. *Nature Reviews Genetics* **3**:137–144.
- Swanson, W. J., Yang, Z., Wolfner, M. F., and Aquadro, C. F., 2001. Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. *Proceedings of the National Academy of Sciences of the United States of America* **98**:2509–2514.
- Vacquier, V. D., Swanson, W. J., and Hellberg, M. E., 1995. What have we learned about sea urchin sperm bindin? *Development, Growth, and Differentiation* **37**:1–10.
- Weir, B. S., and Cockerham, C. C., 1989. Complete characterization of disequilibrium at two loci. Feldman, M. W. (Ed.), *Mathematical Evolutionary Theory*, pp. 86–110. Princeton University Press, Princeton, NJ.
- Wolfner, M. F., 2002. The gifts that keep on giving: Physiological functions and evolutionary dynamics of male seminal proteins in *Drosophila*. *Heredity* **88**:85–93.
- Yund, P. O., 2000. How severe is sperm limitation in natural populations of marine free-spawners? *Trends in Ecology and Evolution* **15**:10–13.
- Zigler, K. S., and Lessios, H. A., 2003. Evolution of bindin in the pantropical sea urchin *Tripneustes*: Comparisons to bindin of other genera. *Molecular Biology and Evolution* **20**:220–231.

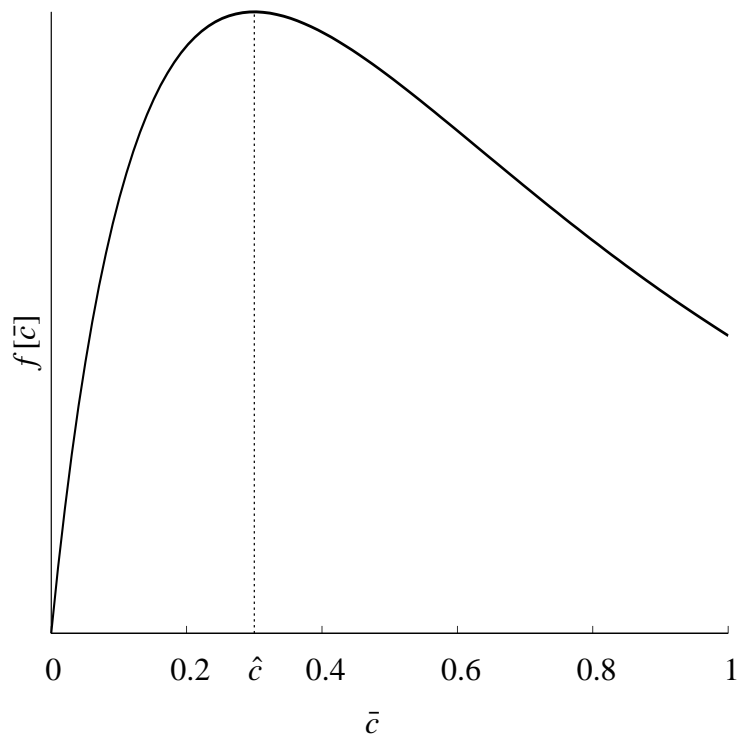


Figure 1: A possible f , $f[\bar{c}] \propto \left(1 - \frac{1}{1 + (\bar{c}/\hat{c})}\right) \exp[-\bar{c}/(2\hat{c})]$, with $\hat{c} = 0.3$. (The vertical scale is irrelevant, because selection depends on relative fitness.)

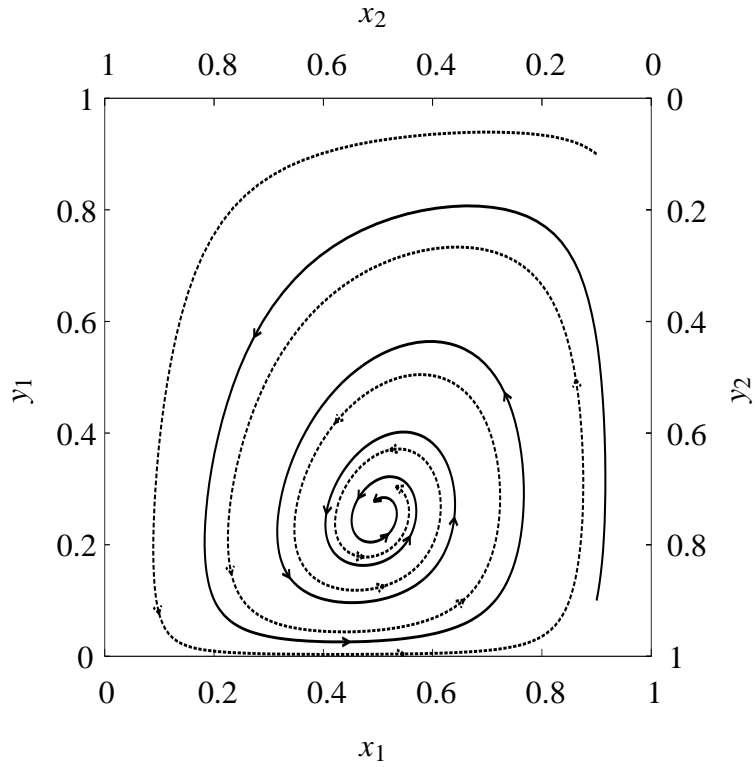


Figure 2: An example of polymorphism maintenance for haploid expression, with f as in Figure 1, $c_{1,1} = 0.7$, $c_{1,2} = 0.5$, $c_{2,1} = 0.4$, $c_{2,2} = 0.6$, and $r = 0.5$. (Each trajectory lasts 500 generations, and an arrowhead is drawn every 50 generations. The dynamics are discrete-time, but continuous curves are drawn for ease of reading.)

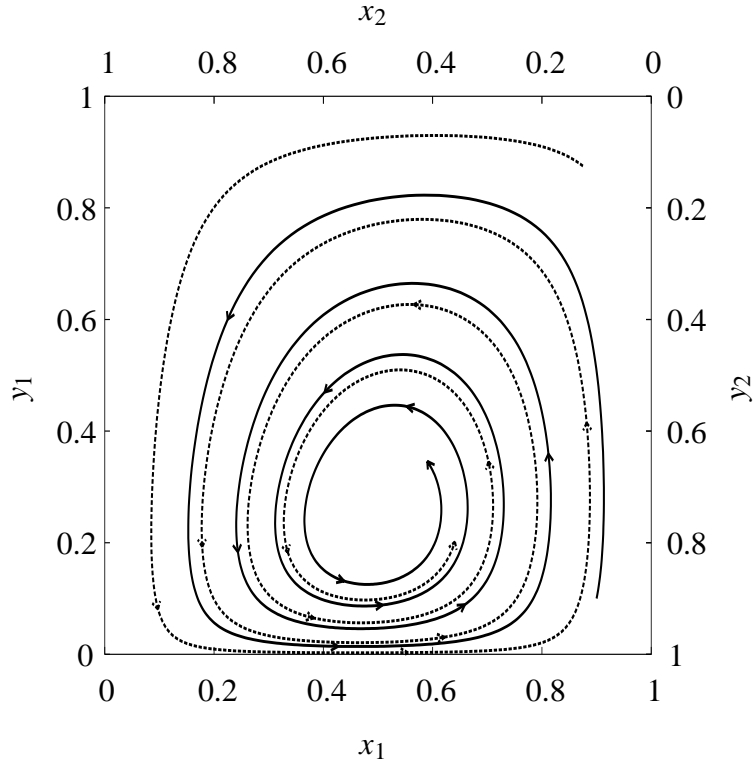


Figure 3a: An example of polymorphism maintenance for diploid expression, with f as in Figure 1, $c_{1/1,1/1} = 0.7$, $c_{1/1,1/2} = 0.6$, $c_{1/1,2/2} = 0.5$, $c_{1/2,1/1} = 0.55$, $c_{1/2,1/2} = 0.55$, $c_{1/2,2/2} = 0.55$, $c_{2/2,1/1} = 0.4$, $c_{2/2,1/2} = 0.5$, $c_{2/2,2/2} = 0.6$, and $r = 0.5$; it follows that $\chi_1 = 0$, $\chi_2 = 0$, $\phi_1 = 0.08$, and $\phi_2 = 0.02$ (see (7a)–(7d)). (Each trajectory lasts 1000 generations, and an arrowhead is drawn every 100 generations.)

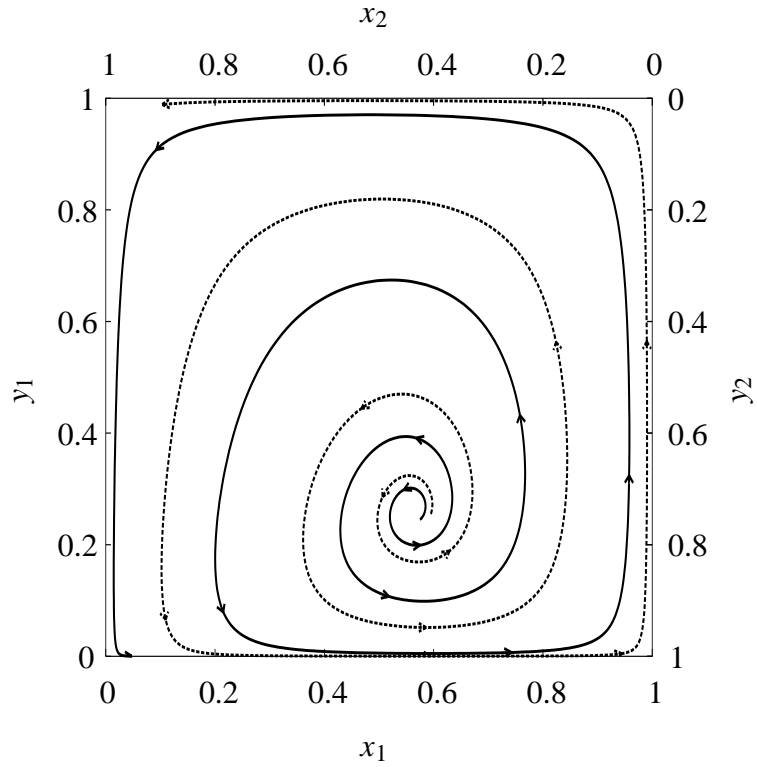


Figure 3b: An example of polymorphism decay for diploid expression, with parameter values as in Figure 3a, except $c_{1/1,1/2} = 0.55$ and $c_{2/2,1/2} = 0.45$; it follows that $\chi_1 = -0.5$ and $\chi_2 = -0.5$.

notation	definition
\hat{c}	optimal compatibility for receptor genotypes ($0 < \hat{c} < 1$)
\bar{c}_i	for haploid expression, average compatibility of R_i
$\bar{c}_{i/k}$	for diploid expression, average compatibility of R_i/R_k
$c_{i,j}$	for haploid expression, compatibility between R_i and L_j ($0 < c_{i,j} \leq 1$)
$c_{i/k,j}$	for haplodiploid expression, compatibility between R_i/R_k and L_j ($0 < c_{i/k,j} \leq 1$)
$c_{i/k,j/\ell}$	for diploid expression, compatibility between R_i/R_k and L_j/L_ℓ ($0 < c_{i/k,j/\ell} \leq 1$)
D	linkage disequilibrium in gametes (after recombination)
$f[\bar{c}]$	fitness of receptor genotype with average compatibility \bar{c}
L	ligand locus
L_j	j th ligand allele ($1 \leq j \leq n_L$)
n_L	number of ligand alleles
n_R	number of receptor alleles
R	receptor locus
R_i	i th receptor allele ($1 \leq i \leq n_R$)
r	frequency of recombination between L and R ($0 < r \leq 1/2$)
U_i	for haploid expression, fitness of R_i
$U_{i/k}$	for diploid expression, fitness of R_i/R_k
V_j	for haploid expression, fitness of L_j
$V_{j/\ell}$	for diploid expression, fitness of L_j/L_ℓ
\bar{W}	mean fitness

(continued on next page)

x_i	frequency of allele R_i
$x_{i/i}$	frequency of genotype R_i/R_i
$x_{i/k}$	1/2 frequency of genotype R_i/R_k for $i \neq k$
y_j	frequency of allele L_j
$y_{j/j}$	frequency of genotype L_j/L_j
$y_{j/\ell}$	1/2 frequency of genotype L_j/L_ℓ for $j \neq \ell$
z_{ij}	frequency of haplotype R_iL_j
$z_{ij/ij}$	frequency of diplotype R_iL_j/R_iL_j
$z_{ij/k\ell}$	1/2 frequency of diplotype R_iL_j/R_kL_ℓ for $i \neq k$ or $j \neq \ell$
ϕ_1	resemblance of R_1/R_2 to R_2/R_2 in interactions with L_1/L_1 ($-1 \leq \phi_1 \leq 1$)
ϕ_2	resemblance of R_1/R_2 to R_1/R_1 in interactions with L_2/L_2 ($-1 \leq \phi_2 \leq 1$)
χ_1	resemblance of L_1/L_2 to L_1/L_1 in interactions with R_1/R_1 ($-1 \leq \chi_1 \leq 1$)
χ_2	resemblance of L_1/L_2 to L_2/L_2 in interactions with R_2/R_2 ($-1 \leq \chi_2 \leq 1$)

Table 1: Glossary of notation. A notation is glossed if it is defined in the main text and used outside the context in which it is defined. Ligand alleles are indexed (from 1 to n_L) by j , ℓ , n , or p , and receptor alleles are indexed (from 1 to n_R) by i , k , m , or o .