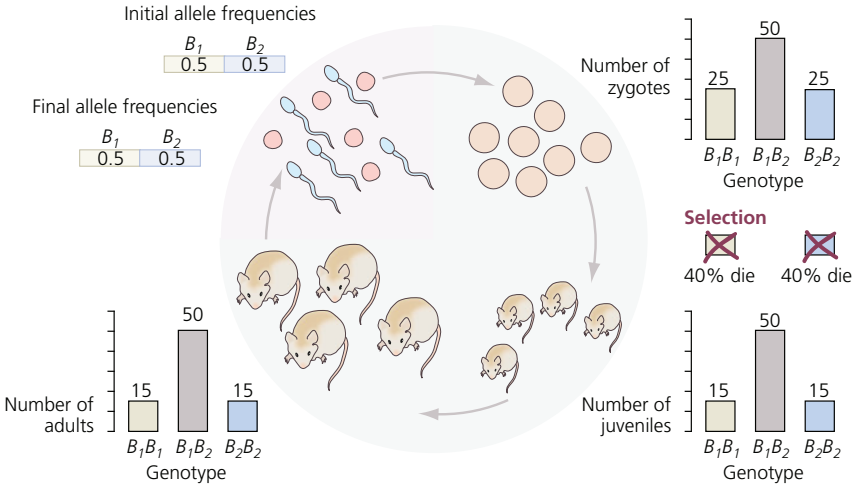

Figure 6.15 Selection can change genotype frequencies so that they cannot be calculated by multiplying the allele frequencies When 40% of the homozygotes in this population die, the allele frequencies do not change. But among the survivors, there are more heterozygotes than predicted under Hardy–Weinberg equilibrium.

We used strong selection in our numerical example to make a point. In fact, selection is rarely strong enough to produce, in a single generation, such a large violation of Hardy–Weinberg conclusion 2. Even if it does, a single bout of random mating will immediately put the genotypes back into Hardy–Weinberg equilibrium. Nonetheless, researchers sometimes find violations of Hardy–Weinberg conclusion 2 that seem to be the result of selection.

Empirical Research on Selection and Genotype Frequencies

Our example comes from research by Atis Muehlenbachs and colleagues (2008), working in the laboratory of Patrick Duffy, on genetic variation for the outcome of falciparum malaria during pregnancy. Falciparum malaria is caused by infection with the single-celled parasite *Plasmodium falciparum*. When a pregnant woman contracts the disease, the parasites invade the placenta via the mother's circulatory system (Karumanchi and Haig 2008). This triggers placental inflammation and may also interfere with placental development (Umbers et al. 2011). The potential complications include spontaneous abortion, premature delivery, low birth weight, and higher risk of infant death.

Pregnancy itself brings an increased risk of malaria infection, particularly a woman's first pregnancy (Karumanchi and Haig 2008). During a first bout of placental malaria, women develop antibodies that confer partial resistance during later pregnancies. Some 125 million women who live in areas affected by malaria become pregnant each year, and malaria infection during pregnancy is estimated to be responsible for an annual toll of 100,000 infant deaths (Umbers et al. 2011).

Muehlenbachs and colleagues (2008) suspected that the outcome of placental malaria hinges on the fetus's genotype at the locus encoding vascular endothelial growth factor receptor 1 (VEGFR1), also known as *fms*-like tyrosine kinase 1 (Flt1). Fetal cells in the placenta release a soluble form of this protein, sVEGFR1, into the mother's circulation. By interacting with vascular endothelial growth factor, VEGFR1 influences both placental development and inflammation.

Copies of the gene for VEGFR1 vary in the length of a two-nucleotide repeat in a region that is transcribed to mRNA but not translated. Alleles cluster into a short group (*S* alleles) and a long group (*L* alleles). Cultured cord blood cells with genotypes *SS* and *SL* produce more VEGFR1 than do *LL* cells.

COMPUTING CONSEQUENCES 6.4

Statistical analysis of allele and genotype frequencies using the χ^2 (chi-square) test

Here we use data from Muehlenbachs and colleagues (2008) to illustrate a method for determining whether genotype frequencies deviate from Hardy–Weinberg equilibrium. The researchers surveyed Tanzanian infants born to first-time mothers during malaria season. The genotype counts (provided by Atis Muehlenbachs and Patrick Duffy, personal communication) were

SS	*SL*	*LL*
16	50	10

The analysis proceeds in five steps:

1. Calculate the allele frequencies. The sample of 76 infants is also a sample of 152 gene copies. All 32 copies carried by the *SS* infants are *S*, as are 50 of the copies carried by the *SL* infants. Thus, the frequency of *S* is

$$\frac{32 + 50}{152} = 0.54$$

The frequency of *L* is

$$\frac{50 + 20}{152} = 0.46$$

2. Calculate the genotype frequencies expected under Hardy–Weinberg equilibrium. If the frequencies of two alleles are p and q, then the expected frequencies of the genotypes are p^2, $2pq$, and q^2. The expected frequencies among the infants are thus

SS	*SL*	*LL*
$0.54^2 = 0.29$	$2 \cdot 0.54 \cdot 0.46 = 0.5$	$0.46^2 = 0.21$

3. Calculate the expected number of infants of each genotype under Hardy–Weinberg equilibrium. This is simply the expected frequency of each genotype multiplied by the total number of infants, 76:

SS	*SL*	*LL*
$0.29 \cdot 76 = 22$	$0.5 \cdot 76 = 38$	$0.21 \cdot 76 = 16$

These expectations are different from the numbers observed (16, 50, and 10). The actual sample contains more heterozygotes and fewer homozygotes. Is it plausible that a difference this large between expectation and reality could arise by chance? Or is the difference statistically significant? Our null hypothesis is that the difference is simply due to chance.

Working with newborn babies of first-time mothers in Muheza, Tanzania, where malaria is a perennial scourge, Muehlenbachs and colleagues (2008) tested their hypothesis in part by using the Hardy–Weinberg equilibrium principle.

The researchers first determined the allele frequencies among 163 infants born from October through April, when the rate of placental malaria was at its annual low. The frequencies were

S	*L*
0.555	0.445

If the population of infants was in Hardy–Weinberg equilibrium, then multiplying these allele frequencies will allow us to predict the genotype frequencies:

SS	*SL*	*LL*
$0.555^2 = 0.308$	$2 \cdot 0.555 \cdot 0.445 = 0.494$	$0.445^2 = 0.198$

These predicted frequencies are, in fact, close to the actual genotype frequencies among the off-season infants:

SS	*SL*	*LL*
$\frac{49}{163} = 0.301$	$\frac{83}{163} = 0.509$	$\frac{31}{163} = 0.190$

The true frequency of heterozygotes is slightly higher than predicted, and the frequencies of homozygotes are slightly lower, but the discrepancies are modest. The infants thus conform to conclusion 2 of the Hardy–Weinberg analysis.

4. Calculate a test statistic. We will use one devised in 1900 by Karl Pearson. It is called chi-square (χ^2).

$$\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

where the symbol $\sum$ indicates a sum taken across all the classes considered. In our data there are three classes: the three genotypes. For our data set

$$\chi^2 = \frac{(16 - 22)^2}{22} + \frac{(50 - 38)^2}{38} + \frac{(10 - 16)^2}{16} = 7.68$$

5. Determine whether the test statistic is significant. χ^2 is defined such that it gets larger as the difference between the observed and expected values gets larger. How likely is it that we could get a χ^2 as large as 7.68 by chance? Most statistical textbooks have a table giving the answer. In Zar's (1996) book. it is called "Critical values of the chi-square distribution."

To use this table, we need to calculate a number called the degrees of freedom for the test statistic. This value for χ^2 is the number of classes minus the number of independent values calculated from the data for use in determining the expected values. For our χ^2 there are three classes: the genotypes. We calculated two values from the data for use in determining the expected values: the total number of individuals, and the frequency of allele *S*. (We also calculated the frequency of *L*, but it is not independent of the frequency of *S*, because the two must sum to 1.) Thus the number of degrees of freedom is 1. (Another formula for calculating the degrees of freedom in χ^2 tests for Hardy–Weinberg equilibrium is

$$df = k - 1 - m$$

where k is the number of classes and m is the number of independent allele frequencies estimated.)

According to the table, the critical value of χ^2 for one degree of freedom and $P = 0.05$ is 3.841. This means there is a 5% chance under the null hypothesis of getting $\chi^2 \geq 3.841$. The probability under the null hypothesis of getting $\chi^2 \geq 7.68$ is therefore (considerably) less than 5%. We reject the null hypothesis and assert that our χ^2 is statistically significant at $P < 0.05$. (In fact, $P < 0.006$.)

The χ^2 test tells us that among infants born during malaria season, the alleles of the gene for VEGFR1 are not in Hardy–Weinberg equilibrium. This indicates that one or more assumptions of the Hardy–Weinberg analysis has been violated. By itself, however, it does not tell us which are being violated, or how.

Muehlenbachs and colleagues then determined the allele frequencies among 76 infants born from May through September, when the rate of placental malaria was at its annual high. The frequencies were nearly the same as among the off-season newborns:

S	*L*
0.539	0.461

If this segment of the population was, like their off-season counterparts, in Hardy–Weinberg equilibrium, then multiplying the allele frequencies will again allow us to predict the genotype frequencies:

SS	*SL*	*LL*
$0.539^2 = 0.291$	$2 \cdot 0.539 \cdot 0.461 = 0.497$	$0.461^2 = 0.213$

This time the predicted values are a poor fit to the actual frequencies:

SS	*SL*	*LL*
$\frac{16}{76} = 0.211$	$\frac{50}{76} = 0.658$	$\frac{10}{76} = 0.132$

There are substantially more heterozygotes than expected, and substantially fewer homozygotes. This discrepancy between prediction and data is statistically significant (see **Computing Consequences 6.4**). The genotypes of the infants born during peak malaria season are in violation of Hardy–Weinberg conclusion 2.

The discovery that genotype frequencies in a population are not in Hardy–Weinberg equilibrium may be a clue that natural selection is at work.

On this and other evidence, Muehlenbachs and colleagues (2008) believe the best explanation for the missing homozygotes is that they did not survive fetal development. A fetus's chance of surviving depends on both its own genotype and whether its mother contracts malaria **(Figure 6.16)**. If the mother does not contract malaria, *SS* infants do somewhat better than others. If, however, the mother does contract malaria, *SL* infants do substantially better than others. Overall, when malaria is common, heterozygotes survive at the highest rate. Consistent with this explanation, where malaria is absent, *S* alleles occur at high frequency.

Figure 6.16 Probability of fetal survival as a function of genotype and placental malaria Inferred from the patterns in maternal and newborn genotype frequencies in Muehlenbachs et al. (2008).

Changes in the Frequency of the *CCR5-Δ32* Allele Revisited

We are now in a position to give a more satisfying answer to the question we raised at the beginning of Section 6.1: Will the AIDS epidemic cause the frequency of the *CCR5-Δ32* allele to increase in human populations? The AIDS epidemic could, in principle, cause the frequency of the allele to increase rapidly, but at present it appears that it will probably not do so in any real population. This conclusion is based on the three model populations depicted in **Figure 6.17** (see **Computing Consequences 6.5** for the algebra). Each model is based on different assumptions about the initial frequency of the *CCR5-Δ32* allele and the prevalence of HIV infection. Each graph shows the predicted change in the frequency of the *Δ32* allele over 40 generations, or approximately 1,000 years.

The model population depicted in Figure 6.17a offers a scenario in which the frequency of the *Δ32* allele could increase rapidly. In this scenario, the initial frequency of the *CCR5-Δ32* allele is 20%. One-quarter of the individuals with genotype +/+ or +/*Δ32* contract AIDS and die without reproducing, whereas all of the *Δ32*/*Δ32* individuals survive. The 20% initial frequency of *Δ32* is approximately equal to the highest frequency reported for any population, a sample of Ashkenazi Jews studied by Martinson et al. (1997). The mortality rates approximate the situation in Botswana, Namibia, Swaziland, and Zimbabwe, where up to 25% of individuals between the ages of 15 and 49 are infected with HIV (UNAIDS 1998). In this model population, the frequency of the *Δ32* allele increases by as much as a few percentage points each generation. By the end of 40 generations, the allele is at a frequency of virtually 100%. Thus, in a human population that combined the highest reported frequency of the *Δ32* allele with the highest reported rates of infection, the AIDS epidemic could cause the frequency of the allele to increase rapidly.

At present, however, no known population combines a high frequency of the *Δ32* allele with a high rate of HIV infection. In northern Europe, many populations have *Δ32* frequencies between 0.1 and 0.2 (Martinson et al. 1997; Stephens et al. 1998), but HIV infection rates are under 1% (UNAIDS 1998). A model population reflecting these conditions is depicted in Figure 6.17b. The initial frequency of the *Δ32* allele is 0.2, and 0.5% of the +/+ and +/*Δ32* individuals contract AIDS and die without reproducing. The frequency of the *Δ32* allele hardly changes at all. Selection is too weak to cause appreciable evolution in such a short time.

In parts of sub-Saharan Africa, as many as a quarter of all individuals of reproductive age are infected with HIV. However, the *Δ32* allele is virtually absent (Martinson et al. 1997). A model population reflecting this situation is depicted in Figure 6.17c. The initial frequency of the *Δ32* allele is 0.01, and 25% of the +/+ and +/*Δ32* individuals contract AIDS and die without reproducing. Again, the frequency of the *Δ32* allele hardly changes at all. When the *Δ32* allele is at low

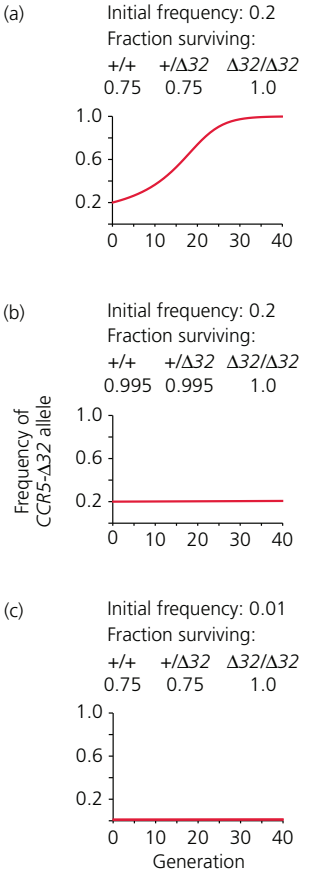

Figure 6.17 Predicted change in allele frequencies at the CCR5 locus under different scenarios (a) If the initial frequency of *CCR5-Δ32* is high and many people become infected with HIV, allele frequencies can change rapidly. (b) In Europe allele frequencies are high, but infection rates are low. (c) In parts of Africa the infection rates are high, but allele frequencies are low.