5 Inbreeding at Several Loci

5.1 Recombinant inbred Strains (RI)

are a special type of inbred strains. They are constructed by an outcross of two established inbred parental strains. From the second generation of F_2 animals, which are not genetically identical, a set of pairs is selected. Each selected pair is used in order to create a new inbred strain via repeated brother–sister mating. Eventually, a set of inbred strains is created, which is called a *recombinant inbred set*. Each strain in a given set is genetically homogeneous within, but is genetically different from the other strains in the set and from the original parental strains. Thus, at each locus for which the original strains differ, the new recombinant inbred strain is homozygous for one of the two alleles. However, the parental source of the allele may vary from one strain to the next within the recombinant inbred set, and at different loci the source of the allele may be different parental strains. For example, if the original parental strains are A/B and a/b homozygous at two loci (i.e., one strain is AA at one locus and BB at another, while the other strain is aa at the first locus and bb at the other), recombinant inbreds can be A/B, A/b, a/B, or a/b homozygous at those two loci. There are $2^3 = 8$ possibilities with three loci, $2^4 = 16$ with four, etc.

The name of the RI set is formed by combining the names of the two parental inbred strains. For example, if the F_1 mice are created by mating an A female and a C57BL/6J (B6) male, then the resulting recombinant inbred set is denoted AXB. If, on the other hand, the female is a B6 and the male is A then the set is called BXA.

Recombinant inbred sets can be used for mapping traits. At each locus the allele is equally likely to have come from each of the parental strains. A major advantage of recombinant inbred sets, and an important motivation for their establishment, is that genotypes are fixed for the entire set. Consequently, the experiment for mapping a trait with recombinant inbreds is conducted by phenotyping mice from the different strains that form the set. Genotyping is unnecessary for the commercially available sets, since their genotypes have already been determined. In addition, since mice from the same strain are genetically identical, one can average the phenotypes of a number of different mice from the same strain, which has the effect of decreasing the environmental variance relative to the genetic variance.

5.2 The Recombination Fraction

Crossovers between chromosomal segments may occur during meiosis. These crossovers cause the gamete that passes to an offspring to be a mosaic of genetic material originating from the two grandparents. A *recombination event* between two loci occurs whenever the number of crossovers is odd. In this case, the genetic material at one locus is from one grandparent, whereas the genetic material at the other locus is from the other grandparent. The probability of a recombination is known as the recombination fraction and is denoted by θ. The value of θ depends on the relative position of the two loci. The closer they are on a chromosome, the closer the recombination frequency is to zero. At the other extreme, for loci on different chromosomes, the recombination frequency is $1/2$, which is equivalent to independent assortment of the chromosomes. It may be shown under some very general assumptions that θ can never be greater than $1/2$.

Consider two loci that are polymorphic with respect to the two parental pure inbred strains. Denote by a the allele of one strain at one locus, and by b the allele of the same strain at the other locus. Similarly, for the other strain, denote by A and B the alleles at the two loci. Within each inbred strain the loci are not polymorphic. In short, we can say that the first strain is a/b homozygote and the second strain is A/B homozygote. Regardless of what crossovers may occur during the formation of the outcross, the gamete that passes on to the F_1 mouse from the first strain is always a/b, and the gamete that passes on from the second strain is always A/B. Consequently, the F_1 mouse must be $(A/B, a/b)$ heterozygote.

The gamete that is passed on from an F_1 parent to its offspring can be any one of the four possible types. If a recombination does not occur during meiosis, then the gamete may be either A/B or a/b with equal probability. If a recombination does occur, then the gamete may be A/b or a/B, again with equal probabilities. We call gametes of the second kind *recombinant* gametes. The probability of recombinant gametes in a given cross is an essential parameter for the assessment of the cross as a resource for QTL mapping. The aim in this section is to evaluate this parameter for the backcross, intercross, and a recombinant inbred strain.

The evaluation of the fraction of recombinants for the backcross and for the intercross is straightforward. For the backcross one of the gametes is inherited from an inbred parent and cannot be recombinant. The other is inherited from the F_1 and is recombinant with probability θ. Consequently, the probability that a random gamete taken from a backcross population is recombinant equals $\theta/2$. For the intercross both gametes come from an

F_1 parent. The probability of random gamete to be recombinant is exactly θ. The situation for the recombinant inbred is more complex. Before attempting an exact mathematical analysis let us examine the fraction of recombinants by simulation.

The functions "`meiosis`" and "`cross`" can be used in order to simulate a recombinant inbred strain. However, in the original programs we considered only a single locus. Modifications are needed in order to track two loci instead of one. Below is an edited version of the original function. In the original version of the function "`meiosis`" the objects "`GF`" and "`GM`" were vectors. In the new version these two objects are matrices with a column dimension of two. The first column represents the QTL, and the second represents another locus, which may be linked to the QTL. The recombination fraction between the two loci is added as a new argument to the function:

```
> meiosis.rec <- function(GF,GM,rec.frac)
+ {
+     N <- nrow(GF)
+     GS <- GF
+     from.GM <- rbinom(N,1,0.5)
+     GS[from.GM==1,1] <- GM[from.GM==1,1]
+     rec <- rbinom(N,1,rec.frac)
+     from.GM <- from.GM*(1-rec) + (1-from.GM)*rec
+     GS[from.GM==1,2] <- GM[from.GM==1,2]
+     return(GS)
+ }
```

In the code for the new function "`meiosis.rec`", the function "`nrow`" returns the number of rows in a matrix. The segregation of the first locus uses exactly the same algorithm as before. The segregation of a maternal allele at the second locus depends on the segregation at the first locus and on the recombination process. A maternal segregation occurs at the second locus if such segregation occurs in the first locus and if there is no recombination, or if there is a paternal segregation at the first locus and a recombination does occur.

The function `cross.rec` is practically identical to the function "`cross`". The only modification, which is really needed only for safety reasons, is that the recombination fraction is explicitly passed to the function "`meiosis.rec`".

```
> cross.rec <- function(fa, mo,rec.frac)
+ {
```

```
+     pat <- meiosis.rec(fa$pat,fa$mat,rec.frac)
+     mat <- meiosis.rec(mo$pat,mo$mat,rec.frac)
+     return(list(pat=pat,mat=mat))
+ }
```

For convenience we write a new function "`rec.count`" that returns the frequency of recombinant gametes in a cross. This function exploits again the function "`table`", which returns in this case the cross table of the frequencies of the different combinations of the levels of "`loc1`" and "`loc2`". In our case, the output of the "`table`" function can be treated as a 2×2 matrix.

```
> rec.count <- function(Fn)
+ {
+     loc1 <- c(Fn$pat[,1],Fn$mat[,1])
+     loc2 <- c(Fn$pat[,2],Fn$mat[,2])
+     cross.tab <- table(loc1,loc2)
+     theta <- (cross.tab[2,1]+cross.tab[1,2])/sum(cross.tab)
+     return(theta)
+ }
```

Finally, we run the simulation and make the plot:

```
> N <- 10^5
> a <- rep("a",N)
> b <- rep("b",N)
> A <- rep("A",N)
> B <- rep("B",N)
> IB1 <- list(pat=cbind(a,b),mat=cbind(a,b))
> IB2 <- list(pat=cbind(A,B),mat=cbind(A,B))
> F1 <- cross.rec(IB1,IB2,0)
> theta.BC <- theta.F2 <- theta.RI <- NULL
> rec.frac <- seq(0,0.5,by=0.05)
> for (theta in rec.frac)
+ {
+     BC2 <- cross.rec(IB2,F1,theta)
+     theta.BC <- c(theta.BC,rec.count(BC2))
+     F2 <- cross.rec(F1,F1,theta)
+     theta.F2 <- c(theta.F2,rec.count(F2))
+     Fn.fa <- cross.rec(F1,F1,theta)
+     Fn.mo <- cross.rec(F1,F1,theta)
```

```
+     for (g in 2:20)
+     {
+         New.fa <- cross.rec(Fn.fa,Fn.mo,theta)
+         New.mo <- cross.rec(Fn.fa,Fn.mo,theta)
+         Fn.fa <- New.fa; Fn.mo <- New.mo
+     }
+     RI <- list(pat=rbind(New.fa$pat,New.mo$pat),
+              mat=rbind(New.fa$mat,New.mo$mat))
+     theta.RI <- c(theta.RI, rec.count(RI))
+ }
> plot(rec.frac,theta.RI,type="l",xlab="theta",
+     ylab="rec. fraction")
> lines(rec.frac,theta.F2,col=2)
> lines(rec.frac,theta.BC,col=3)
> legend(0,0.5,legend=c("RI","F2","BC"),lty=rep(1,3),col=1:3)
```

Examine the three lines in the figure. As expected, the backcross and intercross yield straight lines with slopes 1/2 and 1, respectively. The fraction of recombinant gametes in the recombinant inbred strain is larger and is not linear. The derivative of the line for recombinant inbred strain at zero is about equal to four. The greater rate of recombination for recombinant inbreds arises from the much larger number of meioses involved.

5.3 A Mathematical Derivation of the Recombination Fraction for Recombinant Inbreds

The probability in question can be derived by consideration of the different genotypes in the population and examination of a related Markov chain. Here we outline an alternative approach, which is developed in Kimura. The beauty of this approach is that it allows substantial simplification in the analysis by exploiting the many symmetries that are present.

Consider, in particular, three probabilities, computed for the pair of male and female mice at the gth generation of inbreeding:

$C_g =$ The probability that a randomly selected gamete is of the type A/b.

$S_g =$ The probability that a randomly selected gamete in a randomly selected mouse carries the allele A and, at the same time, the other gamete of that mouse carries the allele b (at the other locus).

$T_g =$ The probability that a random gametes in a randomly selected mouse carries the allele A and a random gamete in another mouse carries the

allele b (again, at the other locus).

Of course, we are interested in C_g, which is equal to one-half the probability of a recombinant gamete. However, keeping track of the other two probabilities will allow us to write down the recursive relations:

$$\begin{aligned} C_g &= (1-\theta)C_{g-1} + \theta S_{g-1} && (6)\\ S_g &= T_{g-1} && (7)\\ T_g &= 0.5\,T_{g-1} + 0.5 \times (0.5\,C_{g-1} + 0.5\,S_{g-1})\ . && (8) \end{aligned}$$

The first relation follows from the fact that the random gamete is inherited from an animal in the previous generation. On the one hand, if no recombination takes place, then the segregated gamete is A/b if and only if it appears in the parent (and is the one selected). On the other hand, if a recombination does take place, then the gamete A/b occurs if the event described in the definition of S_g holds for the parent. The second relation follows simply by the fact that the two gametes of the offspring are a random sample of the gametes of the parents. Recombination is not relevant in this computation, since we are considering only the marginal frequencies of each of the two different loci in the parents. Consider, last, relation (8). The first locus for one offspring and the second locus for the other offspring are either inherited from different parents or from the same parent with equal probabilities. In the former case we get that the probability of the event in question is identical to the same probability for the parents. In the later case, however, there are two possibilities. Either both originate from the same gamete or from the two homologous gametes. Again, each possibility has probability of 1/2. The conditional probability of the event when they originate from the same gamete is C_{g-1} and the conditional probability when they emerge from homologous gametes is S_{g-1}. Note that recombination is again irrelevant, since only marginal probabilities are of concern.

The solution of the recursion (6–8) requires some extra work. Let $u_g = (C_g, S_g, T_g)'$ and observe the relation

$$u_g = Q u_{g-1}, \tag{9}$$

where

$$Q = \begin{pmatrix} 1-\theta & \theta & 0 \\ 0 & 0 & 1 \\ 0.25 & 0.25 & 0.5 \end{pmatrix}.$$

The matrix Q is a transition probability matrix, since its rows sum to 1. Hence it can be regarded as defining a Markov chain. From the theory of

Markov chains it is ensured to have a stationary distribution $\pi = (\pi_1, \pi_2, \pi_3)'$ with the property that $\pi' = \pi' Q$. Solving this system of equations together with the condition that $\sum_i \pi_i = 1$ produces:

$$\pi' = (1, 2\theta, 4\theta)/(1 + 6\theta) \ .$$

By (9) and stationarity we get $\pi' u_g = \pi' Q u_{g-1} = \pi' u_{g-1} = \cdots = \pi' u_0$, for all g. All entries to the vector u_g converge to the same quantity, C_∞, since the population is becoming more and more inbred. Since $C_0 = 0$, $S_0 = 0.5$, and $T_0 = 0.25$ we have $u_0' = (0, 1/2, 1/4)$, and hence

$$C_\infty = \lim_{g \to \infty} \pi' u_g = \pi' u_1 = 2\theta/(1 + 6\theta) \ .$$

Since a recombinant is either A/b or a/B, and these are equally likely, we can obtain the exact formula for the fraction of recombination in a recombinant inbred strain: $4\theta/(1 + 6\theta)$. Observe that the derivative of this fraction at $\theta = 0$ is indeed equal to four. Our work will be completed when we add the exact curve for a recombinant inbred strain to:

```
> lines(rec.frac,4*rec.frac/(1+6*rec.frac),lty=2)
```

Observe, again with reassurance, the excellent agreement between the simulated and the theoretically computed curves.

Homework Question 5.1. *Repeat the analysis of the recombination fraction that was conducted in class for the case of brother-sister mating to the case of selfing. Start by modifying the R code to allow dealing with self-fertilization and plot the recombination fraction in the final generation as a function of the fraction in a single meiosis. Try to compute the theoretical values of the recombination fraction by adapting the analysis that was carried out in class to this situation. Compare the theoretical and simulated curves to each other.*