Chapter 1
Mating Designs for Genetic analysis

Quantitative genetic theory largely deals with allelic effects and allele frequencies. The accuracy of parameter estimation depends on the allelic frequencies. In wild populations, we cannot control the allelic frequencies of the population. Therefore, estimation of genetic effects cannot be guaranteed with the optimal accuracy. In highly designed experiments, we can control the allelic frequencies and thus can design an optimal experiment to ensure that the genetic parameters are estimated with high accuracy. There are many types of mating design for genetic analysis, including backcross (BC), F_2, recombinant inbred line (RIL), double haploid (DH), four-way (FW) cross mating design, diallel cross mating design (DIAL), North Carolina II (NCII) mating design, nested association mapping (NAM) population, multiparent advanced generation intercross (MAGIC) population, full-sib family design (FS), half-sib family design (HS), hierarchical (nested) mating design (NESTED) and random mating design (RANDOM). Data collected from a random mating design experiment are called pedigree data (PEDIGREE), which are common in human genetics, genetic analysis of large animals and tree breeding.

BC (backcross) mating design

Figure 1 shows a sketch of the BC mating design. This design starts from two inbred parents and the F_1 (first filial) hybrid.

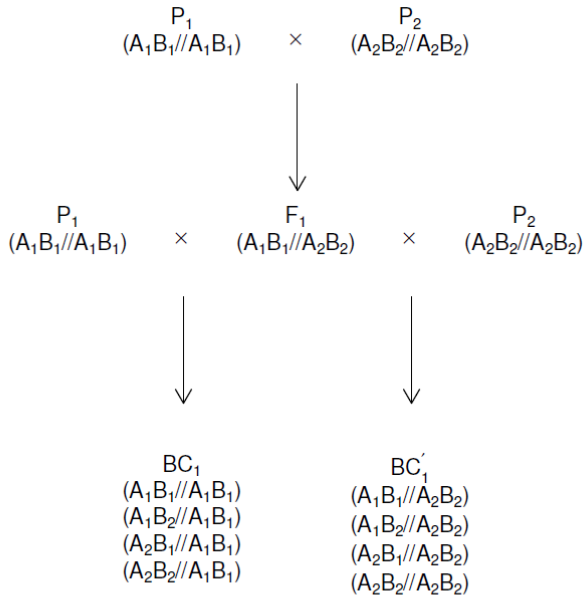

Figure 1. Sketch of the backcross (BC) mating design, where the cross of the F_1 hybrid back to the first parent is called BC_1 and the cross of the F_1 hybrid back to the second parent is called $BC_1^{'}$.

A BC family is generated through crossing the F_1 individual back to one of the two parents. If P_1 is the backcrossed parent, the BC family is called BC_1. If P_2 is the backcrossed parent, the BC is called $BC_1^{'}$. Take BC_1 for example, there are two possible genotypes, A_1A_1 and A_1A_2. It is impossible to estimate the dominance effect because there is not enough degree of freedom to do so. The QTL effect is defined as

$$a' = G_{11} - G_{12} = a - d \qquad \text{(xx)}$$

Apparently, the QTL effect defined this way is the difference between the additive effect and the dominance effect. This effect is equivalent to the additive effect only if the dominance effect is absent. Using a BC family for QTL mapping is not as powerful as using the F_2 family because: (1) for the same sample size, a BC family only carries half the number of meioses as the F_2 family; (2) the additive and dominance effects are confounded. When the dominance effect is absent, we need to double the sample size for a BC design to achieve the same statistical power as the F_2 design. Under the assumption of no dominance, the model appears

$$y_j = \mu + X_j a + e_j \qquad \text{(xxx)}$$

where $X_j = \{1 \quad 0\}$ for the two genotypes, A_1A_1 and A_1A_2. The fact that BC design is not as powerful as the F_2 design can be shown by looking at the variances of X_j in the two different families. The scale of X_j in the F_2 design must be $X_j = \{1 \quad 0 \quad -1\}$ for the three genotypes in order to compare the powers for the two different designs. For the BC design $\text{var}(X_j) = 1/4$, but $\text{var}(X_j) = 1/2$ for the F_2 design. The design with a larger $\text{var}(X_j)$ has more power than the design with a smaller $\text{var}(X_j)$, which explains why the F_2 design is more powerful than the BC design.

F_2 (second filial) mating design

The most popular design of experiments in QTL mapping is the F_2 mating design. An F_2 design is through a line crossing experiment involving two inbred lines (see Figure 2), called P_1 and P_2. The F_1 hybrid from the cross of P_1 and P_1 is then selfing to generate a segregating F_2 family. QTL mapping can be performed using the F_2 family. In terms of allelic frequencies of segregating loci, they are optimal because each parent contributes an equal number of alleles. Let A_1A_1 and A_2A_2 be the genotypes of the two parents, respectively, and A_1A_2 be the genotype of the hybrid. There are three possible genotypes in the progeny of the F_2 family: A_1A_1, A_1A_2 and A_2A_2. The ratio of the three genotypes is $1:2:1$. Let G_{11}, G_{12} and G_{22} be the genotypic values for the three genotypes, respectively.

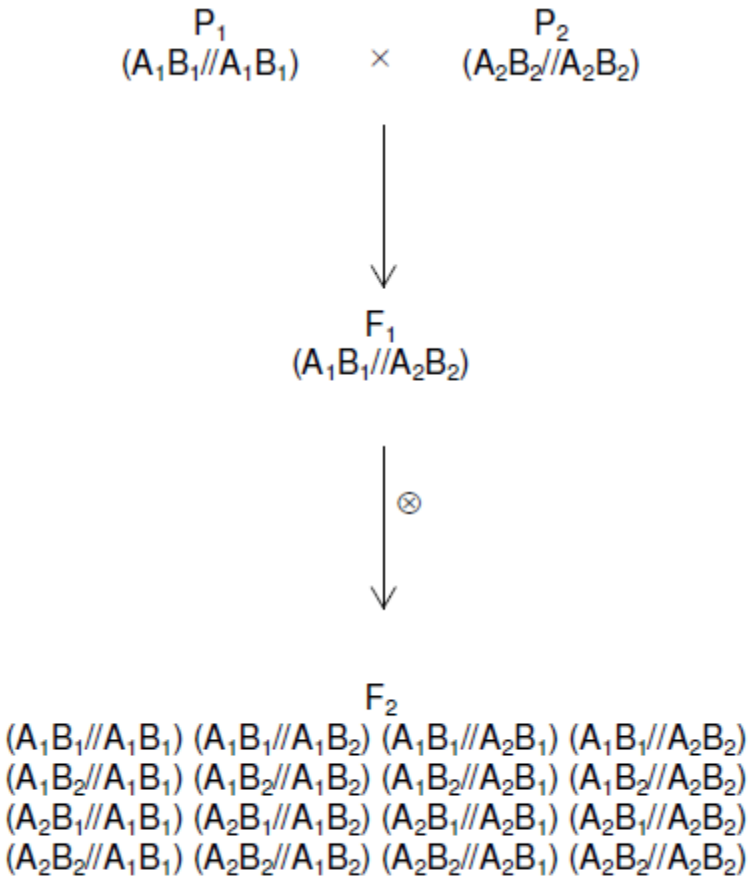

Figure 2. Sketch of the F_2 mating design, where the F_1 hybrid keeps selfing to generate multiple F_2 individuals to form an F_2 family.

The additive and dominance effects of the locus are defined as

$$a = G_{11} - \frac{1}{2}(G_{11} + G_{22}) = \frac{1}{2}(G_{11} - G_{22}) \quad \text{(xx)}$$

and

$$d = G_{12} - \frac{1}{2}(G_{11} + G_{22}) \quad \text{(xx)}$$

respectively. Rather than estimating the genotypic values, we actually estimate and test *a* and *d* in QTL mapping. The linear model for a single QTL is given

$$y_j = \mu + X_j a + W_j d + e_j \quad \text{(xxx)}$$

where y_j is the phenotypic value of individual *j*, μ is the population mean (or intercept), X_j is an indicator variable (for the additive effect) assigned a value of 1, 0 or -1, respectively, for the three genotypes, A_1A_1, A_1A_2 and A_2A_2, W_j is an indicator variable (for the dominance effect) assigned a value of 0, 1 or 0, respectively, for the three genotypes, A_1A_1, A_1A_2 and A_2A_2, e_j is the residual error following a $N(0,\sigma^2)$ distribution. The genotype indicator variables, X_j

and W_j, can be defined in many different scales. The scales are usually chosen for statistical convenience rather than for biological meaningfulness because the scales only affect the estimation of the genetic effects and do not affect the results of statistical tests. If there is no segregation distortion, the following scales for X_j and W_j are recommended (Yang, Tian and Xu 2006), $X_j = \{\sqrt{2} \quad 0 \quad -\sqrt{2}\}$ and $W_j = \{-1 \quad 1 \quad -1\}$ for the three genotypes, A_1A_1, A_1A_2 and A_2A_2. This scale choice leads to $\text{var}(X_j) = \text{var}(W_j) = 1$ and $\text{cov}(X_j, W_j) = 0$, and thus

$$\text{var}(y_j) = a^2 + d^2 + \sigma^2 \qquad \text{(xxx)}$$

which is mathematically more attractive than any other scales.

RIL (recombinant inbred line) mating design

Recombinant inbred line design also involves two inbred parents, the F_1 hybrid and the F_2 family. Each F_2 progeny is undergoing many generations of continuous selfing until all loci are fixed. This may take about 7-8 generations to reach 98% homozygosity. Eventually, each line descended from the cross is an inbred line, but carries genes from different parents across loci. In other words, within each locus, an RIL line carries the same allele from one parent, but between loci, the contributing parents may alternate. Therefore, an RIL carries a mosaic genome of the two parents.

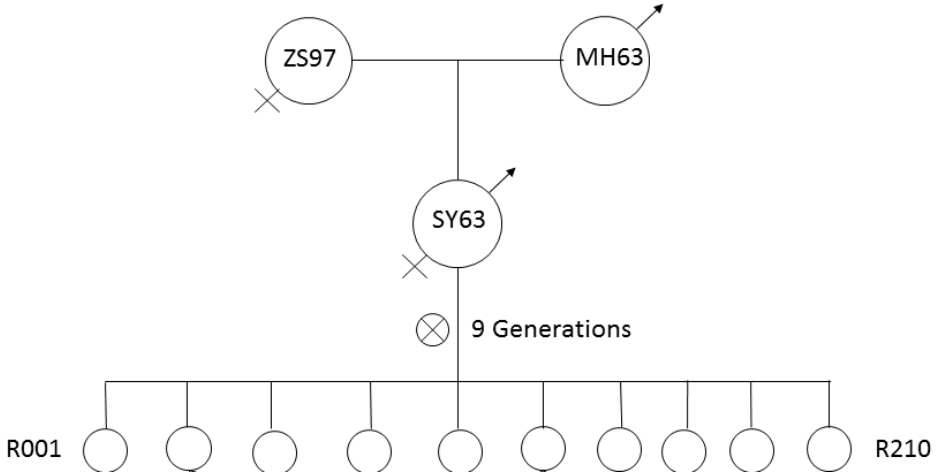

Figure 3. Recombinant inbred lines (RIL) derived from self-fertilization of the hybrid of two inbred parents.

The two homozygotes are A_1A_1 and A_2A_2 with the QTL effect at this locus defined as

$$a' = G_{11} - G_{22} = a - (-a) = 2a \qquad \text{(xx)}$$

Therefore, using the RIL mating design is more powerful than a BC or F_2 design because the QTL effect defined is doubled. If we define $X_j = \{1 \quad -1\}$ for the two homozygotes, A_1A_1 and A_2A_2, the genetic variance at this locus is

$$\text{var}(X_j)(a')^2 = \text{var}(X_j)(2a)^2 = 4a^2 \qquad \text{(xx)}$$

which is much larger than the corresponding genetic variance for the BC design ($a^2/4$) and the genetic variance for the F_2 design ($a^2/2$) under the same scale of variable X_j. In addition, for a sample size n of an F_2 family, there are $2n$ meioses, but the same number of RIL individuals will cumulate many more meioses. The genetic material has altered many times across the genome (loci), leading to a "high" frequency of recombination between loci. Therefore, using the RIL design has both advantages of high power and high resolution (fine mapping) over using the F_2 design. Let r be the recombination frequency between two loci per meiosis (in a BC for example), after many generations of cumulated meioses, the recombination frequency in the RIL will become

$$c_1 = \frac{2r}{1+2r} \qquad \text{(xxx)}$$

This multiple meiosis corrected recombination fraction is larger than the original recombination fraction. Therefore, we will expect to see many more crossovers between two loci in the RIL design than those in a BC family. The genome essentially gets "longer" and thus allows us to do fine mapping. Recombinant inbred lines generated through selfing are called RIL_1. In animals where self-fertilization does not happen, recombinant inbred lines can be generated through continuous brother-sister mating. The sib-mating approach will take more generations (about 20 generations) to reach the same homozygosity as the selfing approach. This type of RIL is called RIL_2. The corresponding correction for the recombination frequency is

$$c_2 = \frac{4r}{1+6r} \qquad \text{(xxx)}$$

Statistical methods of QTL mapping for RIL and BC are identical once r in BC is replaced by c in RIL.

IMF2 (immortalized F2) design

This type of mating design is called IMF2 mating design, which was first designed by Qifa Zhang's group (Hua et al. 2003) in rice. An IMF2 population is generated by random crosses of RILs derived from two parents. The hybrid population mimics an F_2 family (with three possible genotypes of 1:2:1 ratio) but the hybrid genotypes can be repeated indefinitely. For example, if one identifies a super hybrid, he can duplicate this hybrid by selecting the same two RILs to cross again. The genotype of the new hybrid will be exactly the same as the original hybrid because their parents are the same.

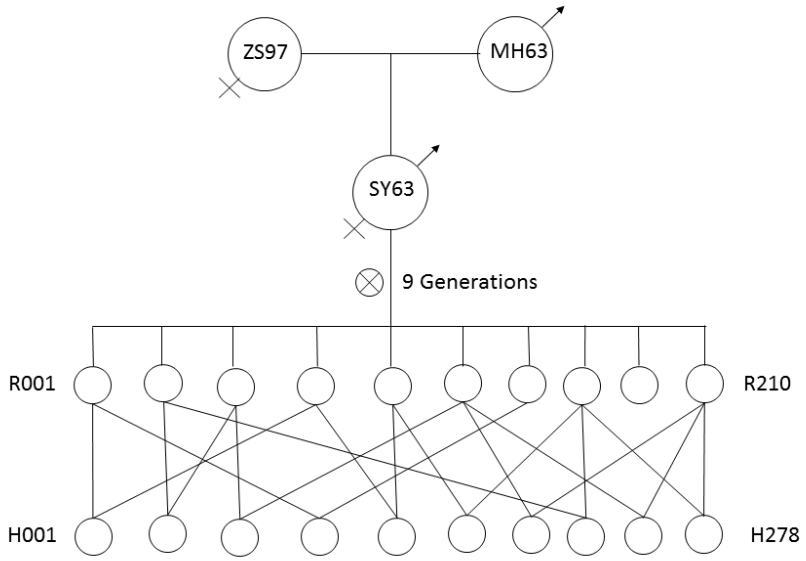

Figure 4. Immortalized F_2 mating design. Hybrids derived from the crosses of RILs (of two parents) are called IMF2 individuals. This hybrid family (H001 – H278) mimics an F_2 family (with three possible genotypes of 1:2:1 ratio) but the hybrid genotypes can be repeated indefinitely.

DH (double haploid or doubled haploid) mating design

A double haploid individual is created by duplicating a gamete (haploid) via chemical treatment. A DH individual is a diploid homozygote for all loci. Starting from the F_1 progeny derived from the cross of two inbred lines, each of the two gametes (haploids) generated by the F_1 is duplicated and the two copies are fused to give two DH individuals (Figure 5). Like the RIL design, there are two possible genotypes in a DH population, A_1A_1 and A_2A_2 with the QTL effect defined as

$$a' = G_{11} - G_{22} = a - (-a) = 2a \qquad \text{(xx)}$$

Therefore, DH design should provide the same power as an RIL design. However, the resolution of DH mapping is equivalent to an F_2 design because the numbers of meioses of the two mating designs are the same.

The mating designs discussed so far all involve only two parents. Therefore, these mating design are called bi-parental mating designs. All biparental mating designs have the same drawback, that is, low genetic diversity. Since only two alleles are involved, if the two parents are not segregating at the locus of interest, we will not be able to detect this locus. Therefore, we will introduce multiparent mating designs in this chapter. Such multiparent populations are called the next generation mapping populations.

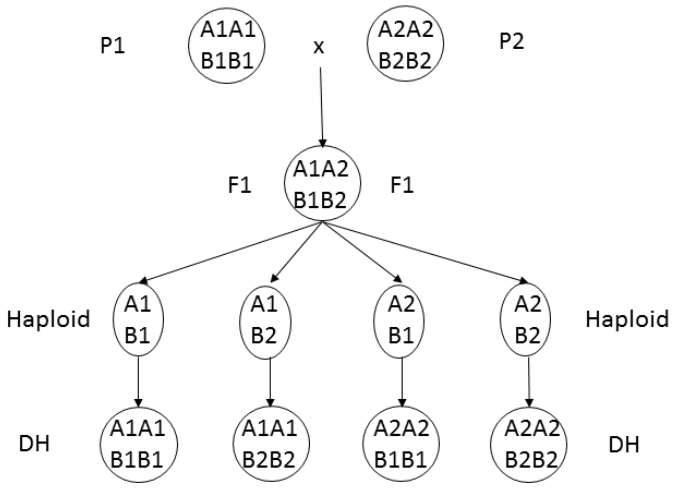

Figure 5. Sketch of a double haploid (DH) mating design.

FW (four-way) cross mating design

The four-way cross mating design requires four inbred lines and two rounds of crossing. In the first round of crossing, $P_1 \times P_2 \to F_1^{(12)}$ and $P_3 \times P_4 \to F_1^{(34)}$, two independent F_1 hybrids are generated. In the second round of crossing, $F_1^{(12)} \times F_1^{(34)} \to FW$, the two different F_1 hybrids are crossing to generate a four-way cross family (Figure 6). The mating design is more clearly described using the genotypic labels. In the first round of cross, we have $A_1A_1 \times A_2A_2 \to A_1A_2$ and $A_3A_3 \times A_4A_4 \to A_3A_4$. In the second round of cross, we get

$$A_1A_2 \times A_3A_4 \to \{A_1A_4, A_1A_4, A_2A_3, A_2A_4\} \qquad \text{(xx)}$$

There are four possible genotypes in the four-way cross family. The labels of the alleles need to be changed again in order to describe the genetic model for the FW cross design.

In the first round of cross, we have $A_1^pA_1^p \times A_2^pA_2^p \to A_1^pA_2^p$ and $A_1^mA_1^m \times A_2^mA_2^m \to A_1^mA_2^m$. Note that the four alleles involved in the FW progeny, A_1, A_2, A_3 and A_4, have been relabeled as A_1^p, A_2^p, A_1^m and A_2^m, respectively, where the superscripts *p* and *m* indicate the paternal and maternal origins of the progeny and the subscripts 1 and 2 indicate the paternal and maternal origins of the parents. With this new notation, we get the FW cross family

$$A_1^pA_2^p \times A_1^mA_2^m \to \{A_1^pA_1^m, A_1^pA_2^m, A_2^pA_1^m, A_2^pA_2^m\} \qquad \text{(xx)}$$

We now assign an allelic value to each allele, say a_1^p, a_2^p, a_1^m and a_2^m, for the four alleles. The corresponding genotypic values are now defined as

$$\begin{bmatrix} G_{11} \\ G_{12} \\ G_{21} \\ G_{22} \end{bmatrix} = \begin{bmatrix} \mu + a_1^p + a_1^m + d_{11} \\ \mu + a_1^p + a_2^m + d_{12} \\ \mu + a_2^p + a_1^m + d_{21} \\ \mu + a_2^p + a_2^m + d_{22} \end{bmatrix} \quad \text{(xxx)}$$

where d_{ij} is the interaction effect between the two alleles involved in the genotype.

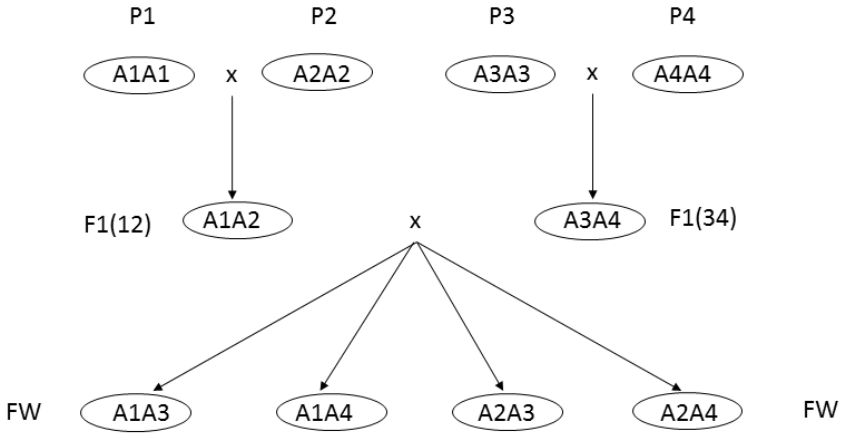

Figure 6. Sketch of a four-way (FW) cross mating design.

The model is over parameterized because we cannot estimate nine parameters using four genotypes. Therefore, some restrictions are required to reduce the number of parameters. Many different schemes of restrictions can be used, but one particular scheme lead to the following reduced parameters (Xu 1998),

$$\begin{bmatrix} \alpha^p \\ \alpha^m \\ \delta \end{bmatrix} = \begin{bmatrix} \frac{1}{2}(a_1^p - a_2^p) \\ \frac{1}{2}(a_1^m - a_2^m) \\ \frac{1}{4}(d_{11} - d_{12} - d_{21} + d_{22}) \end{bmatrix} \quad \text{(xxx)}$$

Including μ, we have four estimable parameters that are expressed as linear contrasts (combinations) of the genotypic values,

$$\begin{bmatrix} \mu \\ \alpha^p \\ \alpha^m \\ \delta \end{bmatrix} = \begin{bmatrix} \frac{1}{4}G_{11} + \frac{1}{4}G_{12} + \frac{1}{4}G_{21} + \frac{1}{4}G_{22} \\ \frac{1}{4}G_{11} + \frac{1}{4}G_{12} - \frac{1}{4}G_{21} - \frac{1}{4}G_{22} \\ \frac{1}{4}G_{11} - \frac{1}{4}G_{12} + \frac{1}{4}G_{21} - \frac{1}{4}G_{22} \\ \frac{1}{4}G_{11} - \frac{1}{4}G_{12} - \frac{1}{4}G_{21} + \frac{1}{4}G_{22} \end{bmatrix} = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \\ \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} \\ \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} \end{bmatrix} \begin{bmatrix} G_{11} \\ G_{12} \\ G_{21} \\ G_{22} \end{bmatrix} \quad \text{(xxx)}$$

The reverse relationship is

$$\begin{bmatrix} G_{11} \\ G_{12} \\ G_{21} \\ G_{22} \end{bmatrix} = \begin{bmatrix} \mu + a^p + a^m + \delta \\ \mu + a^p - a^m - \delta \\ \mu - a^p + a^m - \delta \\ \mu - a^p - a^m + \delta \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha^p \\ \alpha^m \\ \delta \end{bmatrix} \quad \text{(xxx)}$$

Let us define

$$H = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \quad \text{(xx)}$$

We can see that

$$H^{-1} = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \\ \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} \\ \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} \end{bmatrix} \quad \text{(xx)}$$

We now give the model expressed as functions of only the estimable parameters,

$$y_j = X_j \beta + e_j \quad \text{(xxx)}$$

where

$$X_j = \begin{cases} H_1 & \text{for } A_1^p A_1^m \\ H_2 & \text{for } A_1^p A_2^m \\ H_3 & \text{for } A_2^p A_1^m \\ H_4 & \text{for } A_2^p A_2^m \end{cases} \text{ and } \beta = \begin{bmatrix} \mu \\ \alpha^p \\ \alpha^m \\ \delta \end{bmatrix} \quad \text{(xxx)}$$

and H_k is the *k*th row of matrix *H*. Under this scale of definition for X_j, the total phenotypic variance can be partitioned into four components,

$$\text{var}(y_j) = (\alpha^p)^2 + (\alpha^m)^2 + \delta^2 + \sigma^2 \quad \text{(xxx)}$$

where σ^2 is the residual error variance.

FS (full-sib) family

A full-sib family is a family of individuals generated by repeated matings between two parents. Individuals within the family are called full-siblings and they all share the same father and the same mother. The genetic model described in the FW cross design directly applies to QTL mapping in the full-sib family. The father (paternal parent) of the full-sib family is equivalent to the F_1 hybrid used as the paternal parent of a FW cross. The mother (maternal parent) of the full-sib family is equivalent to the F_1 hybrid used as the maternal parent of a FW cross. We can only estimate the allelic difference between the two alleles of the father (α^p), the allelic difference between the two alleles of the mother (α^m) and the interaction effect (δ). The full-sib family design differs from a FW cross is that we need to infer the linkage phases of the markers prior to conducting QTL mapping because we do not necessarily have the genotypic information of the grandparents in the full-sib family. Once the linkage phases are inferred for the full-sib family, the statistical model and method in the FW cross apply to the full-sib family. Figure 7 shows a

sketch of the full-sib design. An even more informative sketch is shown in Figure 8, where it uses the standard symbols for male and female individuals in pedigrees.

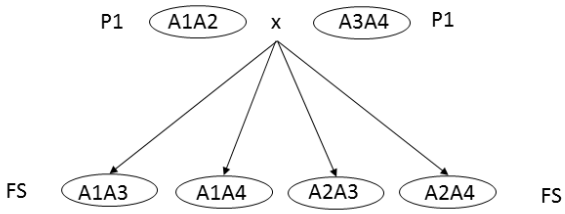

Figure 7. Full-sib (FS) design

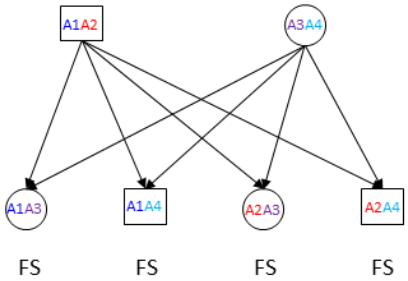

Figure 8. Full-sib (FS) design (an alternative presentation)

HS (half-sib) family

Each member of a half-sib family has a different mother but all share the same father. This type of family is common in large animals such as beef cattle. Half-sib families can also be found in forest trees but the common parent is the female parent. Statistically, we can treat a half-sib family as a BC family. The common parent in the half-sib family is treated as the "F_1" hybrid in the BC. The other parents (all independent) in the half-sib family are treated as the backcrossed "parent." This comparison is hard to understand, but it is true from the statistical model point of view. In a BC family of this mating type $A_1A_2 \times A_1A_1$, we estimate the difference between the two genotypes of the progeny, A_1A_1 and A_1A_2. In fact, we are estimating the difference between the two alleles carried by the F_1 hybrid (A_1A_2). The common parent (A_1A_1) plays no roles other than providing a background for evaluation of the two alleles of the F_1 hybrid. In half-sib QTL mapping, we are estimating the difference between the two alleles of the common parent. The background alleles are provided by all other independent parents. The difference between the two different designs occurs in the different background alleles. A BC design has a uniform or homogeneous background allele while a half-sib family has a heterogeneous background allelic

array. The background alleles play no role in the statistical model. You need to manipulate the data a little bit to "fool" the program. First, you need to infer the linkage phases of all markers for the common parent and label the paternal allele of the common parent by A_1 and the maternal allele by A_2. Secondly, you need to recode the genotypes of the progeny by $A = A_1\_$ and $B = A_2\_$, where the underscore mean a wild card representing the background alleles. We have now relabeled the genotypes of the progeny so that there are only two possible "genotypes" in the progeny. This half-sib family can now be studied using the BC mating design.

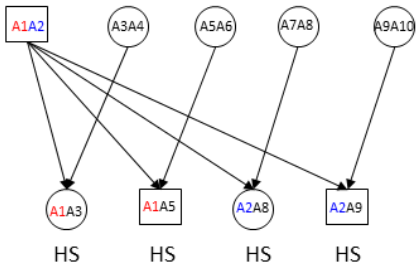

Figure 9. Half-sib (HS) design (equivalent to a BC design but with a heterogeneous background)

DIAL (diallel cross)

A set of diverse (inbred) lines are selected for cross to generate their F2 hybrids. There are two diallel cross designs, full diallel (Figure 10) and half diallel (Figure 11). In the full diallel cross, there are two versions of each cross (the cross and its reciprocal). This type of diallel cross allows investigators to exam maternal or mitochondrial effects (animals) or chloroplast effects (plants). Diagonals are often ignored because they simply represent reproduction of the inbred lines.

Female parent

Male parent	A	B	C	D	E	F	G
A		x	x	x	x	x	x
B	x		x	x	x	x	x
C	x	x		x	x	x	x
D	x	x	x		x	x	x
E	x	x	x	x		x	x
F	x	x	x	x	x		x
G	x	x	x	x	x	x	

Figure 10. Full diallele cross design

Female parent

Male parent	A	B	C	D	E	F	G
A							
B	x						
C	x	x					
D	x	x	x				
E	x	x	x	x			
F	x	x	x	x	x		
G	x	x	x	x	x	x	

Figure 11. Half diallel cross design.

A half diallel cross is simply the half of the full diallel cross (Figure 11). Diallel cross design allows investigator to estimate general combining ability (GCA) and special combining ability (SCA). You cannot use a diallel cross experiment this way to perform QTL mapping or association study because the hybrids are simply F1 individuals. To do QTL mapping or association studies, you need to self these hybrids for at least one more generation, preferably for 7-8 generations to develop recombinant inbred lines within each cross (a sort of nested association mapping population).

NCII (North Carolina II design)

This type of design is often seen in rice and corn breeding where male and female parents are separated. You choose a set of female parents (3 in the example) and a set of male parents (8 in the example). The female and male parents are genetically different (Figure 12). This type of mating design is also called rectangular mating design. Data are collected from the F1 hybrids in the traditional NCII experiment.

Male parent

Female parent	D	E	F	G	H	I	J	K
A	x	x	x	x	x	x	x	x
B	x	x	x	x	x	x	x	x
C	x	x	x	x	x	x	x	x

Figure 12. North Carolina II mating design (also called rectangular mating design)

NAM (nested association mapping) design

This type of design has exactly the same format as the NCII design except that each hybrid will go self-fertilization for multiple generations (7-8 generations) to generate a RIL population. For the example of $3\times8=24$ hybrids example (Figure 13), we will have 28 RIL populations. The RIL populations are nested within the hybrid combinations, explaining why it is called NAM.

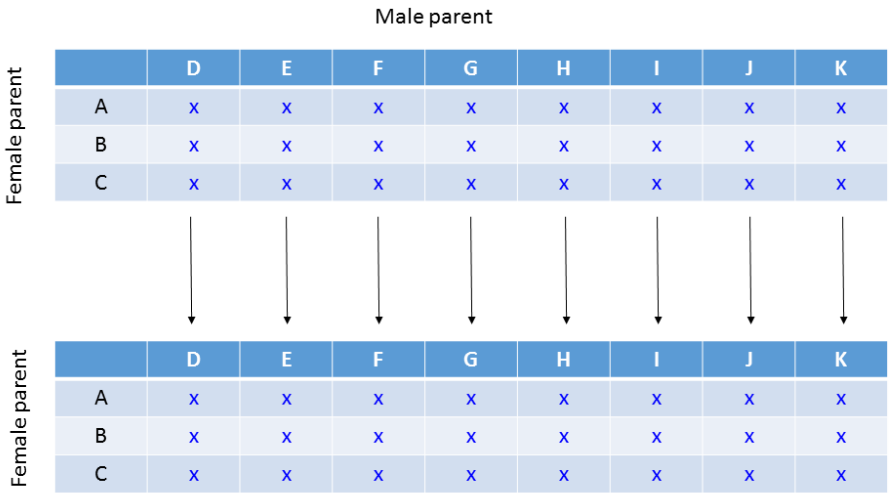

Figure 13. Sketch of a nested association mapping population

The first NAM population was created for maize by Edward Buckler's research team (Yu et al. 2008) using 25 diverse lines of maize to cross with B73 to initiate 25 hybrids in the first generation. Each of the 25 hybrids was then self-fertilized for six generations to produce 200 RILs, making a total of $25\times200=5000$ total RILs. A sketch of the maize NAM population is shown in Figure 4.14. A similar NAM population was created for soybean by Song et al. (2017) with 40 inbred lines crossing with one cultivar. Each hybrid was self-fertilized for 5 generations to generate 14 RILs, making a total of $40\times140=5600$ RILs. The maize and soybean Nam populations are special cases where only one reference line was used. The NAM sketched in Figure 13 represents a more general design where multiple reference lines are used.

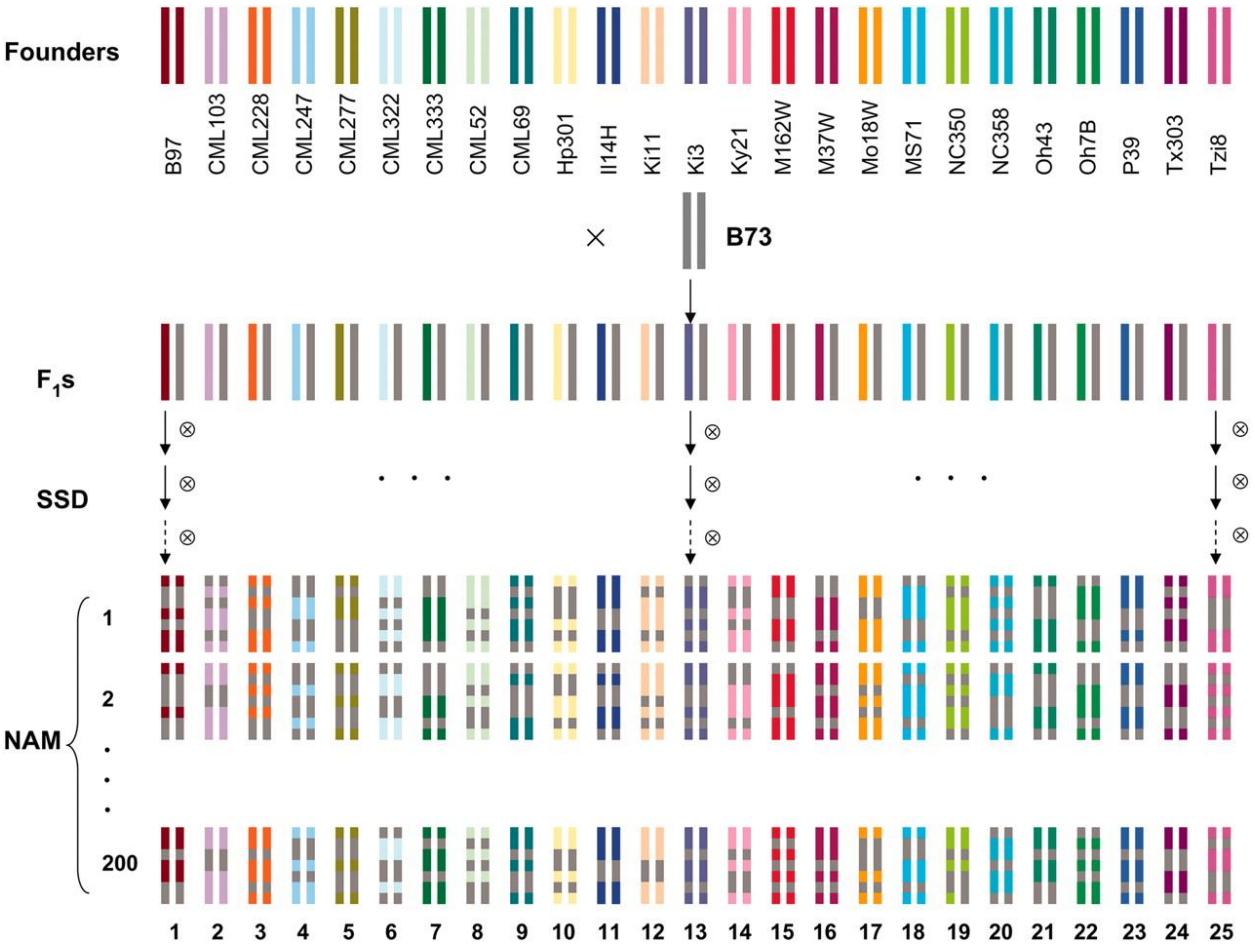

Figure 14. Sketch of the NAM population in maize

MAGIC (multiparent advanced generation intercross) population

The acronym "MAGIC" stands for multiparent advanced generation inter-cross, just trying to make up the word magic! In plants, it is called MAGIC population (Kover et al. 2009). In animals (laboratory mice), however, it is called collaborative cross (CC) (The Complex Trait Consortium 2004), because this type of experiment is often in a large scale and cannot be conducted by a single lab; rather, it requires collaborative work of many research laboratories. Figure 15 shows a MAGIC population initiated with eight parents. It starts with 8 (an even number) inbred lines to form 4 pairs of matings to generate 4 F1 hybrids. The 4 F1 hybrids form 2 pairs of matings to generate 2 (four way) crosses. The last two four-way crossed hybrids to cross again to form the final 8-way cross. The 8-way crosses are then self-fertilized or brother-sister mating for many generations to created recombinant inbred lines. These recombinant inbred lines are collectively called a MAGIC or CC populations (Figure 15). A progeny's genome is a mosaic of 8 genomes (see Figure 16)

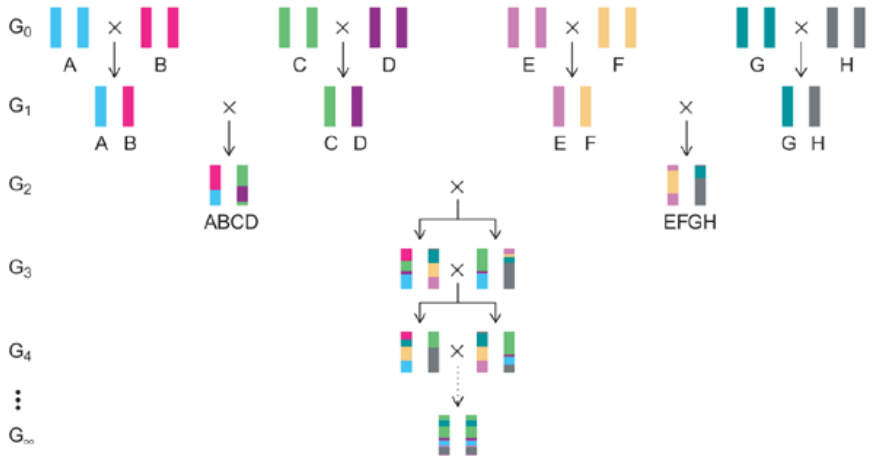

Figure 15. Sketch of a MAGIC or CC population

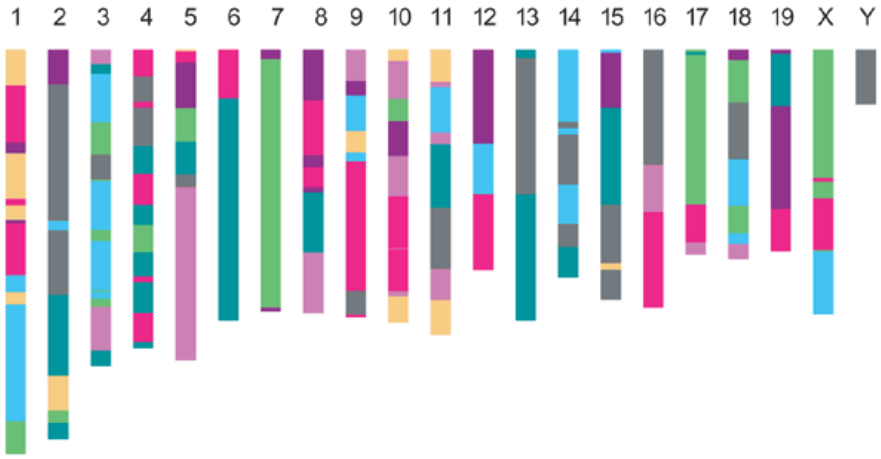

Figure 16. The mosaic genome of a MAGIC or CC individual incorporating genomes of 8 ancestors.

NAM and MAGIC populations serve as intermediate populations between linkage mapping (high LD) and association mapping (low LD). Therefore, we can use these populations to map QTL with relatively high resolution and high statistical power.

Random populations (pedigree data)

Such data are often available in human, large animals and forest trees. We collect the data as they are but we must keep the pedigree relationship of the samples (Figure 17 and Figure 18). Special algorithms are required to map QTL using such random populations.

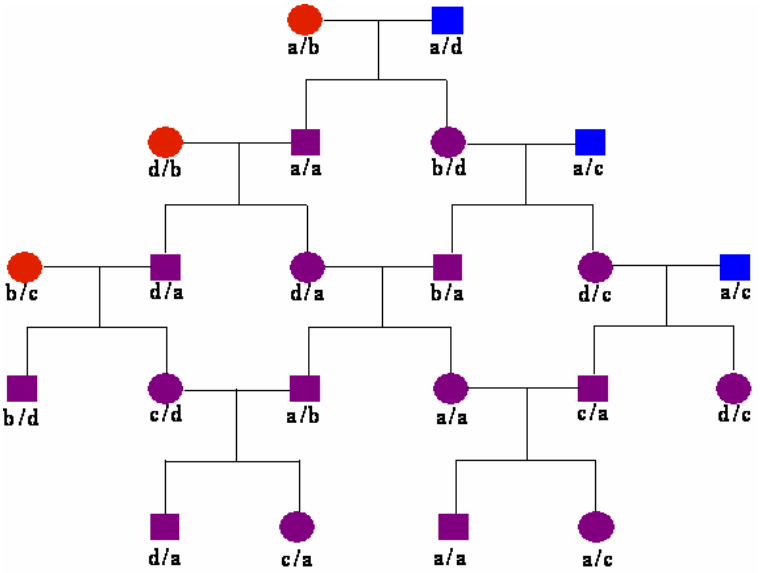

Figure 17. An arbitrary pedigree with 22 members

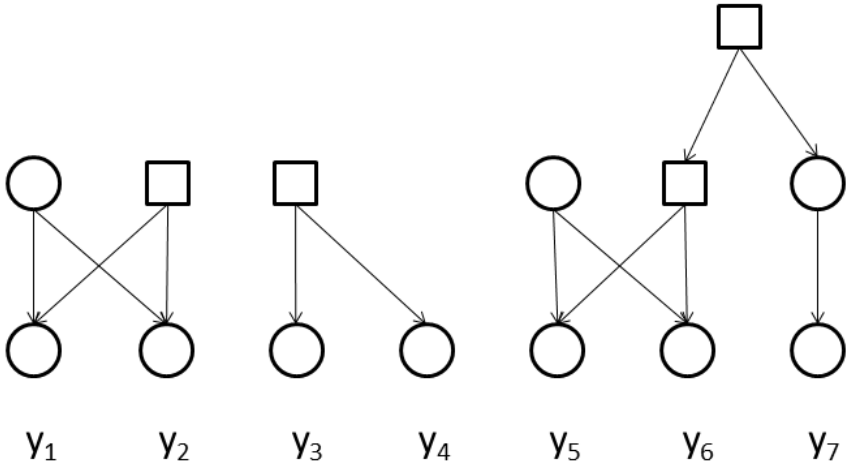

Figure 18. A population consisting of several independent families.

Random populations (unknown genetic relationship, genome-wide association studies)

When genetic relatedness of individuals in a population are unknown, we must assume that they are independent. We then use markers to infer their empirical relatedness and perform genetic analysis. Such a population is the one used in genome-wide association studies (see Figure 19).

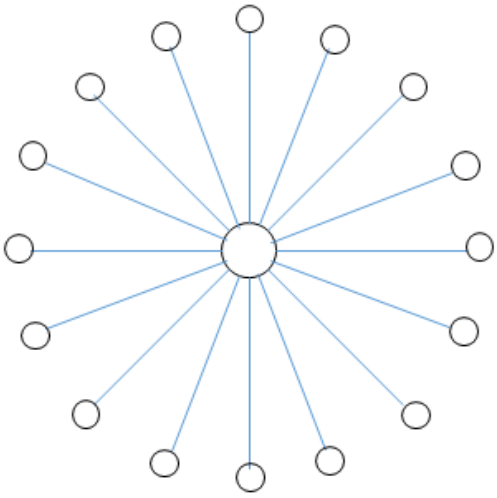

Figure 19. Sketch of a random population (unknown relationship among individuals).