PROPORTION OF INFORMATIVE FAMILIES FOR GENETIC COUNSELING WITH LINKED Marker GENES

Masatoshi Nei

Center for Demographic and Population Genetics,
University of Texas at Houston, Texas 77025 U.S.A.

Summary. The utility of linked marker genes for genetic counseling depends on the gene frequencies and dominance relationship at the marker locus as well as on the recombination value between the marker and disease gene loci. A quantity called the proportion of informative families was introduced to measure the effect of gene frequencies and dominance relationship on the utility of marker genes. This proportion is higher for X-linked genes than for autosomal genes, other things being equal. It is also higher for codominant markers than for marker genes with dominance. It is shown that when gene frequencies are appropriate linked marker genes are quite useful for genetic counseling. Examples are given by using real data on G6PD, colorblindness, the ABO blood groups, etc.

INTRODUCTION

In recent years a great effort has been made to construct a human linkage map. One of the main purposes for making such a map is to use marker genes for genetic counseling. The idea of using linked marker genes for genetic counseling is not new. Hoogvliet (1942) used colorblindness for detecting the female carrier of the hemophilia gene. Edwards (1956) and Renwick (1969) suggested the use of linked marker genes for intrauterine diagnosis of genetic diseases when the disease gene is not expressed. Similarly, marker genes are useful for predicting late-onset genetic diseases before the diseases are manifested. However, no one seems to have done a systematic study of the utility and efficiency of a marker locus for genetic counseling, though special cases were investigated by Murphy and Chase (1975) and Rivas and Connally (1977). To evaluate the utility of marker loci for genetic counseling two different criteria are required. One is the accuracy of prediction or the probability with which an individual with marker gene contracts the disease in question. This probability depends on the recombination value between the marker and disease genes and how the linkage phase of the parent is determined. The other is the proportion of individuals or families (informative families) in which a particular marker

Received May 7, 1979
locus can be used for genetic counseling. This proportion depends on the frequencies and dominance of marker alleles. Obviously, the marker locus must be polymorphic; otherwise it has no utility for genetic counseling.

The linkage phase of a counselee (parent) for X-linked or autosomal dominant diseases can be determined with probability one if information on the genotypes of grandparents is available. In this case the accuracy of predicting a genetic disease depends solely on the recombination value, and it can easily be computed. However, computation of the proportion of informative families is somewhat complicated. The purpose of this paper is to develop mathematical formulae for computing this proportion for various situations. In this paper I shall consider only the case where the linkage phase can be determined with certainty and neglect the effect of new mutations. Other cases where the linkage phase of counselee is determined through his or her sibs or children will be treated in a separate paper. In these cases a somewhat different approach is necessary. The effect of mutation will also be considered in a subsequent paper.

X-LINKED RECESSIVE DISEASES

Let us first consider how marker genes are used for predicting genetic diseases. Let \( D \) and \( d \) be the normal and disease alleles respectively at an X-linked locus, and \( M_1 \) and \( M_2 \) be the alleles at the marker locus. We denote by \( r \) the recombination value between the two loci. We consider the case of codominant alleles at the \( M \) locus, so that the three possible genotypes in females are identifiable. Since female patients for X-linked recessive genes are very rare, we ignore them. If a mother is heterozygous for \( d \), one-half of her male children are expected to contract the disease. Let the genotype of the mother be \( DM_1/dM_2 \) (Fig. 1). Then, the proportion \( 1 - r \) of her sons with allele \( M_2 \) is expected to develop the disease. Therefore, if \( r \) is small, the probability of predicting the disease before it is expressed is very high. Clearly, marker genes are useful for increasing the predictability of genetic diseases. It is noted that marker genes are also useful for predicting the carrier status of a female offspring. In the above example, if the \( DM_1/dM_2 \) mother marries a husband with genotype \( DM_1 \), \( 1 - r \) of her daughters with genotype \( M_1M_2 \) will carry the \( d \) gene.

In the above discussion we assumed that the mother is a double heterozygote. Actually, this is essential; if she is not a double heterozygote, the marker locus has no utility for increasing the predictability of genetic diseases. There are two types of double heterozygotes, \textit{i.e.} \( DM_1/dM_2 \) and \( DM_2/dM_1 \). Both of these heterozygotes are informative for genetic counseling. However, to make a prediction we must know whether allele \( d \) is linked with \( M_1 \) or \( M_2 \). Knowledge of double heterozygosity alone is not sufficient. In general, we cannot determine the linkage phase of a mother by examining her genotype.

In the case of X-linked genes the linkage phase of a mother can be determined by examining the genotype of grandfather (see Fig. 1). If his genotype is \( dM_2 \), then
the mother’s genotype must be $DM_1/dM_2$, neglecting new mutation. On the other hand, if his genotype is $dM_1$, then the mother’s genotype is $DM_2/dM_1$. With certain genetic diseases, the heterozygous condition of the $D$ locus is detected by some biochemical techniques. In this case the linkage phase of the mother can be determined even if the grandfather does not have the disease gene. If the grandfather is normal, the gene $d$ carried by the mother must have come from the grandmother, but the mother’s linkage phase can be determined by examining the grandfather. Namely, if his genotype is $DM_1$, then the mother’s genotype must be $DM_2/dM_2$. On the other hand, if his genotype is $DM_2$, then the mother’s genotype must be $DM_2/dM_1$.

The question we now face is: How often can we tell the linkage phase of a mother with gene $d$ if she visits a genetic counselor? In other words what is the proportion of informative mothers? We consider this question for two different situations. The first situation is one in which a mother comes to know her carrier status because her father had the genetic disease, whereas in the second situation her carrier status is revealed by a biochemical test or some other similar technique in genetic screening. The answer to our question depends on whether there is dominance at the marker locus or not. We shall first consider the case of no dominance and then the case of dominance.

(1) Codominant marker genes

Affected grandfather. Let $x_1$ and $x_2$ be the frequencies of alleles $M_1$ and $M_2$ at the $M$ locus, respectively, in a random mating population. We consider the probability that a mother heterozygous for the $d$ gene is also heterozygous at the $M$ locus, given that the grandfather has allele $d$. For simplicity, throughout this paper we assume that the gene frequencies at the $D$ and $M$ loci have reached the equilibrium value and there is linkage equilibrium between the two loci. The grandfather may have either allele $M_1$ or $M_2$ at the $M$ locus. The probability that he has $M_1$ is $x_1$, and the probability that he has $M_2$ is $x_2$. The probability that the mother is $DM_1/dM_2$, given that she is heterozygous for the $d$ gene and the grandfather is $dM_2$, is $x_1$. Therefore, the probability that the affected grandfather has allele $M_2$ at the $M$ locus and the mother is $DM_1/dM_2$ is given by $x_1x_2$. Similarly, the probability
that the affected grandfather has allele $M_1$ at the $M$ locus and the mother is $DM_2/dM_1$ is $x_1x_2$. Thus, the total probability that a mother is informative or the proportion of informative mothers is

$$I = 2x_1x_2.$$  \hfill (1)

In the above computation we assumed that there are only two alleles at the $M$ locus, but the above result can easily be extended to the case of multiple alleles. Namely, if there are $n$ alleles and the frequency of the $i$-th allele ($M_i$) is $x_i$, then the proportion of informative mother is

$$I = \sum_{i=1}^{n} x_i (1-x_i)$$

$$= 1 - \sum x_i x_i.$$  \hfill (2)

It is noted that the maximum value of $I$ is attained when all allele frequencies are equal, i.e., $x_i = 1/n$. Thus, the maximum value of $I$ for $n$ alleles is $(n-1)/n$. This value is 0.5, 0.67, 0.75, 0.80, and 0.83 for $n=2$, 3, 4, 5, and 6, respectively. Therefore, the utility of marker genes is quite high if the number of alleles is large and allele frequencies are nearly equal.

Normal grandfather. When a mother comes to know her carrier status through genetic screening, her father may carry either gene $D$ or $d$ at the $D$ locus. In most instances, however, if the grandfather carries gene $d$, the mother will know her carrier status without having a genetic test. Therefore, we consider only the case where the grandfather is normal. In this case, if we neglect the possibility of new mutation, gene $d$ in the mother must have come from her mother (grandmother), who was heterozygous for $d$, but the linkage phase of the mother can still be determined by examining her father, as mentioned earlier. Clearly, the probability that the mother is $DM_1/dM_2$ and the grandfather is $DM_1$, on the condition that the mother is heterozygous for the $d$ gene and the grandfather is normal is $x_1x_2$. Similarly, the probability that the mother is $DM_1/dM_2$ and the grandfather is $DM_2$ on the same condition is $x_1x_2$. Therefore, the total frequency of informative mothers is the same as that of the case of affected grandfather and given by expression (1) or (2).

(2) Dominant markers

Affected grandfather. If there is dominance for the marker genes, we need information on the genotypes of both grandfather and grandmother to determine the linkage phase of the mother. Figure 2 shows pedigrees in which the linkage phase of the mother can be determined. In this figure $M$ and $m$ denote the dominant and recessive alleles at the $M$ locus, and the only thing we know about the mother before examination of the $M$ locus is that she is a carrier because her father (grandfather) had the disease. However, the double heterozygosity and linkage phase of the mother can be determined simultaneously if we know the phenotypes of the grandfather and grandmother for the $M$ locus. In the example given in Fig. 2(a) the mother's phenotype for the $M$ locus is $M$, so that $d$ and $M$ must have come from
the grandfather and be present on the same chromosome, whereas $D$ and $m$ must have come from the grandmother. It is noted that chromosome $Dm$ may come from either the $Dm/Dm$ or $Dm/dm$ grandmother. However, there is no need to distinguish between the two cases, since we are interested in the probability that the mother is a double heterozygote, given that she is heterozygous for the $d$ gene. In the case of Fig. 2(a) the genotype of the grandfather for the $M$ locus is $M$, whereas that of the grandmother is $mm$. Thus, the probability of observing this combination of grandfather and grandmother is $(1-x)x^2$, where $x$ is the frequency of allele $m$. It is noted that when the grandfather has allele $M$ at the $M$ locus, the grandmother's genotype for this locus must be $mm$. Otherwise, we cannot determine the linkage phase of the mother. Therefore, the expected frequency of informative mothers is also given by $(1-x)x^2$.

When the grandfather has allele $m$, the grandmother's genotype at the $M$ locus can be either $MM$ or $Mm$ (Fig. 2b), but only the gamete with $M$ produces a heterozygous mother for the $M$ locus. Therefore, the expected frequency of informative mothers is given by $(1-x)x$. If we consider both $M$ and $m$ grandfathers, the proportion of informative mothers is

$$I = x(1-x^2).$$

The maximum value of this $I$ is 0.385, which is attained when $x$ is $\sqrt{1/3}$. Therefore, dominant marker genes are less efficient than codominant genes.

Normal grandfather. We consider the case where the genotypes of grandfather and grandmother at the $D$ locus are $D$ and $Dd$, respectively. In this case, the requirement for the $M$ locus for determining the mother's linkage phase is the same as that for the case of affected grandfather. Therefore, the proportion of informative mothers is given by (3).

**AUTOSOMAL DOMINANT DISEASES**

(1) Codominant markers

In the case of autosomal dominant diseases, information on the genotypes of both parents as well as those of grandparents is required for linked marker genes to be useful. This can be seen from example (a) in Fig. 3, in which $D$ denotes a dominant disease gene. In this example the genotype of the marker locus is $M_1M_2$ for the mother and $M_1M_1$ for the father. Therefore, if $r$ is the recombination value
between the \( D \) and \( M \) loci, the offspring with genotype \( M_1M_2 \) is expected to have the dominant disease gene \( D \) with probability \( 1-r \).

**Case of two alleles.** The frequency of dominant disease genes is generally so rare that the homozygotes for \( D \) are virtually nonexistent. Therefore, we assume that all affected individuals are heterozygous. From Fig. 3, it is clear that both affected and normal grandparents can be either homozygous or heterozygous for the marker locus. If we consider the case where the affected grandparent is homozygous for allele \( M_1 \), the probability that the affected parent's genotype \( (DM_1/dM_2) \) can be determined is \( x_1^2x_2 \). Similarly, when the affected grandparent is \( M_2M_3 \), the probability that the affected parent is informative is \( x_1x_3^2 \). On the other hand, if the affected grandparent is \( M_1M_3 \), the probability is \( x_1x_2(x_1^2+x_2^2) \). Therefore, the total frequency of informative parents is

\[
I = 2x_1x_2(1 - x_1x_2).
\]  

In their study on autosomal codominant markers, Murphy and Chase (1975) thought that the spouse of the affected parent must be homozygous for a marker gene, and gave formula \( I = 2x_1x_2(1 - x_1x_2)(x_1^2 + x_2^2) \). In fact, if \( DM_1/dM_2 \) mates with \( M_1M_3 \) and the genotype of their offspring is \( M_1M_2 \), the marker genes are indeed useless for genetic counseling, as will be seen from Table 1. However, if the genotype is \( M_1M_3 \), he or she is expected to carry gene \( D \) with probability \( 1-r \), whereas an offspring with genotype \( M_2M_4 \) is expected to carry the disease gene with probability \( r \). Thus, the marker genes are useful for genetic counseling for 50 percent of the offspring. We therefore regard this type of family as informative.

If we include this type of family, the frequency of informative families is given by (4). The maximum value of this frequency is 0.375 when \( x_1 = 0.5 \). This is much larger than the value given by Murphy and Chase's formula, i.e. 0.192. Compared with the case of X-linked recessives, however, the \( I \) value is considerably smaller.

**Case of multiple alleles.** In the presence of multiple alleles at the marker locus the frequency of informative parents may be computed in the following way. We first note that the affected parent must be heterozygous for the marker locus and one of the genes comes from the grandfather and the other from the grandmother. Obviously, the chance that the affected parent is heterozygous for the \( i \)-th and \( j \)-th
Table 1. Relative frequencies of different genotypes from the mating $DM_1/dM_4 \times dM_2/dM_3$.

<table>
<thead>
<tr>
<th>Gamete</th>
<th>$dM_1$</th>
<th>$dM_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$DM_1$</td>
<td>$(1-r)/2$</td>
<td>$DM_1/dM_4$</td>
</tr>
<tr>
<td>$DM_4$</td>
<td>$r/2$</td>
<td>$DM_4/dM_4$</td>
</tr>
<tr>
<td>$dM_1$</td>
<td>$r/2$</td>
<td>$dM_1/dM_4$</td>
</tr>
<tr>
<td>$dM_2$</td>
<td>$(1-r)/2$</td>
<td>$dM_2/dM_4$</td>
</tr>
</tbody>
</table>

alleles ($M_iM_j$) is $2x_i x_j$. However, when both grandfather and grandmother are heterozygous for these alleles, the linkage phase of the affected parent cannot be determined. Therefore, the probability of this event must be subtracted from $2x_i x_j$.

The chance that both grandfather and grandmother are heterozygotes is $4x_i^2x_j^2$, but an affected parent from such a mating becomes heterozygous only with probability $1/2$. Therefore, the probability that an affected parent with $M_iM_j$ is informative is $2x_i x_j - 2x_i^2x_j^2$. Thus, the total frequency of informative parents is

$$I = 2\sum_{i<j}(x_i x_j - x_i^2 x_j^2)$$

$$= 1 - \sum x_i^2 - (\sum x_i)^2 + \sum x_i a.$$  \hspace{1cm} (5)

The maximum value of $I$ is obtained when all allele frequencies are equal, i.e., $x_i = 1/n$. The maximum values for $n = 2, 3, 4, 5$, and 6 are 0.375, 0.593, 0.703, 0.768, and 0.810, respectively. Therefore, if there are many alleles, marker genes can be quite useful.

(2) Dominant markers with two alleles

Dominance reduces the frequency of informative parents considerably, since it makes the distinction of genotypes difficult. Furthermore, if an affected parent mates with a dominant homozygote for a marker locus, marker genes are totally uninformative. In the case where $DM/dm$ or $Dm/dM$ mates with $dM/dm$, the family is only partially informative. Namely, in the mating $Dm/dM \times dM/dm$, the offspring with phenotype $m$ (genotype $mm$) is expected to carry disease gene $D$ with probability $1-r$, whereas the offspring with phenotype $M$ is expected to carry $D$ with probability $(1+r)/3$. Thus, marker genes are not very useful if the offspring shows a dominant phenotype. Moreover, the distinction between $MM$ and $Mm$ for a spouse is not possible. Therefore, unless the frequency of $m$ is very high, such "partially informative families" are not really informative. For this reason, I shall not consider these partially informative families in this paper.

In the case of two alleles there are only two types of informative families. They are shown in Fig. 4. It is easy to show that the probability of obtaining these families or the frequency of informative families on the condition that the affected parent is heterozygous for the $D$ gene and his or her spouse is $dd$ is

$$I = 2(1-x)x^4,$$  \hspace{1cm} (6)
Fig. 4. Pedigrees in which autosomal dominant markers may be used for genetic counseling. Symbols in parentheses stand for the phenotype of each individual.

1. $dA/dO$, $dA/dA$, $dA/dB \times DO/dO$
   \begin{align*}
   & (A) \quad (A) \quad (AB) \quad (D,O) \\
   & DO/dA \times dO/dO, dB/dO, dB/dB \\
   & (D,A) \quad (B) \quad (B)
   \end{align*}

2. $dO/dO$, $dB/dO \times DA/dO$, $dA/dA$
   \begin{align*}
   & (O) \quad (B) \quad (D,A) \quad (D,A) \\
   & DA/dO \times dO/dO, dB/dO, dB/dB \\
   & (D,A) \quad (B) \quad (B)
   \end{align*}

3. $dO/dO$, $dB/dO \times DA/dB$
   \begin{align*}
   & (O) \quad (B) \quad (D,AB) \\
   & DA/dO \times dO/dA, dB/dO, dB/dB \\
   & (D,A) \quad (B) \quad (B)
   \end{align*}

4. $dB/dO$, $dB/dB$, $dA/dB \times DO/dO$
   \begin{align*}
   & (B) \quad (B) \quad (AB) \quad (D,O) \\
   & DO/dB \times dO/dO, dA/dO, dA/dA \\
   & (D,B) \quad (B) \quad (A)
   \end{align*}

5. $dO/dO$, $dB/dO \times DA/dO$, $DB/dB$
   \begin{align*}
   & (O) \quad (A) \quad (D,B) \quad (D,B) \\
   & DB/dO \times dO/dA, dA/dO, dA/dA \\
   & (D,B) \quad (B) \quad (A)
   \end{align*}

6. $dO/dO$, $dA/dO \times DB/dA$
   \begin{align*}
   & (O) \quad (B) \quad (D,AB) \\
   & DB/dO \times dO/dO, dA/dO, dA/dA \\
   & (D,B) \quad (O) \quad (A)
   \end{align*}

7. $dA/dO$, $dA/dA$, $dA/dB \times DB/dO$, $DB/dB$
   \begin{align*}
   & (A) \quad (AB) \quad (D,B) \quad (D,B) \\
   & DB/dA \times all genotypes \\
   & (D,AB)
   \end{align*}

8. $dA/dO$, $dA/dA \times DB/dA$
   \begin{align*}
   & (A) \quad (AB) \quad (D,AB) \\
   & DB/dA \times all genotypes \\
   & (D,AB)
   \end{align*}

9. $dB/dO$, $dB/dB$, $dA/dB \times DA/dO$, $DA/dA$
   \begin{align*}
   & (B) \quad (B) \quad (AB) \quad (D,A) \quad (D,A) \\
   & DA/dB \times all genotypes \\
   & (D,AB)
   \end{align*}

10. $dB/dO$, $dB/dB \times DA/dB$
    \begin{align*}
    & (B) \quad (B) \quad (D,AB) \\
    & DB/dA \times all genotypes \\
    & (D,AB)
    \end{align*}

Fig. 5. Informative families for the ABO locus. Symbols in parentheses stand for the phenotype of each individual. The phenotype for the disease locus is given only for the affected.
where \( x \) is the gene frequency of \( m \). \( I \) has a maximum value of 0.164 when \( x \) is 0.8.

In this connection it should be noted that for X-linked genes and autosomal codominant genes the frequency of informative families is the same as the frequency of informative parents.

(3) ABO blood groups as markers

In the presence of dominance it is difficult to derive a general formula for the case of multiple alleles, since the dominance relationship among alleles may vary from locus to locus. In the following I consider only the ABO blood group locus. Since this locus is highly polymorphic in most human populations, it is expected to be important in the future as a marker locus. This locus is known to have four alleles \( A_1, A_2, B, \) and \( O \), but we treat it as a three-allele locus, pooling \( A_1 \) and \( A_2 \) as a single class of \( A \), since in many populations the frequencies of \( A_1 \) and \( A_2 \) are not known. It also simplifies the mathematical treatment considerably.

Figure 5 shows all informative families possible for this locus. They are classified into ten different types according to the genotypes of the affected parent and grandparent. In general we do not know the genotypes of the individuals involved, but if we know their phenotypes, the linkage phase of the parent with the disease gene can be determined. If we denote the gene frequencies of \( A, B, \) and \( O \) by \( x_1, x_2, \) and \( x_3 \), respectively, the total frequency of informative families is given by

\[
I = x_1(1-x_1)^2x_3(1-x_1+x_3) + x_2(1-x_2)^2x_3(1-x_2+x_3) + 2x_1x_2(1-x_1x_2).
\]

NUMERICAL EXAMPLES

Despite a great effort for making a human linkage map in recent years, the number of genes which are known to be linked with genetic diseases is still small. Table 2 shows examples of these linkages. Let us now consider how useful these linkages are for genetic counseling.

X-linked genes. McKusick (1975) lists about 100 X-linked genes, but only four genes (or gene complexes) i.e., those for colorblindness, G6PD, the Xg blood groups, and the Xm serum groups, are known to be polymorphic. Among these, the gene for colorblindness, G6PD and the Xm serum groups are closely linked with the hemophilia gene (Table 2). However, the anti-Xm serum is not available now, since the rabbit which made it died some years ago. Therefore, only colorblindness and G6PD can be used for predicting hemophilia males or female carriers. The G6PD gene is known to be polymorphic in areas where malaria is endemic. In some areas of Saudi Arabia the frequency of G6PD deficiency genes is about 0.50 (Mourant et al., 1976). Since the heterozygotes for this gene cannot be identified, we can see from formula (3) that 37.5 percent of hemophilia males in this area can
Table 2. Examples of linked genes in man and their recombination values. All autosomal disease genes are dominant except the last two. From Race and Sanger (1975) and others.

<table>
<thead>
<tr>
<th>Loci involved</th>
<th>Recombination value</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td></td>
</tr>
<tr>
<td>X-linked genes</td>
<td></td>
</tr>
<tr>
<td>Colorblindness—Hemophilia</td>
<td>~0</td>
</tr>
<tr>
<td>G6PD—Hemophilia</td>
<td>5</td>
</tr>
<tr>
<td>Xm—Hemophilia</td>
<td>7</td>
</tr>
<tr>
<td>Xg—X-born khitthyosis</td>
<td>10</td>
</tr>
<tr>
<td>Xg—Ocular albinism</td>
<td>15</td>
</tr>
<tr>
<td>Xg—Retinoschisis</td>
<td>27</td>
</tr>
<tr>
<td>Xg—Fabry’s disease</td>
<td>27</td>
</tr>
<tr>
<td>II. Autosomal genes</td>
<td></td>
</tr>
<tr>
<td>Lu—Myotonic dystrophy</td>
<td>13</td>
</tr>
<tr>
<td>Se—Myotonic dystrophy</td>
<td>8</td>
</tr>
<tr>
<td>Rh—Elliptocytosis-1</td>
<td>~3</td>
</tr>
<tr>
<td>ABO—Nail patella</td>
<td>10</td>
</tr>
<tr>
<td>MN—Sclerotylosis</td>
<td>4</td>
</tr>
<tr>
<td>Fy—Zonular cataract</td>
<td>~0</td>
</tr>
<tr>
<td>i—Congenital cataract</td>
<td>~0</td>
</tr>
<tr>
<td>GPT—Epidermolysis bullosa (Ogna type)</td>
<td>5</td>
</tr>
<tr>
<td>HLA(B)—Adrenal hyperplasia (III)</td>
<td>~0</td>
</tr>
<tr>
<td>Gm—α₁ antitrypsin deficiency</td>
<td>10</td>
</tr>
</tbody>
</table>

be diagnosed in terms of this polymorphism. In some areas of southwestern India the frequency of this deficiency is about 0.10. Therefore, the frequency of informative mothers is 9.9 percent.

In northern Europe (e.g. England) the G6PD deficiency gene is virtually nonexistent, whereas the frequency of colorblindness genes is about 0.07. Colorblindness is recessive, so that the frequency of informative mothers becomes 7 percent from formula (3). Therefore the colorblindness polymorphism is not very useful for genetic counseling.

**Autosomal genes.** Among autosomal genes, the secretor polymorphism (Se—se) can be used for detecting myotonic dystrophy. Myotonic dystrophy is a dominant disease and expressed generally in middle age. At the Se locus allele se is recessive to allele Se, and the frequency of se in Europe is about 0.48. Thus, the frequency of informative families when this locus is used is 2.7 percent from formula (6). This indicates that autosomal dominant markers are not very useful for genetic counseling even if they are quite polymorphic. For detecting myotonic dystrophy, the Lutheran blood groups are a little more informative than the Se locus. The frequency of the recessive allele Lu₄ for this locus in Europe is 0.96. Thus, the frequency of informative families is 3.3 percent.
Table 3. Gene frequencies at the ABO locus and the frequency of informative families \( (I) \) in several human populations. Gene frequency data are from Mourant et al. (1976).

<table>
<thead>
<tr>
<th>Population</th>
<th>Gene frequency</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>English</td>
<td>.26</td>
<td>.06</td>
</tr>
<tr>
<td>Russian (Moscow)</td>
<td>.26</td>
<td>.16</td>
</tr>
<tr>
<td>Hindus</td>
<td>.15</td>
<td>.29</td>
</tr>
<tr>
<td>Japanese</td>
<td>.28</td>
<td>.17</td>
</tr>
<tr>
<td>American Indian (Maya)</td>
<td>.04</td>
<td>.01</td>
</tr>
<tr>
<td>Australian Aborigines</td>
<td>.10</td>
<td>.08</td>
</tr>
</tbody>
</table>

Generally speaking, the ABO locus is much more useful for genetic counseling than the Se and Lu Loci. This locus is known to be fairly closely linked with the gene for nail patella, a dominant genetic disease. Table 3 shows the gene frequencies at this locus for various human populations and the frequencies of informative families. It is clear that the ABO locus is quite useful except in the Maya tribe of American Indians.

The MN locus, which is codominant and closely linked to the sclerostosis gene (dominant), is also quite informative. For example, the frequencies of alleles \( M \) and \( N \) in Japan are 0.543 and 0.457, respectively. The probability of informative parents is 0.373, which is higher than that for the ABO blood groups.

**DISCUSSION**

We have seen that the utility of linked marker genes for genetic counseling depends on the gene frequencies and dominance relationship at the marker locus as well as on the recombination value between the marker and disease loci. If a marker locus has many codominant alleles, it is quite useful. In the past ten years many polymorphic enzymes were discovered in human populations by using electrophoresis. The genes for these enzymes are generally codominant, so that if they are polymorphic, they should be very useful for genetic counseling. The loci for blood groups often involve dominance, but there are many highly polymorphic loci such as MNSs and Rh. These loci seem to be as useful as those for enzyme loci. To use this technique for genetic counseling, however, more effort should be made to find new marker loci and establish linkage relationships between marker and disease loci.

In the present paper I have shown how to compute the proportion of informative families. If this proportion is computed for all marker genes that are linked with a particular disease locus, we will be able to know which marker locus or loci are most useful for genetic counseling. The gene frequencies at marker loci are known to vary from population to population. Therefore, the frequencies of informative families should be computed for each population separately.
If a particular disease gene is known to be linked with a number of marker genes, it is possible to use all the marker genes simultaneously to increase the reliability of diagnosis of a disease. The computation of the utility of marker genes for this case seems to be quite complicated. Even the identification of informative families is not simple. However, if we consider the cost of having patients with genetic diseases, it is certainly worth studying.

In the present study I assumed linkage equilibrium between the disease and marker loci. Since these two loci are generally unrelated functionally, this assumption seems to be satisfactory in most cases. In relatively small populations, however, linkage disequilibrium may be generated by random genetic drift. If there is linkage disequilibrium, a more detailed analysis has to be done by considering gamete frequencies rather than gene frequencies.

I thank Aravinda Chakravarti for helpful discussions. This work was supported by the Public Health Service Research Grant GM 19513 and GM 20293. An original version of this paper was presented at the 41st meeting of the International Statistical Institute, New Delhi, India, 1977.

REFERENCES