Ordered Genotypes: An Extended ITO Method and a General Formula for Genetic Covariance

Feng Dai[1] and Daniel E. Weeks[1,2]

Departments of [1]Biostatistics and [2]Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh

Traditionally, the stochastic ITO transition matrices provide a simple general method for obtaining the joint genotype distribution and genotypic correlations between any specified pair of noninbred relatives. The ITO method has been widely used in modern genetic analysis; however, since it was originally derived for unordered genotypes, it is not very useful in some new applications—for example, when one is modeling genomic imprinting and must keep track of the parental origin of alleles. To address these new, emerging problems, here we extend the ITO method to handle ordered genotypes. Our extended method is applied to calculate the covariance in unilineal and bilineal relatives under genomic imprinting, and some generalized linear functions of the transition matrices are given. Since the ITO method is limited to biallelic loci and to unilineal and bilineal relatives, we derive a general formula for calculating the genetic covariance using ordered genotypes for any type of relative pair.

The "ITO" paper by Li and Sacks[1] provides an elegant algorithm for deriving joint genotype probabilities between pairs of relatives. With the ITO method, given the genotype of an individual, it is possible to derive the conditional probability of the genotypes of any noninbred relative of that individual: $P(G_2|G_1)$, where G_i denotes the genotype of the ith person. The ITO method was extended to handle multiple alleles and was generalized for inbred populations.[2] The ITO method was generalized for multiple loci and was also extended to handle consanguinity.[3]

Although the ITO method has been widely used to solve various problems in human genetics, it uses only unordered genotypes; that is, the genotypes are unordered in the sense that maternal and parental contributions are not distinguished. However, some new applications require the use of ordered genotypes; for example, when one is modeling genomic imprinting, one must keep track of the parental (ordered) origin of alleles. Genomic imprinting occurs when the functional activity of a person's allele depends on whether it was inherited maternally or paternally. Strong genomic imprinting renders an imprinted locus effectively haploid and thereby causes certain genetic diseases, including disorders affecting cell growth, development, and behavior.[4] From the point of view of quantitative genetics, the effect of genomic imprinting is to make the phenotypical values of reciprocal heterozygotes different, which means various basic values of genetic quantities, as well as correlations, are not the same as the standard values. This difference may be crucial, especially in human quantitative genetics.

Although Campbell and Elston[3] did attempt to extend the ITO method to handle ordered genotypes, their extension is flawed, because of an incorrect assumption that does not generalize (as we explain in detail below). Li[5] revised the 4 × 4 Li-Sacks matrices[1] to 2 × 2 matrices by focusing on allele identity by descent (IBD) instead of genotype IBD. However, Li[5] still did not consider ordered genotypes.

More-accurate modeling of underlying biological processes should lead to more-accurate and more-powerful inferences. In the present study, we extended the ITO method to handle ordered genotypes. We derived some generalized linear functions of the transition matrices for deriving the probabilities of an individual's genotype, conditional on a relative's genotype. In the application part of this work, our extended method is applied to calculate the covariance between both unilineal and bilineal relatives under imprinting. Although the ITO approach is pleasing in terms of its clarity and understandability, it is difficult to extend it to handle loci with multiple alleles, as well as to handle very complex inbred relative pairs. Therefore, we also derive a completely

Received December 22, 2005; accepted for publication March 7, 2006; electronically published April 25, 2006.

Address for correspondence and reprints: Feng Dai, Parran Hall, Room 309B, Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, 130 DeSoto Street, Pittsburgh, PA 15261. E-mail: fed1@pitt.edu

Am. J. Hum. Genet. 2006;78:1035–1045. 0002-9297/2006/7806-0013$15.00

general formula for the genetic covariance, using ordered genotypes for any type of relative pair; this uses the approach of Gillois,[6] as more recently elucidated by Lange.[7] The resulting covariance equations can be easily applied in a variance component-based linkage analysis that takes genomic imprinting into account (not shown here).

Here, we notationally distinguish ordered and unordered genotypes. Accordingly, we let *A/a*, with a slash, represent an *unordered* genotype and let *A|a*, with a vertical bar, denote an *ordered* genotype. Here, without loss of generality, the maternal allele is listed to the left of the vertical bar and the paternal allele is listed to the right.

Two diploid outbred related individuals may share (1) both genes IBD, (2) one gene IBD, or (3) no genes IBD. If we denote transition matrices as matrices of conditional probabilities, then the three basic transition matrices corresponding to the number of identical genes shared in common by the two relatives are as follows.[1]

The matrix *I*:

	G_2 =		
G_1	*A/A*	*A/a*	*a/a*
A/A	1	0	0
A/a	0	1	0
a/a	0	0	1

The matrix *T*:

	G_2 =		
G_1	*A/A*	*A/a*	*a/a*
A/A	p	q	0
A/a	$\frac{p}{2}$	$\frac{1}{2}$	$\frac{q}{2}$
a/a	0	p	q

The matrix *O*:

	G_2 =		
G_1	*A/A*	*A/a*	*a/a*
A/A	p^2	$2pq$	q^2
A/a	p^2	$2pq$	q^2
a/a	p^2	$2pq$	q^2

Following convention, in these three matrices, p and q represent the allele frequencies of A and a in the population, respectively, with $p + q = 1$. The first matrix, *I*, $P(G_2|G_1$,share 2 IBD), gives the genotype transition probabilities for two relatives when they share two alleles IBD with person 1's genotype given. In such a case, their genotypes are necessarily identical. When one is given to be A/A, the other must be the A/A, etc. The second matrix, *T*, $P(G_2|G_1$,share 1 IBD), gives the transition probabilities from one relative to the other when they share one gene IBD. Suppose the given relative is of genotype A/a (second row of *T*), then the other relative must share allele A or allele a in common with probability 0.5 plus a random allele from the population. The third matrix, *O*, $P(G_2|G_1$,share 0 IBD), gives the conditional probabilities when the two individuals do not have any alleles IBD in common. Hence, they are genetically unrelated individuals. Regardless of the genotype of one individual, the probabilities of the other individual having the genotypes (A/A, A/a, a/a) remain simply p^2, $2pq$, and q^2, respectively, under Hardy-Weinberg equilibrium (HWE).

With these three basic matrices, it is then straightforward to find the joint distribution and correlation between any pair of unilineal relatives (unordered genotypes) in a random-mating population.[1] For autosomal genes, the general expression of the transition matrix for a specific pair of relatives is given as

$$R = c_I I + c_T T + c_o O \ ,$$

where c_I, c_T, and c_o are the probabilities that the two specified relatives share both, one, and no genes IBD, respectively, with $c_I + c_T + c_o = 1$.

Since the original ITO matrices were derived for unordered genotypes, they are not useful when one is modeling imprinting, since one must then keep track of parental origin. Thus, to take genomic imprinting into consideration, we must track where the IBD gene comes from by using ordered genotypes (where we list the maternal allele first). Since the heterozygote *A|a* may have a different genotypic value from *a|A*, Campbell and Elston[3] introduced four basic transition matrices with dimension 4 × 4 (instead of three 3 × 3 matrices), which are listed below. Each matrix element represents the probability of sibling 2 having the specific ordered genotype conditional on the genotype of sibling 1. The subscripts m and f represent male and female. We use the same notation *I* and *O* as defined in the 3 × 3 matrices discussed above. The matrix S_m, $P(G_2|G_1$,share 1 allele IBD through father), specifies the probabilities of sibling 2's genotypes conditional on sibling 1 sharing one allele IBD through the father. Similarly, the matrix S_f, $P(G_2|G_1$,share 1 allele IBD through mother), specifies the probabilities of sibling 2's genotypes conditional on sibling 1 sharing one allele IBD through the mother.

The matrix I:

	G_2 =			
G_1	$A\|A$	$A\|a$	$a\|A$	$a\|a$
$A\|A$	1	0	0	0
$A\|a$	0	1	0	0
$a\|A$	0	0	1	0
$a\|a$	0	0	0	1

The matrix S_m:

	G_2 =			
G_1	$A\|A$	$A\|a$	$a\|A$	$a\|a$
$A\|A$	p	0	q	0
$A\|a$	0	p	0	q
$a\|A$	p	0	q	0
$a\|a$	0	p	0	q

The matrix S_f:

	G_2 =			
G_1	$A\|A$	$A\|a$	$a\|A$	$a\|a$
$A\|A$	p	q	0	0
$A\|a$	p	q	0	0
$a\|A$	0	0	p	q
$a\|a$	0	0	p	q

The matrix O:

	G_2 =			
G_1	$A\|A$	$A\|a$	$a\|A$	$a\|a$
$A\|A$	p^2	pq	pq	q^2
$A\|a$	p^2	pq	pq	q^2
$a\|A$	p^2	pq	pq	q^2
$a\|a$	p^2	pq	pq	q^2

Campbell and Elston[3] proposed that the transition matrix R for any specified pair of relatives could be derived by the formula

$$R = c_I I + \frac{c_T}{2} S_m + \frac{c_T}{2} S_f + c_o O \ .$$

However, their formula for R is incorrect for some pairs of relatives, as we explain in detail below. Furthermore, a more complete derivation of the ITO method for ordered genotypes requires the specification of two additional matrices, T_f and T_m, which concern parent-offspring transitions. These matrices are as follows.

The matrix T_m:

	$G_{offspring}$ =			
G_{father}	A/A	A/a	a/A	a/a
A/A	p	0	q	0
A/a	$\frac{p}{2}$	$\frac{p}{2}$	$\frac{q}{2}$	$\frac{q}{2}$
a/A	$\frac{p}{2}$	$\frac{p}{2}$	$\frac{q}{2}$	$\frac{q}{2}$
a/a	0	p	0	q

The matrix T_f:

	$G_{offspring}$ =			
G_{mother}	A/A	A/a	a/A	a/a
A/A	p	q	0	0
A/a	$\frac{p}{2}$	$\frac{q}{2}$	$\frac{p}{2}$	$\frac{q}{2}$
a/A	$\frac{p}{2}$	$\frac{q}{2}$	$\frac{p}{2}$	$\frac{q}{2}$
a/a	0	0	p	q

The derivations of the matrices T_m, $P(G_{offspring}|G_{father})$, and T_f, $P(G_{offspring}|G_{mother})$, are straightforward. For example, if the genotype of the mother is $A|a$, the conditional probabilities that the offspring's genotypes are $A|A$, $A|a$, $a|A$, and $a|a$ are $\frac{p}{2}$, $\frac{q}{2}$, $\frac{p}{2}$, and $\frac{q}{2}$, respectively. The reason is that the alleles A and a from the mother each have a 50% chance of being transmitted to her offspring.

Li[5] showed how the 4 × 4 S_m and S_f matrices[3] can be derived as "'external tensors" of the 2 × 2 matrices. However, the matrices T_f and T_m cannot be derived via outer products. We have shown that $T_f(T_m)$ can be computed as a weighted sum of the $S_f(S_m)$ matrix and a permuted version of $S_f(S_m)$, where the middle two rows are switched (details omitted).

Now, with the above six ordered genotype 4 × 4 transition matrices, we can derive the conditional probabilities for two specified outbred relatives, in which we track the origin of both alleles at a locus. We first consider some simple unilineal relatives. For parent and offspring pair, since they share one gene identical by descent, the transition probabilities from parent to offspring for unordered genotypes are the elements of matrix T.[1] However, in the ordered genotype method, we have to consider separately the transition of paternal and maternal alleles. The elements of matrix T_m are now the transition probabilities from father to offspring, and the elements of matrix T_f are the transition probabilities from mother to offspring. Note that with ordered geno-

types, $P(G_{\text{father}}|G_{\text{offspring}})$ is not equal to $P(G_{\text{offspring}}|G_{\text{father}})$, whereas these two are equal with unordered genotypes.

Consider the transition probabilities from maternal grandmother to grandchild GC through mother M. If the genotype of the maternal grandmother is $A|A$, the resulting genotype of her daughter M will be $A|?$ (where ? is some unknown allele), so the total conditional probability that the grandchild GC is $A|a$ with allele A inherited from the mother M is

$$\left(\frac{1}{2}+\frac{1}{2}p\right)q \ .$$

The reason is that there is a 50% chance that grandchild GC receives the grandmaternal A allele from her mother M and a 50% chance that she receives her maternal grandfather's allele, which is A with probability p. So the conditional probability of the grandchild's ordered genotype being $A|a$ is then

$$\left(\frac{1}{2}+\frac{1}{2}p\right)q \ ,$$

where q is the probability of grandchild GC randomly inheriting the second allele (allele a) from his/her father.

It is much easier to understand if we do the above derivation with matrix manipulation. The probabilities of the mother are the elements of first row of matrix T_{f}, since we are conditioning on the grandmother's $A|A$ genotype. For each genotype of the mother M, the elements of second column of T_{f} are the probabilities of the genotype of the grandchild GC being $A|a$. The total conditional probability for the grandchild having genotype $A|a$ given grandmother's $A|A$ genotype via the grandchild's mother is the product of the first row and second column of the transition matrix T_{f}

$$(p \ \ q \ \ 0 \ \ 0)\begin{pmatrix} q \\ \frac{q}{2} \\ \frac{q}{2} \\ 0 \end{pmatrix} = pq+\frac{q^2}{2} = \left(\frac{1}{2}+\frac{1}{2}p\right)q \ .$$

By the same algorithm, the conditional probabilities for a grandchild given a specific genotype for the maternal grandmother are given by the elements of the product matrix $T_{\text{f}} \times T_{\text{f}} = T_{\text{f}}^2$. In the same manner, the conditional probabilities of a grandchild's genotypes given a specific genotype for the maternal grandfather (via his/her mother) are given by the elements of the product matrix $T_{\text{m}} \times T_{\text{f}}$.

For the above grandmother-grandchild pair relationship, Campbell and Elston's[3] formula (p. 229)

$$R = \frac{S_{\text{m}}}{4}+\frac{S_{\text{f}}}{4}+\frac{O}{2}$$

gives

$$\left(\frac{1}{4}+\frac{1}{2}p\right)q$$

as the probability of the genotype of the grandchild being $A|a$ given that the genotype of the grandmother is $A|A$, which is clearly wrong. For another example, for an aunt-niece pair connected through the mother, $c_T = \frac{1}{2}$, but c_T does not split in half as Campbell and Elston suggest; rather, all its "weight" goes on the T_{f}, and the correct matrix is

$$R = \frac{T_{\text{f}}}{2}+\frac{O}{2} \ .$$

Li and Sacks[1] showed that

$$T^2 = \frac{T}{2}+\frac{O}{2} \ ,$$

which means a grandparent-grandchild pair shares one gene IBD and no gene IBD with an equal 50% chance. T^2 also gives the conditional probabilities for half siblings. They also showed that, in general,

$$T^{n+1} = \left(\frac{1}{2}\right)^n T+\left[1-\left(\frac{1}{2}\right)^n\right]O \ ,$$

where $n+1$ is the total number of generations between the two relatives. When ordered genotypes are used, similar equations hold. For example,

$$\begin{Bmatrix} T_{\text{m}}^2 = \frac{1}{2}T_{\text{m}}+\frac{1}{2}O, & T_{\text{f}}^2 = \frac{1}{2}T_{\text{f}}+\frac{1}{2}O \\ T_{\text{m}}T_{\text{f}} = \frac{1}{2}T_{\text{f}}+\frac{1}{2}O, & T_{\text{f}}T_{\text{m}} = \frac{1}{2}T_{\text{m}}+\frac{1}{2}O \\ T_{\text{m}}T_{\text{f}}T_{\text{f}} = \frac{1}{4}T_{\text{f}}+\frac{3}{4}O, & T_{\text{m}}T_{\text{f}}T_{\text{m}} = \frac{1}{4}T_{\text{m}}+\frac{3}{4}O \\ T_{\text{f}}T_{\text{m}}T_{\text{m}} = \frac{1}{4}T_{\text{m}}+\frac{3}{4}O, & T_{\text{f}}T_{\text{m}}T_{\text{f}} = \frac{1}{4}T_{\text{f}}+\frac{3}{4}O \end{Bmatrix} ,$$

where T_{m}^2 gives the conditional probabilities for half siblings who have same father but different mothers and

T_f^2 gives the conditional probabilities for half siblings who have same mother but different fathers. And, in general,

$$\left\{\begin{array}{c} T_{\mathrm{m(f)}}^{n+1} = (\frac{1}{2})^n T_{\mathrm{m(f)}} + \left[1 - \left(\frac{1}{2}\right)^n\right] O \\ (T_{i_1} T_{i_2} \cdots T_{i_n}) T_{\mathrm{m(f)}} = \left(\frac{1}{2}\right)^n T_{\mathrm{m(f)}} + \left[1 - \left(\frac{1}{2}\right)^n\right] O, \\ i_j \in \{\mathrm{m,f}\}, j = 1, \ldots, n \\ T_{\mathrm{m(f)}}^{n+1} \to O, \text{ as } n \to \infty \end{array}\right\},$$

where $n + 1$ is the number of generations between two relatives. When n is infinitely large, the conditional probabilities for two relatives are given by the elements of the matrix O; that is, the two relatives could be treated as two random samples from the general human population who are all unrelated to each other.

Now, we model bilineal relatives. Let us first consider the simple but most important type: full siblings. Since full siblings have a 25% chance of sharing two genes IBD and a 25% chance of sharing no genes IBD, they therefore have a 50% chance of sharing one gene IBD. However, we are dealing with ordered genotypes, so the two siblings have a 25% chance of sharing one maternal allele IBD and no paternal allele IBD, and vice versa. Thus, the transition matrix for full siblings is

$$\mathcal{S} = \frac{1}{4} I + \frac{1}{4} S_{\mathrm{m}} + \frac{1}{4} S_{\mathrm{f}} + \frac{1}{4} O \ . \qquad (1)$$

Other relatives who can share two genes IBD are double–first cousins whose parents are members of two sibships. There are six types of sibships in the general population, because of six different mating types.[1] With use of the same algorithm mentioned in figure 1 of the work of Li and Sacks[1] but with maternal and paternal alleles labeled, we derive the conditional matrix for double–first cousins, which is

$$D = \mathcal{S}^2 = \frac{1}{16} I + \frac{3}{16} S_{\mathrm{m}} + \frac{3}{16} S_{\mathrm{f}} + \frac{9}{16} O \ .$$

Next, we try to model the relationship for some unlineal relatives in which the conditional matrix $\mathcal{S}$ for full siblings and the matrices T_{m} and T_{f} are involved. We use the broad sense of the term "avuncular," which includes uncle-nephew, uncle-niece, aunt-nephew, and aunt-niece relationships. The conditional probabilities of a nephew's genotypes are then given by the product of $\mathcal{S}T_{\mathrm{m(f)}}$ conditional on the uncle's genotypes. Conversely, the conditional probabilities of the uncle's genotypes are given by elements of the product of $T_{\mathrm{m(f)}}\mathcal{S}$ conditional on the genotypes of the nephew. Whether T_{m} or T_{f} is the transition matrix involved depends on whether the mother or the father of the nephew is the "connecting" relative. We verified that $\mathcal{S}T_{\mathrm{m(f)}} = T_{\mathrm{m(f)}}\mathcal{S}$, which indicates that the uncle-nephew matrix is the same as the nephew-uncle transition matrix.[1] Through further multiplication of matrices, we can also prove "the most remarkable property"[1(p352)]; that is,

$$\mathcal{S}T_{\mathrm{m(f)}} = T_{\mathrm{m(f)}}\mathcal{S} = T_{\mathrm{m(f)}}^2 \ ,$$

which indicates that uncle-nephew relationships are the same as those for grandparent-grandchild or half siblings[1]; whether the transition matrix T_{m} or T_{f} is involved depends on whether the uncle is the nephew's paternal uncle or the nephew's maternal uncle. Extension of the above equations results in other important matrices,

$$T_{m(f)} S T_{\mathrm{m(f)}} = T_{\mathrm{m(f)}}^3 = \frac{1}{4} T_{\mathrm{m(f)}} + \frac{3}{4} O \ ,$$

whose elements give the conditional probabilities for first cousins and the probabilities for the great-grandchild conditional on one given great-grandparent.

Next, as an illustration of the utility of our extended ordered-genotype ITO method, we now derive equations for the genetic covariance between siblings and for the covariance between parent and child, with genomic imprinting considered. We begin with the standard genetic model and extend it to consider the case in which a locus is subject to imprinting. To derive the covariance formulas, it is necessary first to define the QTL model and its variance components. Here, we briefly review the results of Spencer.[8] Assume that an unobserved major gene has two alleles, allele A and allele a, with $P(A) = p$ and $P(a) = q$. In the standard genetic model, the genotypic value is a if the genotype is A/A, that of the a/a homozygote is $-a$, and that of heterozygote A/a is d. However, under imprinting, different genotypic values are possible for the two possible heterozygotes: d_1 for $A|a$ and d_2 for $a|A$ (fig. 1). It is usually assumed that $a \geqslant d_1$ and $d_2 \geqslant -a$. We have $d_1 = a$ ($d_2 = -a$) when there is complete

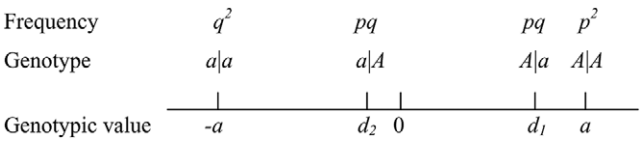

Figure 1 Genotypic values of the four possible genotypes

Table 1

Values of Genetic Components of Variance under Genomic Imprinting

Name	Expression	Definition
μ	$a(p-q)+(d_1+d_2)pq$	Mean phenotype of a population in HWE
α_{m}	$a+d_2q-d_1p$	Average effect of a gene substitution for males
α_{f}	$a+d_1q-d_2p$	Average effect of a gene substitution for females
$\sigma^2_{A\mathrm{m}}$	$2pq\alpha^2_{\mathrm{m}}$	Additive genetic variance for males
$\sigma^2_{A\mathrm{f}}$	$2pq\alpha^2_{f}$	Additive genetic variance for females
σ^2_D	$pq[pq(d_1+d_2)^2+(d_1-d_2)^2]$	Dominance genetic variance
$\sigma^2_{D\mathrm{m}}$	$pq[pq(d_1+d_2)^2+(d_1-d_2)^2]$	Dominance genetic variance for males
$\sigma^2_{D\mathrm{f}}$	$pq[pq(d_1+d_2)^2+(d_1-d_2)^2]$	Dominance genetic variance for females
σ^2_G	$pq[2\alpha_m\alpha_f+pq(d_1+d_2)^2+(d_1-d_2)^2]$	Overall genetic variance
$\sigma_{AD\mathrm{m}}$	$pq\alpha_{\mathrm{m}}(d_1-d_2)$	Covariance between dominance deviation and breeding value for males
$\sigma_{AD\mathrm{f}}$	$pq\alpha_{\mathrm{f}}(d_2-d_1)$	Covariance between dominance deviation and breeding value for females

NOTE.—The table reflects the work of Spencer.[8]

inactivation of the maternally (paternally) derived allele. A measure of imprinting is denoted as[9]

$$I=\frac{d_1-d_2}{2}\ .$$

Spencer[8] derived many useful genetic components of variance under imprinting, which are summarized in table 1. When $d_1=d_2=d$—that is, there is no imprinting ($I=0$)—the above various genetic values "revert" to their standard values. We further show that

$$\sigma^2_{A\mathrm{m}}+\sigma^2_{A\mathrm{f}}=2\sigma^2_a$$

and

$$\sigma^2_D+\sigma_{AD\mathrm{m}}+\sigma_{AD\mathrm{f}}=\sigma^2_d$$

(no constraints on I), where σ^2_a and σ^2_d are the additive genetic variance and dominance genetic variance, respectively, under the standard genetic model with no imprinting (details not presented here).

With the above definitions of different genetic variance components (table 1), we now begin to derive the covariance between siblings and between parent and offspring under genomic imprinting, using our transition matrices. Spencer[8] derived the covariance between parent and offspring under genomic imprinting. However, he did not derive the covariance between a pair of full siblings. As an illustration of the utility of our ordered ITO method, we first derive the covariance between sib pairs and also include a short part on deriving covariance between parent and offspring, to verify that our results match Spencer's results.

Let k_0 denote the probability of sibling 1 and sibling 2 sharing no alleles IBD, $k_{1\mathrm{m}}$ denote the probability of sibling 1 and sibling 2 sharing one paternal allele IBD, $k_{1\mathrm{f}}$ denote the probability of sibling 1 sharing one maternal allele IBD with sibling 2, and k_2 denote the probability of sibling 1 and sibling 2 sharing two alleles IBD. The probability of sibling 2 having a particular genotype (G_2) given the genotype of sibling 1 (G_1) can be calculated as

$$\begin{aligned}P(G_2|G_1)&=\sum_{i=0}^{2}P(G_2|G_1,\text{share } i \text{ alleles IBD})\\&\quad\times P(\text{share } i \text{ alleles IBD})\\&=k_2I+k_{1\mathrm{m}}S_{\mathrm{m}}+k_{1\mathrm{f}}S_{\mathrm{f}}+k_0O\ ,\end{aligned}$$

which, for full siblings, is the same as matrix $\mathcal{S}$ (in eq. [1] above).

Given the above conditional matrix, the joint probability of sibling 2 having a certain genotype s_2 and sibling 1 having another certain genotype s_1, $P(s_1,s_2)$ can be derived by multiplying the specific element of the above matrix $P(G_2|G_1)$ by the probability of having the certain genotype for sibling 1. The genetic covariance between a pair of siblings can be derived as

$$\begin{aligned}\mathrm{Cov}(s_1,s_2)&=E(s_1s_2)-E(s_1)E(s_2)\\&=\sum_{s_1}\sum_{s_2}s_1s_2P(s_1,s_2)-\mu^2\\&=\frac{k_{1\mathrm{m}}+k_2}{2}\sigma^2_{A\mathrm{m}}+\frac{k_{1\mathrm{f}}+k_2}{2}\sigma^2_{A\mathrm{f}}\\&\quad+k_2(\sigma^2_D+\sigma_{AD\mathrm{m}}+\sigma_{AD\mathrm{f}})\ .\end{aligned}\qquad(2)$$

For full siblings, $k_0=1/4$, $k_{1m}=1/4$, $k_{1f}=1/4$, and $k_2=1/4$; we get the following model from equation (2):

$$\mathrm{Cov}(s_1,s_2)=\frac{1}{4}\sigma^2_{A\mathrm{m}}+\frac{1}{4}\sigma^2_{A\mathrm{f}}+\frac{1}{4}(\sigma^2_D+\sigma_{AD\mathrm{m}}+\sigma_{AD\mathrm{f}})\ .\qquad(3)$$

As stated earlier, $\sigma^2_{A\mathrm{m}}+\sigma^2_{A\mathrm{f}}=2\sigma^2_a$ and $\sigma^2_D+\sigma_{AD\mathrm{m}}+$

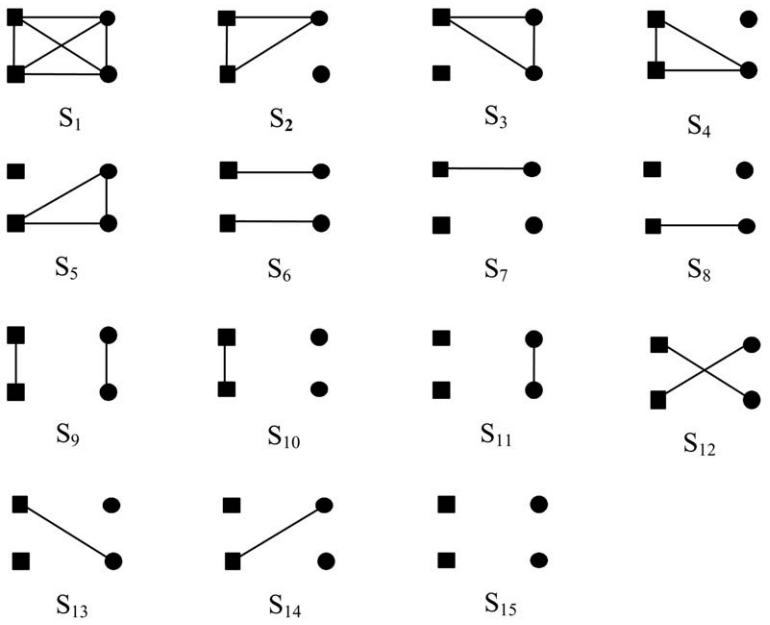

Figure 2 The 15 possible detailed identity states for individuals *i* and *j* with ordered genotypes. The squares represent maternal alleles, and the circles represent paternal alleles. Within each state, the top two symbols are *i*'s alleles, and the lower two symbols are *j*'s alleles. Lines connect alleles that are IBD (in the style of Sobel et al.[12] and Jacquard[13]).

$\sigma_{ADf} = \sigma_d^2$ (σ_a^2 and σ_d^2 are additive genetic variance and dominance genetic variance, respectively, under the standard genetic model). Thus, equation (3) simplifies to be exactly the standard genetic model, defined as

$$\text{Cov}(s_1,s_2) = \frac{1}{2}\sigma_a^2 + \frac{1}{4}\sigma_d^2 .$$

For half siblings, we distinguish between half siblings who share a mother ($k_0 = 1/2$, $k_{1f} = 1/2$, and $k_2 = 0$) and half siblings who share a father ($k_0 = 1/2$, $k_{1m} = 1/2$, and $k_2 = 0$). Equation (2) gives, respectively,

$$\text{Cov}(s_1,s_2) = \frac{1}{4}\sigma_{Af}^2$$

and

$$\text{Cov}(s_1,s_2) = \frac{1}{4}\sigma_{Am}^2 . \quad (4)$$

In the next section, we derive equations for the genetic covariance for a parent-offspring pair as follows. Let o and p_f be denoted as the genotypic values of offspring and mother, respectively. The joint probability of the offspring having a certain genotype and the mother having another genotype, $P(o,p_f)$, can be derived by multiplying the specific element of the matrix T_f—$P(G_{\text{offspring}}|G_{\text{mother}})$—by the probability that the mother has that genotype. Then, the covariance σ_{op_f} is calculated as

$$\begin{aligned}\text{Cov}(o,p_f) &= E(op_f) - E(o)E(p_f)\\ &= \sum_o \sum_{p_f} op_f P(o,p_f) - \mu^2\\ &= \frac{1}{2}\sigma_{Af}^2 + \frac{1}{2}\sigma_{ADf} . \end{aligned} \quad (5)$$

Similarly, we can derive the covariance equation for father and offspring:

$$\text{Cov}(o,p_m) = \frac{1}{2}\sigma_{Am}^2 + \frac{1}{2}\sigma_{ADm} . \quad (6)$$

Although historically attractive, the ITO method we generalized here is limited to biallelic loci and to unilineal and bilineal relatives. To tackle these limitations, we now introduce a more general way of calculating the covariance for any relative pair under imprinting. This is a generalization of the approach used by Gillois,[6] which was more recently summarized by Lange,[7] whose notation we use here.

First, suppose there are two or more alleles, let the

kth allele have population frequency p_k and the ordered genotype $k|l$ have the trait value $w_{k|l}$, then we can write

$$w_{k|l} = \alpha_k + \beta_l + \delta_{k|l} ,$$

where α_k is the additive impact of the maternal allele, β_l is the additive impact of the paternal allele, and $\delta_{k|l}$ is the residual departure from additivity. Under imprinting, the identity $w_{k|l} = w_{l|k}$ does not necessarily hold. No generality is lost if we adjust the trait mean to be 0, so that $\sum_k \sum_l w_{k|l} p_k p_l = 0$. The allelic contributions α_k and β_l are chosen to minimize the deviations $\delta_{k|l} = w_{k|l} - \alpha_k - \beta_l$. One way of doing this is to minimize the sum of squares

$$\sum_k \sum_l \delta_{k|l}^2 p_k p_l = \sum_k \sum_l (w_{k|l} - \alpha_k - \beta_l)^2 p_k p_l ,$$

which is achieved by taking $\alpha_k = \sum_l w_{k|l} p_l$ and $\beta_l = \sum_k w_{k|l} p_k$ (see appendix A).

Next, suppose individuals i and j are relatives. The covariance $\mathrm{Cov}(X_i,X_j)$ between the trait values X_i and X_j of i and j can be computed in the following steps. Here, we use the 15 detailed identity states of Gillois,[6] Harris,[10] and Jacquard.[11] Figure 2 (in the style of Sobel et al.[12] and Jacquard[13]) shows the 15 detailed identity states that are possible when maternal and paternal alleles are distinguished. The states vary from sharing no alleles IBD, S_{15}, to sharing all four alleles IBD, S_1.

Conditioning on these detailed identity states of the two relatives and using the identities $\sum_k \alpha_k p_k = 0$, $\sum_l \beta_l p_l = 0$, $\sum_k \delta_{k|l} p_k = 0$, and $\sum_l \delta_{k|l} p_l = 0$, we can deduce

$$\begin{aligned}
E(X_i,X_j) = {} & (\delta_1 + \delta_2 + \delta_4 + \delta_9 + \delta_{10}) \sum_k \alpha_k^2 p_k \\
& + (\delta_1 + \delta_3 + \delta_5 + \delta_9 + \delta_{11}) \sum_k \beta_k^2 p_k \\
& + (2\delta_1 + \delta_2 + \delta_3 + \delta_4 + \delta_5 + 2\delta_{12} + \delta_{13} + \delta_{14}) \sum_k \alpha_k \beta_k p_k \\
& + (2\delta_1 + \delta_2 + \delta_4) \sum_k \alpha_k \delta_{k|k} p_k \\
& + (2\delta_1 + \delta_3 + \delta_5) \sum_k \beta_k \delta_{k|k} p_k \\
& + \delta_1 \sum_k \delta_{k|k}^2 p_k \\
& + \delta_6 \sum_k \sum_l \delta_{k|k} \delta_{l|l} p_k p_l \\
& + \delta_9 \sum_k \sum_l \delta_{k|l}^2 p_k p_l \\
& + \delta_{12} \sum_k \sum_l \delta_{k|l} \delta_{l|k} p_k p_l ,
\end{aligned} \tag{7}$$

where δ_i is the probability of the ith detailed identity state.[11] A detailed derivation of equation (7) is given in appendix A. When there is no imprinting, equation (7) reduces to the general covariance equation derived by Gillois.[6]

Since trait means $E(X_i) = E(X_j) = 0$, the covariance $\mathrm{Cov}(X_i,X_j) = E(X_i,X_j)$, which is given in equation (7). If we assume that neither i nor j is inbred, we have $\delta_1 = \delta_2 = \ldots = \delta_8 = 0$. The covariance $\mathrm{Cov}(X_i,X_j)$ then simplifies to

$$\begin{aligned}
\mathrm{Cov}(X_i,X_j) = {} & (\delta_9 + \delta_{10}) \sum_k \alpha_k^2 p_k + (\delta_9 + \delta_{11}) \sum_k \beta_k^2 p_k \\
& + (2\delta_{12} + \delta_{13} + \delta_{14}) \sum_k \alpha_k \beta_k p_k \\
& + \delta_9 \sum_k \sum_l \delta_{k|l}^2 p_k p_l + \delta_{12} \sum_k \sum_l \delta_{k|l} \delta_{l|k} p_k p_l .
\end{aligned} \tag{8}$$

When there are two alleles, we can rewrite the summations in terms of our notation used above:

$$\left\{ \begin{array}{c}
\displaystyle\sum_k \alpha_k^2 p_k = \frac{1}{2}\sigma_{\mathrm{Af}}^2 \\
\displaystyle\sum_k \beta_k^2 p_k = \frac{1}{2}\sigma_{\mathrm{Am}}^2 \\
\displaystyle\sum_k \alpha_k \beta_k p_k = \frac{1}{2}\sigma_{\mathrm{Am}}^2 + \sigma_{AD\mathrm{m}} = \frac{1}{2}\sigma_{\mathrm{Af}}^2 + \sigma_{AD\mathrm{f}} \\
\displaystyle\sum_k \sum_l \delta_{k|l}^2 p_k p_l = \sum_k \sum_l \delta_{k|l} \delta_{l|k} p_k p_l = \sigma_d^2
\end{array} \right\} .$$

Thus, from our general equation (8), we obtain the same covariances as we derived above for full siblings (eq. [3]), half siblings (eq. [4]), mother-offspring pairs (eq. [5]), and father-offspring pairs (eq. [6]). Furthermore, our equation (7) can be used to generalize the variance-components model developed by Shete et al.[14] to handle all possible types of inbred relative pairs.

Once the detailed identity coefficients[11] are computed, any relative-to-relative covariance is expressible in terms of the theoretical variances and covariances defined above. An algorithm for computing these detailed identity coefficients (under the assumption that the entire pedigree structure connecting the two individuals is known) was derived by Nadot and Vaysseix.[15]

In summary, in this work, we extended the ITO method,[1,3] to handle ordered genotypes in an attempt to generalize this simple but useful method. We also showed that Campbell and Elston's previous formula[3] for the transition matrix R is incorrect for some pairs of relatives. In practice, a more complete derivation of the ITO method for ordered genotypes requires the specification of two additional matrices, T_{m} and T_{f}, which we derived

in this work. By tracking the paternal or maternal origin of each allele, we now have six basic transition matrices, with the help of which it is possible to derive conditional probabilities between two specified outbred relatives when we need to distinguish the two forms of the heterozygotes. In addition to providing an algorithm for deriving conditional probabilities with use of ordered genotypes, the ITO approach can be used to derive formulas for the genetic covariance between a pair of relatives. To complement the more limited ITO approach, we also derived a completely general formula, for the genetic covariance, using ordered genotypes; this formula is applicable to multiallelic loci and to any type of inbred relative pair.

We illustrated the utility of the extended ITO approach and our general covariance formula by using them to derive the genetic covariance under imprinting between parent and offspring and between sib pairs. The derived formulas for the covariance between a parent's and an offspring's genotypic values are the same as those given in the study by Spencer.[8] The consistency of our equations with previous work proves the applicability of our proposed method for the calculation of covariance between two relatives when we have to deal with ordered genotypes—for example, when we try to model genomic imprinting in human quantitative genetic analysis. Also, it should be noted that our work could help accurately test genetic hypotheses or predict risk (for genetic counseling), given a known genetic model.[3] Our extended ordered-genotype ITO method and our general covariance formula, with their easy applicability, will be helpful in modeling the complex relationship between relatives under the important biological phenomena (genomic imprinting) that need further statistical attention.

Acknowledgments

This work was supported by the University of Pittsburgh and by National Institutes of Health grants R01DK059642 and R01DC005630. We thank our reviewers for their helpful comments, and, in particular, thank Kenneth Lange for suggesting the elegant approach for deriving the general covariance formula. This work is dedicated to our late esteemed colleague, C. C. Li, one of the codevelopers of the original ITO method.

Appendix A

Covariance between Individuals *i* and *j* under Imprinting

Let the kth allele have population frequency p_k and the ordered genotype $k|l$ have trait value $w_{k|l}$, then we can write $w_{k|l} = \alpha_k + \beta_l + \delta_{k|l}$, where α_k is the additive impact of the maternal allele, β_l is the additive impact of the paternal allele, and $\delta_{k|l}$ is the residual departure from additivity. Under imprinting, the identity $w_{k|l} = w_{l|k}$ does not necessarily hold. No generality is lost if we take the trait mean

$$\sum_k \sum_l w_{k|l} p_k p_l = 0 \ .$$

To minimize

$$\sum_k \sum_l \delta_{k|l}^2 = \sum_k \sum_l (w_{k|l} - \alpha_k - \beta_l)^2 p_k p_l \ ,$$

the partial derivative with respect to α_k was taken, which gives

$$-2 \sum_k \sum_l (w_{k|l} - \alpha_k - \beta_l) p_k p_l = -2 \sum_k \sum_l \delta_{k|l} p_k p_l = 0 \ ,$$

which is true only if

$$\left\{ \begin{array}{l} \sum_k \delta_{k|l} p_k = 0, \sum_l \delta_{k|l} p_l = 0 \\ \sum_k \alpha_k p_k = 0, \sum_l \beta_l p_l = 0 \end{array} \right\} .$$

Using the above facts and because

$$\begin{aligned} 0 &= \sum_l \delta_{k|l} p_l \\ &= \sum_l w_{k|l} p_l - \sum_l \alpha_k p_l - \sum_l \alpha_l p_l \\ &= \sum_l w_{k|l} p_l - \alpha_k \ , \end{aligned}$$

we can conclude that $\alpha_k = \sum_l w_{k|l} p_l$. Similarly, $\beta_l = \sum_k w_{k|l} p_k$ also holds.

If X_i and X_j are traits of individuals i and j, since $E(X_i) = E(X_j) = 0$, then the covariance is

$$\begin{aligned} \mathrm{Cov}\,(X_i,X_j) &= E(X_i,X_j) \\ &= \delta_1 \sum_k (\alpha_k + \beta_k + \delta_{k|k})^2 p_k + \delta_2 \sum_k \sum_l (\alpha_k + \beta_k + \delta_{k|k})(\alpha_k + \beta_l + \delta_{k|l}) p_k p_l \\ &\quad + \delta_3 \sum_k \sum_l (\alpha_k + \beta_k + \delta_{k|k})(\alpha_l + \beta_k + \delta_{l|k}) p_k p_l \\ &\quad + \delta_4 \sum_k \sum_l (\alpha_k + \beta_l + \delta_{k|l})(\alpha_k + \beta_k + \delta_{k|k}) p_k p_l \\ &\quad + \delta_5 \sum_k \sum_l (\alpha_k + \beta_l + \delta_{k|l})(\alpha_l + \beta_l + \delta_{l|l}) p_k p_l \\ &\quad + \delta_6 \sum_k \sum_l (\alpha_k + \beta_k + \delta_{k|k})(\alpha_l + \beta_l + \delta_{l|l}) p_k p_l \\ &\quad + \delta_7 \sum_k \sum_l \sum_m (\alpha_k + \beta_k + \delta_{k|k})(\alpha_l + \beta_m + \delta_{l|m}) p_k p_l p_m \\ &\quad + \delta_8 \sum_k \sum_l \sum_m (\alpha_k + \beta_l + \delta_{k|l})(\alpha_m + \beta_m + \delta_{m|m}) p_k p_l p_m \\ &\quad + \delta_9 \sum_k \sum_l (\alpha_k + \beta_l + \delta_{k|l})^2 p_k p_l \\ &\quad + \delta_{10} \sum_k \sum_l \sum_m (\alpha_k + \beta_l + \delta_{k|l})(\alpha_k + \beta_m + \delta_{k|m}) p_k p_l p_m \\ &\quad + \delta_{11} \sum_k \sum_l \sum_m (\alpha_k + \beta_l + \delta_{k|l})(\alpha_m + \beta_l + \delta_{m|l}) p_k p_l p_m \\ &\quad + \delta_{12} \sum_k \sum_l (\alpha_k + \beta_l + \delta_{k|l})(\alpha_l + \beta_k + \delta_{l|k}) p_k p_l \\ &\quad + \delta_{13} \sum_k \sum_l \sum_m (\alpha_k + \beta_l + \delta_{k|l})(\alpha_m + \beta_k + \delta_{m|k}) p_k p_l p_m \\ &\quad + \delta_{14} \sum_k \sum_l \sum_m (\alpha_k + \beta_l + \delta_{k|l})(\alpha_l + \beta_m + \delta_{l|m}) p_k p_l p_m \\ &\quad + \delta_{15} \sum_k \sum_l \sum_m \sum_n (\alpha_k + \beta_l + \delta_{k|l})(\alpha_m + \beta_n + \delta_{m|n}) p_k p_l p_m p_n \\ &= (\delta_1 + \delta_2 + \delta_4 + \delta_9 + \delta_{10}) \sum_k \alpha_k^2 p_k + (\delta_1 + \delta_3 + \delta_5 + \delta_9 + \delta_{11}) \sum_k \beta_k^2 p_k \\ &\quad + (2\delta_1 + \delta_2 + \delta_3 + \delta_4 + \delta_5 + 2\delta_{12} + \delta_{13} + \delta_{14}) \sum_k \alpha_k \beta_k p_k \\ &\quad + (2\delta_1 + \delta_2 + \delta_4) \sum_k \alpha_k \delta_{k|k} p_k + (2\delta_1 + \delta_3 + \delta_5) \sum_k \beta_k \delta_{k|k} p_k \\ &\quad + \delta_1 \sum_k \delta_{k|k}^2 p_k + \delta_6 \sum_k \sum_l \delta_{k|k} \delta_{l|l} p_k p_l \\ &\quad + \delta_9 \sum_k \sum_l \delta_{k|l}^2 p_k p_l + \delta_{12} \sum_k \sum_l \delta_{k|l} \delta_{l|k} p_k p_l \ . \end{aligned}$$

References

1. Li CC, Sacks L (1954) The derivation of joint distribution and correlation between relatives by the use of stochastic matrices. Biometrics 10:347–360
2. Richardson WH (1964) Frequencies of genotypes of relatives, as determined by stochastic matrices. Genetica 35:323–354
3. Campbell MA, Elston RC (1971) Relatives of probands: models for preliminary genetic analysis. Ann Hum Genet 35:225–236
4. Reik W, Walter J (2001) Genomic imprinting: parental influence on the genome. Nat Rev Genet 2:21–32
5. Li W (1998) An exact calculation of the probability of identity-by-descent in two-locus models using an extension of the Li-Sacks' method. Am J Hum Genet 63:A297
6. Gillois M (1965) La relation d'identité en génétique. Ann Inst Henri Poincaré B 2:1–94
7. Lange K (2002) Mathematical and statistical methods for genetic analysis, 2nd ed. Springer-Verlag, New York
8. Spencer HG (2002) The correlation between relatives on the supposition of genomic imprinting. Genetics 161:411–417
9. Shete S, Amos CT (2002) Testing for genetic linkage in families by a variance-components approach in the presence of genomic imprinting. Am J Hum Genet 70:751–757
10. Harris DL (1964) Genotypic covariance between inbred relatives. Genetics 50:1319–1348
11. Jacquard A (1974) The genetic structure of populations. Springer-Verlag, New York
12. Sobel E, Sengul H, Weeks DE (2001) Multipoint estimation of identity-by-descent probabilities at arbitrary positions among marker loci on general pedigrees. Hum Hered 52:121–131
13. Jacquard A (1972) Genetic information given by a relative. Biometrics 28:1101–1114
14. Shete S, Zhou X, Amos CI (2003) Genomic imprinting and linkage test for quantitative-trait loci in extended pedigrees. Am J Hum Genet 73:933–938
15. Nadot R, Vaysseix G (1973) Apparentement et identité: algorithme du calcul des coefficients d'identité. Biometrics 29:347–359