Improved Inference of Relationship for Pairs of Individuals

Michael P. Epstein, William L. Duren, and Michael Boehnke

Department of Biostatistics, University of Michigan, Ann Arbor

Linkage analyses of genetic diseases and quantitative traits generally are performed using family data. These studies assume the relationships between individuals within families are known correctly. Misclassification of relationships can lead to reduced or inappropriately increased evidence for linkage. Boehnke and Cox (1997) presented a likelihood-based method to infer the most likely relationship of a pair of putative sibs. Here, we modify this method to consider all possible pairs of individuals in the sample, to test for additional relationships, to allow explicitly for genotyping error, and to include X-linked data. Using autosomal genome scan data, our method has excellent power to differentiate monozygotic twins, full sibs, parent-offspring pairs, second-degree (2°) relatives, first cousins, and unrelated pairs but is unable to distinguish accurately among the 2° relationships of half sibs, avuncular pairs, and grandparent-grandchild pairs. Inclusion of X-linked data improves our ability to distinguish certain types of 2° relationships. Our method also models genotyping error successfully, to judge by the recovery of MZ twins and parent-offspring pairs that are otherwise misclassified when error exists. We have included these extensions in the latest version of our computer program RELPAIR and have applied the program to data from the Finland-United States Investigation of Non-Insulin-Dependent Diabetes Mellitus (FUSION) study.

Introduction

Valid inference of genetic linkage requires correct relationship specification for the pairs of individuals in the study. Misclassification of relationships because of false paternity, unknown adoption, or sample switches can lead to a loss of power to detect linkage, caused either by inclusion of pairs who are less closely related than assumed or by exclusion of families through apparent failures of Mendelian inheritance. False evidence for linkage can be created if misclassification is due to sample duplications or incorrect assignment of monozygotic twins as full sibs. Therefore, it is important to ensure that the putative relationship of a given pair of individuals is correct.

Boehnke and Cox (1997) introduced a likelihood-based method for inferring the most likely relationship for putative sib pairs. They calculate the multipoint likelihood of the marker data for each pair conditional on each of four possible relationships: full sibs, monozygotic twins, half sibs, and unrelated pairs. To do so, they assume no genetic interference, so that the identity-by-descent (IBD) states at an ordered map of markers represent a nonhomogeneous Markov chain. The multipoint likelihood depends on population marker-allele

frequencies, intermarker distances, and the presumed relationship of the pair. The inferred relationship of the pair is that which maximizes this multipoint likelihood. Simulations revealed that this method yields accurate identification of these relationships under a wide range of marker number, heterozygosity, and intermarker distance. A FORTRAN 77 program, RELPAIR, was written to evaluate the multipoint likelihoods and to assess the most likely relationship between different putative sib pairs (Duren et al. 1997). A similar method was proposed independently by Göring and Ott (1997).

McPeek and Sun (2000) presented a similar method that also tested additional relationships within a family. The multipoint likelihood of the data for a given relationship is calculated using the same general method as Boehnke and Cox (1997). McPeek and Sun (2000) then use a likelihood-ratio statistic to test the putative relationship of a relative pair. For more distant relationships, such as avuncular and first-cousin relationships, the likelihood calculation is complicated by the fact that the IBD states at an ordered map of markers are not a Markov chain (Feingold 1993). McPeek and Sun (2000) solved this problem by creating augmented IBD processes that are Markovian for these more distant relationships.

Although these methods are useful for relationship identification, they can be improved and extended in several ways. First, rather than considering only pairs within the same family, we may wish to test all possible pairs of individuals in our sample. Testing all possible pairs may identify apparently independent families that are, in fact, related; it may also identify related indi-

viduals erroneously classified as unrelated because of sample switches or duplications. Second, we might allow explicitly for genotyping error rather than assuming that a relative pair is correctly genotyped for every marker under analysis. Failure to account for genotyping error even when only a few errors are present can lead to erroneous classification of MZ twins as full sibs and of parent-offspring pairs as grandparent-grandchild pairs. Third, we might want to include X-linked marker data in the multipoint probability calculations. For specific relationship-sex combinations, X-linked data may be particularly informative.

We have extended the method of Boehnke and Cox (1997) in each of these three ways and to test additional relationships. We have implemented these extensions in the computer program RELPAIR, version 2.0. For avuncular and first-cousin relationships, we approximate the likelihood by assuming (incorrectly) that the original IBD processes are Markovian. This approximate likelihood requires less computation time than the exact likelihood and has been shown to be an adequate substitute (McPeek and Sun 2000). Using simulated data, we examine classification rates of all tested relationships as a function of marker number and heterozygosity and intermarker distance, assess the importance of modeling genotyping error when it is present, and determine the value of incorporating X-linked marker data. We also illustrate the use of our method by application to data from the Finland–United States Investigation of Non-Insulin-Dependent Diabetes Mellitus (FUSION) study (Valle et al. 1998).

Materials and Methods

Assumptions and Definitions

We assume that a relative pair is typed for a collection of M codominant markers. Let θ_k be the recombination fraction between markers k and $k + 1$ ($1 \leq k \leq M - 1$). We assume that θ_k is known without error and is the same for both sexes. Also, let $\psi_k = \theta_k^2 + (1 - \theta_k)^2$. If X-linked data are included, we assume that the sex of each individual is known. Let X_k be the pair of genotypes at marker k, and $X = (X_1, X_2, ..., X_M)$ be all the genotype data for the pair. Finally, let I_k be the number of alleles shared IBD by the pair at marker k.

Probability of the Marker Data

We wish to calculate $P(X|R)$, the probability of the marker genotype data X for a pair of relationship R. We infer the relationship R^*, which maximizes $P(X|R)$. If R^* is different from the putative relationship, R_0, then the level of support for R^* over R_0 can be summarized conveniently by the likelihood ratio $P(X|R^*)/P(X|R_0)$.

To calculate $P(X|R)$, let $\alpha_k(i|R) = P(X_1, X_2, ..., X_{k-1}, I_k = i|R)$ be the joint probability of the marker data at the first $k - 1$ markers and that the pair shares i alleles IBD at marker k given relationship R. For the first marker ($k = 1$), $\alpha_1(i|R) = P(I_1 = i|R)$ is a simple function of R. For example, for an autosomal marker and $i = (0,1,2)$, $\alpha_1(i|\text{Full Sibs}) = (1/4, 1/2, 1/4)$, $\alpha_1(i|\text{Parent-Offspring}) = (0,1,0)$, and $\alpha_1(i|\text{First Cousins}) = (3/4, 1/4, 0)$. Analogous terms may be calculated for the X-linked case and are sex specific. For example, $\alpha_1(i|\text{Brothers}) = (1/2, 1/2, 0)$, whereas $\alpha_1(i|\text{Sisters}) = (0, 1/2, 1/2)$.

To evaluate $\alpha_k(i|R)$ for subsequent markers, we assume no genetic interference, so that, for most of the relationships we consider, the IBD states $I_1, I_2, ..., I_M$ form a (hidden) nonhomogeneous Markov chain. For such a chain, according to Baum's (1972) forward algorithm:

$$\alpha_{k+1}(j|R) = \sum_i \alpha_k(i|R) P(X_k|I_k = i)$$
$$\times\ P(I_{k+1} = j|I_k = i; R) . \qquad (1)$$

Here, $P(X_k|I_k = i)$ is the conditional probability of the data at marker k, given that the pair shares i alleles IBD at marker k. These probabilities are given in table 1 for an autosomal marker (Thompson 1975) and in table 2 for an X-linked marker. In the latter case, the probabilities are, again, sex specific. Note that $P(X_k|I_k = i)$ is independent of the relationship R. $P(I_{k+1} = j|I_k = i; R)$ denotes the transition probability that a pair of relationship R shares j alleles IBD at marker $k + 1$, given they share i alleles IBD at marker k. Transition probabilities for different relationships are presented in table 3 for autosomal data (Risch 1990) and in tables 4–6 for X-linked data for female-female, male-male, and male-female pairs, respectively.

The joint likelihood of the marker data conditional

Table 1

Probabilities for Ordered Autosomal Genotype Pairs

	$P(X_k|I_k)$ FOR		
GENOTYPE X_k	$I_k = 0$	$I_k = 1$	$I_k = 2$
(aa,aa)	p_a^4	p_a^3	p_a^2
(aa,ab)	$2p_a^3 p_b$	$p_a^2 p_b$	0
(aa,bb)	$p_a^2 p_b^2$	0	0
(aa,bc)	$2p_a^2 p_b p_c$	0	0
(ab,ab)	$4p_a^2 p_b^2$	$p_a p_b(p_a + p_b)$	$2p_a p_b$
(ab,ac)	$4p_a^2 p_b p_c$	$p_a p_b p_c$	0
(ab,cd)	$4p_a p_b p_c p_d$	0	0

NOTE.—X_k denotes the genotypes for the relative pair at an autosomal marker locus k. a, b, c, and d are distinct alleles at that locus with population frequencies p_a, p_b, p_c, and p_d.

on relationship R, $P(X|R)$, is obtained by the final summation as

$$P(X|R) = \sum_i \alpha_M(i|R)P(X_M|I_M = i) \ .$$

For avuncular and first-cousin relationships, (1) is only approximately correct, since, for those relationships, $I_1, I_2, ..., I_M$ are not a Markov chain (Feingold 1993).

Genotyping Error

To allow for genotyping error, we assume that each marker genotype is determined correctly for certain with probability $1 - \epsilon$, and is determined at random according to population genotype frequencies with probability ϵ. To allow for this random-genotype-error model in the calculation of $P(X|R)$ in equation (1), the only component altered is $P(X_k|I_k)$. If each member of the pair is correctly genotyped for marker k, $P(X_k|I_k)$ is the same as before. However, if either member is randomly genotyped for marker k, then the pair is effectively unrelated. Hence,

$$P(X_k|I_k = j; \epsilon)$$
$$= (1 - \epsilon)^2 P(X_k|I_k = j; \epsilon = 0)$$
$$+ [1 - (1 - \epsilon)^2]P(X_k|I_k = 0; \epsilon = 0) \quad (2)$$

This model was used previously by Broman and Weber (1998).

Simulations

To determine the accuracy of our method in identification of relationships, we performed computer simulations. Marker data were generated for 100,000 relative pairs for each of the following relationships: monozygotic twins (MZ), full sibs (FS), parent-offspring (PO), grandparent-grandchild (GG), half sibs (HS), avuncular (AV), first cousins (FC), and unrelated (UN). We simulated maps of genetic markers with either two or four equally frequent alleles spaced at 5-, 10-, or 20-cM intervals. The positioning of the markers began at the telomere of the short arm of chromosome 1 and proceeded down the chromosome. When no more markers could be placed on chromosome 1, the next marker was placed on the telomere of the short arm of chromosome 2, and so on, along the entire autosomal genome or until the number of markers desired was placed. When X-linked data were included, this positioning process continued along the X chromosome. We used chromosome lengths from Morton (1991) and Kosambi's (1944) mapping function to relate map distance and recombination fraction. If a 10-cM map is assumed, a total

of 399 autosomal and 23 X-linked simulated markers can be placed along the genome. To investigate the impact of genotyping error, we considered random-genotyping rates ϵ of 0 or .01.

For a given simulated relationship R, each of the 100,000 relative pairs was analyzed, using our multipoint method for each of the relationships listed above. Our method inferred the relationship that maximized the multipoint likelihood. For data simulated with error, we analyzed the data four times: assuming no random genotyping ($\epsilon = 0$), assuming the true genotyping-error rate ($\epsilon = .01$), underestimating the true genotyping-error rate ($\epsilon = .001$), and overestimating the true genotyping-error rate ($\epsilon = .02$).

Application to FUSION Data Set

The FUSION data to which we applied our method consist of 580 families with 2,118 genotyped individuals. The genome scan included 456 autosomal markers with average heterozygosity of 0.773 and average intermarker distance of ~9 cM. Allele frequencies were estimated by gene counting, ignoring family relationships. Marker order and intermarker distances were estimated using MultiMap (Matise et al. 1994) on a combination of FUSION and CEPH data.

To reduce misclassification, we limited our analyses to pairs of individuals that shared $\geqslant 100$ genotyped markers in common. Under this criterion, RELPAIR analyzed a total of 2,206,937 pairs of individuals in the data set. Of these pairs, 2,647 were within-family comparisons and consisted of 1,477 putative full sibs and 1,170 putative parent-offspring pairs. The remaining 2,204,290 pairs were between-family comparisons of putative unrelated pairs. For each pair of individuals, RELPAIR tested all eight relationships discussed in this

Table 2

Probabilities for Ordered X-Linked Genotype Pairs

	$P(X_k\|I_k)$ AT	
GENOTYPE X_k	$I_k = 0$	$I_k = 1$
Male-male pairs:		
(a,a)	p_a^2	p_a
(a,b)	$p_a p_b$	0
Male-female pairs:		
(a,aa)	p_a^3	p_a^2
(a,ab)	$2p_a^2 p_b$	$p_a p_b$
(a,bb)	$p_a p_b^2$	0
(a,bc)	$2p_a p_b p_c$	0

NOTE.—For female-female pairs, probabilities are the same as autosomal probabilities (table 1). X_k denotes the genotypes for the relative pair at an X-linked marker locus k. a, b, and c are distinct alleles at that locus with population frequencies p_a, p_b, and p_c.

Table 3

Autosomal Transition Probabilities

RELATIONSHIP AND I_k	$P(I_{k+1}\|I_k)$ FOR		
	$I_{k+1} = 0$	$I_{k+1} = 1$	$I_{k+1} = 2$
MZ twins:			
0	0	0	0
1	0	0	0
2	0	0	1
Full sibs:			
0	ψ^2	$2\psi(1-\psi)$	$(1-\psi)^2$
1	$\psi(1-\psi)$	$\psi^2 + (1-\psi)^2$	$\psi(1-\psi)$
2	$(1-\psi)^2$	$2\psi(1-\psi)$	ψ^2
Parent-offspring:			
0	0	0	
1	0	1	
Grandparent-grandchild:			
0	$1-\theta$	θ	
1	θ	$1-\theta$	
Half sibs:			
0	ψ	$1-\psi$	
1	$1-\psi$	ψ	
Avuncular:			
0	$\frac{\theta}{2} + \psi(1-\theta)$	$1 - \frac{\theta}{2} - \psi(1-\theta)$	
1	$1 - \frac{\theta}{2} - \psi(1-\theta)$	$\frac{\theta}{2} + \psi(1-\theta)$	
First cousins:			
0	$\frac{1}{3}\left[2 + \frac{\theta^2}{2} + \psi(1-\theta)^2\right]$	$\frac{1}{3}\left[1 - \frac{\theta^2}{2} - \psi(1-\theta)^2\right]$	
1	$1 - \frac{\theta^2}{2} - \psi(1-\theta)^2$	$\frac{\theta^2}{2} + \psi(1-\theta)^2$	
Unrelated:			
0	1	0	
1	0	0	

paper and inferred the relationship that maximized the multipoint probability of the marker data. We allowed for genotyping error by assuming a random-genotyping rate of .01 in all analyses.

Results

Autosomal Data

Table 7 shows the relationship classification rates for autosomal markers with four equally frequent alleles spaced at 10-cM intervals, with the assumption of no genotyping error. The standard errors for these estimated rates are

$$\sqrt{\frac{p(1-p)}{100,000}},$$

where p is the estimated classification rate and 100,000 is the number of replicates. Results are presented for 200 markers (which is referred to as a "half-genome scan") and for 399 markers (a "full-genome scan") using the chromosome-length estimates of Morton (1991).

Relationship-misclassification rates decreased with increasing number of markers or increasing intermarker distance (given a fixed number of markers) (data not shown). Even for a half-genome scan, the estimated misclassification rates for MZ-twin, full-sib, and parent-offspring pairs are only .0000, .0020, and .0000, respectively. Our multipoint method also yields reasonably accurate classification rates of first cousins and unrelated pairs. Using data from a half-genome scan yields misclassification rates of .1507 for first cousins and .0657 for unrelated pairs, whereas a full-genome scan reduces these rates to .0432 and .0157, respectively.

Our method has more difficulty distinguishing among the three tested second-degree (2°) relationships: grandparent-grandchild, half-sib, and avuncular. When a full-genome scan is used, the misclassification rates for grandparent-grandchild, half-sib, and avuncular pairs are .2788, .6282, and .3777, respectively. Although our method has poor ability to distinguish between these three 2° relationships, it has excellent ability to correctly classify grandparent-grandchild, half-sib, and avuncular relationships as 2° relationships. For a full-genome scan, classification rates of grandparent-grandchild, half-sib, and avuncular relationships as 2° relationships are .9867, .9731, and .9684, respectively (underlined region of table 7).

When misclassification of 2° relationships occurs, grandparent-grandchild and avuncular pairs are most

Table 4

X-Linked Transition Probabilities for Female-Female Pairs

Relationship and I_k	$P(I_{k+1}\mid I_k)$ for		
	$I_{k+1}=0$	$I_{k+1}=1$	$I_{k+1}=2$
Sisters:			
0	0	0	0
1	0	ψ	$1-\psi$
2	0	$1-\psi$	ψ
Paternal grandmother-granddaughter and paternal half sisters:			
0	0	0	
1	0	1	
Paternal aunt-niece and first cousins connected through brothers:			
0	ψ	$1-\psi$	
1	$1-\psi$	ψ	
Maternal aunt-niece:			
0	$\psi(1-\theta)$	$1-\psi(1-\theta)$	
1	$\frac{1}{3}[1-\psi(1-\theta)]$	$1-\frac{1}{3}[1-\psi(1-\theta)]$	
First cousins connected through sisters:			
0	$\frac{1}{5}\left[2+2\psi+\psi(1-\theta)^2\right]$	$\frac{1}{5}\left[3-2\psi-\psi(1-\theta)^2\right]$	
1	$\frac{1}{3}\left[3-2\psi-\psi(1-\theta)^2\right]$	$\frac{1}{3}\left[2\psi+\psi(1-\theta)^2\right]$	
First cousins connected through a brother and sister:			
0	$1-\frac{1}{3}[1-\psi(1-\theta)]$	$\frac{1}{3}[1-\psi(1-\theta)]$	
1	$1-\psi(1-\theta)$	$\psi(1-\theta)$	

NOTE.—Probabilities for MZ twins, mother-daughter pairs, maternal half sisters, maternal grandmother-granddaughter pairs, and unrelated pairs are the same as for the autosomal case (table 3).

often incorrectly classified as half sibs, and half sibs are usually incorrectly classified as avuncular pairs. The inability to distinguish these three 2° relationships can be traced to their similarities in IBD sharing: all three pairs share, on average, 1/4 of their autosomal genome IBD. The transition probabilities are the only components of the autosomal multipoint likelihood that vary among these relationships. Figure 1 shows the IBD sharing transition probabilities between two markers for the three relationships as a function of the recombination fraction θ. The transition probabilities for half sibs and avuncular pairs have similar values across all values of θ, with half sibs intermediate between avuncular and grandparent-grandchild. These observations about the transition probabilities help explain the 2° relative misclassification rates. We are also using only an approximation of the likelihood for the avuncular relationship, which likely results in a modest increase in the misclassification rates for our tested 2° relationships.

Effect of Random-Genotyping Error

Table 8 shows relationship-classification-rate estimates when marker data are simulated with a random-genotype-error rate of .01. As expected, the failure to incorporate genotype error in the model when it exists in the data leads to erroneous classification of nearly all MZ twins as full sibs and nearly all parent-offspring pairs as grandparent-grandchild pairs. However, allowing for a .01 random-genotyping rate in our model re-

sults in correct classification of all MZ twin and parent-offspring pairs.

For the 2° relationships, the introduction of a .01 random-genotyping rate and the failure to model it leads to increased misclassification rates for grandparent-grandchild pairs and half sibs but slightly decreased misclassification rates for avuncular pairs (tables 7 and 8). Genotyping error leads to perception of more-frequent changes in IBD sharing for the relative pair along the genome. Since we expect avuncular pairs to have more shifts in IBD sharing than half sibs and half sibs to have more changes in sharing than grandparent-grandchild pairs, genotype error favors the avuncular relationship over the half-sib relationship and the half-sib relationship over the grandparent-grandchild relationship. Incorporation of a .01 genotyping-error rate in our model restores the classification rates of all 2° relationships essentially to the level seen when marker data were simulated with no genotyping error (tables 7 and 8).

The classification-rate estimates for full sibs, first cousins, and unrelated pairs remain essentially unchanged when random-genotype error is introduced in a full-genome scan. This is not surprising, since these pairs have inheritance patterns distinct from the other pairs we considered and, unlike MZ twins and parent-offspring pairs, these pairs need not share alleles IBD. Therefore, the introduction of genotype error does not override the information supporting the true relationship.

Table 8 also shows the effect of underestimating and

Table 5

X-Linked Transition Probabilities for Male-Male Pairs

	$P(I_{k+1}\|I_k)$ FOR	
RELATIONSHIP AND I_k	$I_{k+1} = 0$	$I_{k+1} = 1$
MZ twins:		
0	0	0
1	0	1
Brothers:		
0	ψ	$1 - \psi$
1	$1 - \psi$	ψ
Father-son, paternal grandfather-grandson, paternal half brothers, paternal uncle-nephew, first cousins connected through brothers, and first cousins connected through a brother and sister:		
0	1	0
1	0	0
Maternal uncle-nephew:		
0	$1 - \frac{1}{3}[1 - \psi(1-\theta)]$	$\frac{1}{3}[1 - \psi(1-\theta)]$
1	$1 - \psi(1-\theta)$	$\psi(1-\theta)$
First cousins connected through sisters:		
0	$\frac{1}{5}\left[2 + 2\psi + \psi(1-\theta)^2\right]$	$\frac{1}{5}\left[3 - 2\psi - \psi(1-\theta)^2\right]$
1	$\frac{1}{3}\left[3 - 2\psi - \psi(1-\theta)^2\right]$	$\frac{1}{3}\left[2\psi + \psi(1-\theta)^2\right]$

NOTE.—Probabilities for maternal half brothers, maternal grandfather-grandson pairs, and unrelated pairs are the same as for the autosomal case (table 3).

overestimating the true genotyping-error rate. Assumption of an error rate of .001 when the true rate is .01 results in misclassification-rate estimates of .0000 for MZ twins and only .0007 for parent-offspring pairs. The misclassification rates for all other tested relationships were similar to the case where we assumed no genotyping error in our model. Assumption of an error rate of .02 when the true rate is .01 yields similar classification results, for most relationships, compared with those yielded under the assumption of the true genotyping-error rate, .01. Only the tested 2° relationships appear to be affected by the overestimation of the true genotyping-error rate. The misclassification-rate estimate for grandparent-grandchild pairs and half sibs decreases, whereas the misclassification-rate estimate increases for avuncular pairs. When we overestimate the genotyping-error rate, the model adjusts for more perceived shifts in IBD sharing for the relative pair along the genome than actually are expected under the true genotyping-error rate. Since we expect grandparent-grandchild pairs to have fewer shifts in IBD sharing than half sibs and also expect half sibs to have fewer changes in sharing than avuncular pairs, overestimating the true genotyping-error rate favors the grandparent-grandchild relationship over the half-sib relationship and the half-sib relationship over the avuncular relationship.

Autosomal and X-linked Data

Table 9 shows classification rates for selected female-female, male-male, and male-female pairs, using a full 10-cM autosomal genome scan together with 23 additional X-linked markers spaced at 10-cM intervals. We limit our attention both in the table and in the text to those relationships most affected by the inclusion of X-linked data.

Female-Female Pairs

X-linked data significantly improve our ability to infer paternal half sisters. The misclassification rate decreases from .6282 (table 7) for an autosomal genome scan to .1546 (table 9) when X-linked data are included. Those that were misclassified previously as avuncular now are classified correctly, since paternal half sisters must share one allele IBD across the entire X genome. Thus, they must share half of their X-linked genome IBD. This is in contrast to maternal and paternal aunt-niece relationships, which are expected to share 3/8 and 1/4 of their X-linked genome IBD, respectively. Paternal half-sister pairs that were misclassified previously as grandparent-grandchild remain misclassified as that relationship when X-linked data are included. X-linked data will not help in this case, since paternal grandmother-granddaughter pairs must also share one allele IBD across the entire X genome.

This advantage in the classification of paternal half sisters comes at the price of a modest decrease in our ability to classify maternal aunt-niece pairs correctly. This misclassification rate increases from .3777 to .4575 as more pairs are misclassified as half sibs. Because of the limited number of X-linked markers, some maternal aunt-niece pairs may share one allele IBD at every marker. This leads to misclassification of the pair as pa-

Table 6

X-Linked Transition Probabilities for Different Relationships for Male-Female Pairs

	$P(I_{k+1} | I_k)$ FOR	
RELATIONSHIP AND I_k	$I_{k+1} = 0$	$I_{k+1} = 1$
Brother-sister, paternal uncle-niece:		
0	ψ	$1 - \psi$
1	$1 - \psi$	ψ
Paternal half brother-sister, paternal grandfather-granddaughter, paternal grandmother-grandson, paternal aunt-nephew, first cousins connected through brothers, and first cousins connected through a brother and sister where male cousin is brother's offspring and female cousin is sister's offspring:		
0	1	0
1	0	0
Maternal aunt-nephew:		
0	$\psi(1 - \theta)$	$1 - \psi(1 - \theta)$
1	$\frac{1}{3}[1 - \psi(1 - \theta)]$	$1 - \frac{1}{3}[1 - \psi(1 - \theta)]$
Maternal uncle-niece and first cousins connected through a brother and sister where male cousin is sister's offspring and female cousin is brother's offspring:		
0	$1 - \frac{1}{3}[1 - \psi(1 - \theta)]$	$\frac{1}{3}[1 - \psi(1 - \theta)]$
1	$1 - \psi(1 - \theta)$	$\psi(1 - \theta)$
First cousins connected through sisters:		
0	$\frac{1}{5}[2 + 2\psi + \psi(1 - \theta)^2]$	$\frac{1}{5}[3 - 2\psi - \psi(1 - \theta)^2]$
1	$\frac{1}{3}[3 - 2\psi - \psi(1 - \theta)^2]$	$\frac{1}{3}[2\psi + \psi(1 - \theta)^2]$

NOTE.—Probabilities for father-daughter, mother-son, maternal grandmother-grandson, maternal grandfather-granddaughter, maternal half brother-sister, and unrelated relationships are the same as for the autosomal case (table 3).

ternal half sisters. We can remedy this situation by typing more X-linked markers.

Male-Male Pairs

X-linked data substantially improve our ability to distinguish maternal half brothers. The misclassification rate decreases from .6282 (table 7) to .3890 (table 9) as pairs previously misclassified as avuncular now are classified correctly. This is because maternal half brothers have IBD sharing trends distinct from those seen in avuncular relationships. We expect maternal half brothers to share 1/2 of their X chromosome IBD, whereas maternal and paternal uncle-nephew pairs expect to share 1/4 and 0, respectively.

Male-Female Pairs

X-linked data decrease the misclassification rates of the different avuncular relationships but increase misclassification rates for grandparent-grandchild and half-sib pairs. Many pairs previously classified as grandparent-grandchild or half sib now are identified as avuncular. The main reason is that, for maternal uncle-niece and maternal aunt-nephew pairs, the female is expected to share 1/4 and 3/4 of the male X chromosome IBD, respectively. The females in the other 2° male-female relationships expect to share either zero or 1/2 of the male X-chromosome IBD. The inheritance patterns

of both maternal uncle-niece pairs and maternal aunt-nephew pairs are distinct enough from the other 2° male-female relationships that we can classify them accurately. However, random increases or decreases in allele sharing along the X chromosome will lead to misclassification of many of these other 2° male-female relationships as one of these two particular avuncular relationships.

Effect of Random X-linked Genotyping Error

Results when X-linked marker data are simulated with a random-genotyping rate of .01 reveal trends similar to those seen for autosomal data (data not shown). Unaccounted genotyping errors result both in misclassification of MZ twins and parent-offspring pairs and in increased classification of many 2°-relative pairs as avuncular. Accounting for error restores the classification rates to very near the levels seen when data were simulated without error.

Analysis of FUSION Data Set

For the within-family comparisons, RELPAIR identified 3 of 1,477 putative full sibs as MZ twins (or sample duplications), 20 as 2° relatives, 1 as first cousins, and 6 as unrelated. RELPAIR also classified 8 of the 1,170 putative parent-offspring pairs as unrelated. The three MZ-twin pairs are most likely true MZ twins and not duplications, since they reported the same birth dates.

Table 7

Classification-Rate Estimates for 200 or 399 Autosomal Markers with Four Equally Frequent Alleles Spaced at 10-cM Intervals and No Genotype Error

	CLASSIFICATION-RATE ESTIMATE FOR INFERRED RELATIONSHIP R^*							
R	MZ	FS	PO	GG	HS	AV	FC	UN
200:								
MZ	**1.0000**	.0000	.0000	.0000	.0000	.0000	.0000	.0000
FS	.0000	**.9980**	.0000	.0007	.0007	.0006	.0000	.0000
PO	.0000	.0000	**1.0000**	.0000	.0000	.0000	.0000	.0000
GG	.0000	.0008	.0000	<u>.6536</u>	<u>.2047</u>	<u>.0832</u>	.0576	.0002
HS	.0000	.0005	.0000	<u>.2184</u>	<u>.2774</u>	<u>.4166</u>	.0869	.0001
AV	.0000	.0005	.0000	<u>.1390</u>	<u>.2342</u>	<u>.5321</u>	.0942	.0000
FC	.0000	.0000	.0000	.0217	.0158	.0456	**.8493**	.0677
UN	.0000	.0000	.0000	.0000	.0000	.0000	.0657	**.9343**
399:								
MZ	**1.0000**	.0000	.0000	.0000	.0000	.0000	.0000	.0000
FS	.0000	**1.0000**	.0000	.0000	.0000	.0000	.0000	.0000
PO	.0000	.0000	**1.0000**	.0000	.0000	.0000	.0000	.0000
GG	.0000	.0000	.0000	<u>.7212</u>	<u>.2277</u>	<u>.0378</u>	.0133	.0000
HS	.0000	.0000	.0000	<u>.1445</u>	<u>.3718</u>	<u>.4568</u>	.0269	.0000
AV	.0000	.0000	.0000	<u>.0692</u>	<u>.2769</u>	<u>.6223</u>	.0316	.0000
FC	.0000	.0000	.0000	.0044	.0060	.0144	**.9568**	.0184
UN	.0000	.0000	.0000	.0000	.0000	.0000	.0157	**.9843**

NOTE.—Classification rates of 2° relatives are underlined. Entries in boldface are the probabilities of correct classification.

The 20 2° relative pairs are due to the presence of half sibs in 14 pedigrees. The 14 unrelated pairs are due to two confirmed genotype sample switches, one confirmed genotype reassignment, and one case of false paternity. Further investigation of other family members suggests the first-cousin pair is most likely an unrelated pair that is misclassified by RELPAIR.

For the between-family comparisons, RELPAIR identified 1 of the 2,204,290 putative unrelated pairs as MZ twins, 5 as full sibs, 17 as 2° relatives, 8 as parent-offspring pairs, and 5,330 as first cousins. The MZ-twin pair has been confirmed as a sample duplication. Two of the three genotype sample switches found in the within-family comparisons explain all of the full-sib and some of the 2°-relative and parent-offspring pairs. The other 2°-relative and parent-offspring pairs are the result of two pairs of related families in the data set. Because of the increased chance of error when we analyze a large number of pairs, we suspect that the majority of the putative pairs identified by RELPAIR as first cousins are unrelated.

Discussion

Overview

We have extended the method of Boehnke and Cox (1997) to test all possible pairs of individuals, to test additional relationships, to allow for random-genotyping error, and to include X-linked data. Assuming a half (200 markers) or full (399 markers) 10-cM autosomal

genome scan, our method accurately classifies monozygotic twins, full sibs, parent-offspring pairs, 2° relatives, first cousins, and unrelated pairs. Our method is also computationally efficient. When a SUN Enterprise 450 workstation is used, the classification of 100,000 relative pairs requires only 2 min of computation time.

The primary limitation of our method is its inability to distinguish accurately among the 2° relationships, particularly if only autosomal data are used. Ages of the individuals within a putative 2° relationship and the results for other pairs of relatives within the same family may assist in verification of the true relationship of the relative pair. We are currently working on statistical methods for improving 2°-relationship classification rates that utilize the marker data of additional relatives within the family. However, even if there are no additional relatives to analyze, the inclusion of X-linked data will help, as it improves classification accuracy for certain sex combinations of 2° relationships caused by differences in expected X-linked IBD sharing.

Our method also accommodates genotyping error effectively, to judge by the near-complete restoration of classification accuracy for all relationships considered when genotyping error is allowed for in the analysis. Since the true underlying genotyping-error rate will be unknown beforehand, some consideration is required in choosing an assumed rate. To avoid misclassification of MZ twins and parent-offspring pairs, we suggest assumption of a positive error rate. The assumed error rate could reflect the empirical error rate produced by
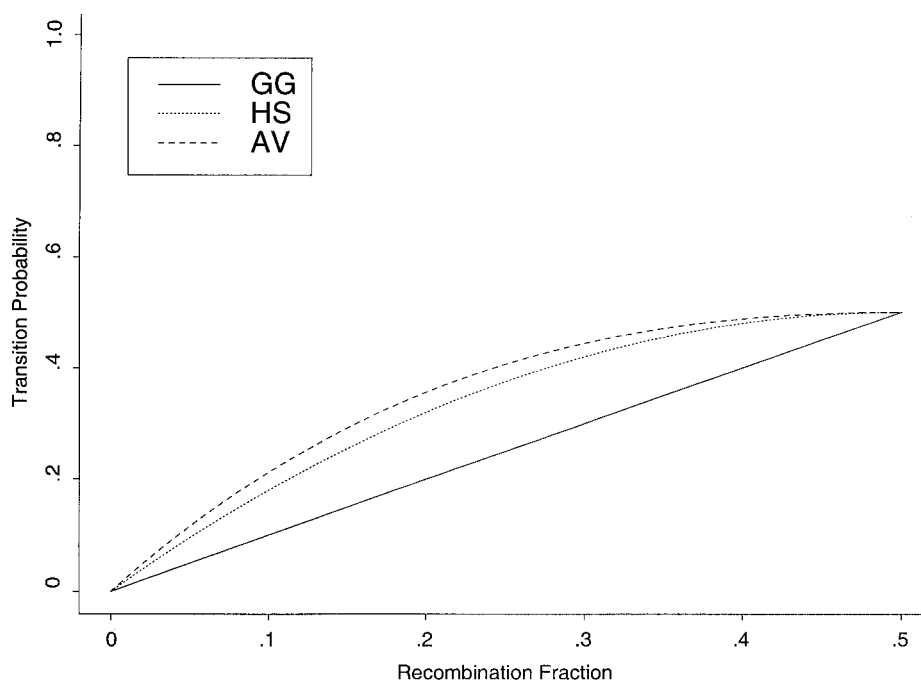
Figure 1 Autosomal transition probabilities for grandparent-grandchild (GG), half-sib (HS), and avuncular (AV) pairs. $P(I_{k+1} = 1 | I_k = 0) = P(I_{k+1} = 0 | I_k = 1)$ is shown. Note that $P(I_{k+1} = 0 | I_k = 0) = 1 - P(I_{k+1} = 1 | I_k = 0)$ and $P(I_{k+1} = 1 | I_k = 1) = 1 - P(I_{k+1} = 0 | I_k = 1)$.

one's genotyping facility. However, as our results have shown, our method is robust to sensible under- or overestimation of the true random-genotyping rate, so long as the assumed rate is not zero.

The random-genotype model we have used to allow for genotype error certainly is not realistic. Scoring heterozygotes as homozygotes, scoring homozygotes as heterozygotes, or displacing both alleles of a genotype are common errors in actual data. However, the random-genotyping error has the virtue of computational simplicity, and previous work suggests that it works very well at detecting errors generated by these and other, more realistic error mechanisms (Douglas et al. 2000).

Throughout this paper, we have assumed that estimated intermarker recombination fractions are always correct and are the same for both males and females. Violations of these assumptions might be expected to lead to higher misclassification rates for many of the tested relationships. To determine the effect of map uncertainty, we simulated 399 autosomal markers and placed them at alternating 8- and 12-cM distances along the genome. We then analyzed the data assuming a constant intermarker distance of 10 cM. Results (not shown) revealed that our method is quite robust to recombination-fraction misspecification for all tested relation-

ships, since no classification rate decreased by more than ~1%.

The chromosome-map lengths used in these analyses come from Morton (1991). To investigate the impact of assumed map length, we repeated some simulations, using chromosome lengths from Broman et al. (1998). Assuming a 10-cM intermarker distance, we placed a total of 359 autosomal and 19 X-linked markers along the genome, using the autosomal sex-averaged maps and the female X-linked map of Broman et al. (1998). We performed analyses using these 359 autosomal markers as a full-genome scan. Compared with the full-genome scan using the maps of Morton (1991) that used 399 autosomal markers, the only classification rates affected were those of the 2° relationships. The misclassification rates increased from .2788 to .3155 for grandparent-grandchild pairs, from .6282 to .6484 for half sibs, and from .3777 to .3856 for avuncular pairs. Results using X-linked data showed similar trends (data not shown).

In principle, our method easily can be extended to test other relationships, such as second cousins or great-grandparent-greatgrandchild; one need only derive the IBD initial conditions and transition probabilities for these relationships. In practice, the ability to classify these relationships accurately will depend on their sim-

Table 8

Classification-Rate Estimates for 399 Autosomal Markers with Four Equally Frequent Alleles Spaced at 10-cM Intervals and a True Random-Genotype Rate of .01

ASSUMED ERROR RATE AND R	CLASSIFICATION-RATE ESTIMATES FOR INFERRED RELATIONSHIP R^*							
	MZ	FS	PO	GG	HS	AV	FC	UN
0:								
MZ	**.0009**	.9991	.0000	.0000	.0000	.0000	.0000	.0000
FS	.0000	**1.0000**	.0000	.0000	.0000	.0000	.0000	.0000
PO	.0000	.0023	.0739	.9238	.0000	.0000	.0000	.0000
GG	.0000	.0000	.0000	**.6342**	.2805	.0640	.0214	.0000
HS	.0000	.0000	.0000	.1153	**.3362**	.5073	.0412	.0000
AV	.0000	.0000	.0000	.0570	.2426	**.6553**	.0450	.0000
FC	.0000	.0000	.0000	.0034	.0044	.0107	**.9600**	.0215
UN	.0000	.0000	.0000	.0000	.0000	.0000	.0162	**.9838**
.001:								
MZ	**1.0000**	.0000	.0000	.0000	.0000	.0000	.0000	.0000
FS	.0000	**1.0000**	.0000	.0000	.0000	.0000	.0000	.0000
PO	.0000	.0000	.9993	.0007	.0000	.0000	.0000	.0000
GG	.0000	.0000	.0000	**.6338**	.2621	.0774	.0267	.0000
HS	.0000	.0000	.0000	.1206	**.3384**	.5031	.0380	.0000
AV	.0000	.0000	.0000	.0597	.2436	**.6527**	.0440	.0000
FC	.0000	.0000	.0000	.0034	.0045	.0116	**.9598**	.0207
UN	.0000	.0000	.0000	.0000	.0000	.0000	.0161	**.9839**
.01:								
MZ	**1.0000**	.0000	.0000	.0000	.0000	.0000	.0000	.0000
FS	.0000	**1.0000**	.0000	.0000	.0000	.0000	.0000	.0000
PO	.0000	.0000	**1.0000**	.0000	.0000	.0000	.0000	.0000
GG	.0000	.0000	.0000	**.7082**	.2271	.0484	.0163	.0000
HS	.0000	.0000	.0000	.1586	**.3544**	.4573	.0296	.0000
AV	.0000	.0000	.0000	.0820	.2689	**.6151**	.0340	.0000
FC	.0000	.0000	.0000	.0055	.0057	.0158	**.9530**	.0200
UN	.0000	.0000	.0000	.0000	.0000	.0000	.0188	**.9812**
.02:								
MZ	**1.0000**	.0000	.0000	.0000	.0000	.0000	.0000	.0000
FS	.0000	**1.0000**	.0000	.0000	.0000	.0000	.0000	.0000
PO	.0000	.0000	**1.0000**	.0000	.0000	.0000	.0000	.0000
GG	.0000	.0000	.0000	**.7675**	.1856	.0356	.0113	.0000
HS	.0000	.0000	.0000	.2066	**.3616**	.4103	.0214	.0000
AV	.0000	.0000	.0000	.1108	.2933	**.5718**	.0240	.0000
FC	.0000	.0000	.0000	.0081	.0083	.0202	**.9464**	.0170
UN	.0000	.0000	.0000	.0000	.0000	.0000	.0198	**.9802**

NOTE.—Classification rates of 2° relatives are underlined. Entries in boldface are the probabilities of correct classification.

ilarities in IBD sharing to those of other tested relationships. For example, since greatgrandparent-greatgrandchild and first-cousin relationships are 3° relationships and have similar IBD sharing across the autosomal genome, our method will have difficulty distinguishing between them. Also, as the relationships tested become more distant, our method will have trouble distinguishing these pairs from unrelated pairs. The inclusion of X-linked data may help to infer some of these relationships for specific sex combinations.

Marker-Allele Frequencies and Map Density

In the Results section, we focused on markers with heterozygosity .75 (four equally frequent alleles), which is typical of microsatellite markers frequently used in gene-mapping studies. We also performed analyses assuming markers with heterozygosity .50 (two equally frequent alleles), which is representative of the most highly informative single-nucleotide polymorphisms (SNPs). For MZ-twin and parent-offspring pairs, we found that a full 10-cM autosomal genome scan with these biallelic markers yields misclassification rates of .0000 for both relationships (results not shown). More biallelic markers are required to achieve a given classification rate for the more distant tested relationships. Compared with a 10-cM autosomal genome scan with four equally frequent alleles, a biallelic marker genome scan at 4-cM density attains the same misclassification rate (.0000) for full sibs, a 3-cM biallelic genome scan attains approximately the same rates for tested 2° re-

Table 9

Classification-Rate Estimates for Selected 2° Relationships for 399 Autosomal and 23 X-linked Markers with Four Equally Frequent Alleles Spaced at 10-cM Intervals

	CLASSIFICATION-RATE ESTIMATES FOR INFERRED RELATIONSHIP R^*			
R	GG	HS	AV	FC
Female-female pairs:				
Paternal half sisters	.1456	**.8454**	.0047	.0044
Maternal half sisters	.1299	**.3434**	.5004	.0263
Paternal aunt-niece	.0612	.2652	**.6442**	.0294
Maternal aunt-niece	.0649	.3728	**.5425**	.0198
Male-male pairs:				
Paternal half brothers	.1444	**.3432**	.4881	.0243
Maternal half brothers	.1226	**.6110**	.2436	.0228
Paternal uncle-nephew	.0692	.2549	**.6485**	.0274
Maternal uncle-nephew	.0624	.3067	**.6038**	.0271
Male-female pairs:				
Paternal grandfather-granddaughter	**.7089**	.2089	.0665	.0157
Maternal grandfather-granddaughter	**.6921**	.1516	.1439	.0124
Paternal grandmother-grandson	**.7103**	.2072	.0666	.0159
Maternal grandmother-grandson	**.6921**	.1491	.1459	.0129
Paternal half brother-sister	.1427	**.3414**	.4915	.0244
Maternal half brother-sister	.1179	**.2709**	.5916	.0196
Paternal uncle-niece	.0560	.1895	**.7318**	.0227
Maternal uncle-niece	.0633	.2053	**.7048**	.0266
Paternal aunt-nephew	.0707	.2496	**.6532**	.0265
Maternal aunt-nephew	.0584	.0947	**.8342**	.0127

NOTE.—No 2° relative pair was misclassified as MZ twins, full sibs, parent-offspring, or unrelated. Entries in boldface are the probabilities of correct classification.

lationships, and a 4-cM density yields similar rates for first cousins and unrelated pairs (results not shown).

The optimal data for our method would be an infinitely dense map of fully informative markers. To approximate this situation, we placed markers with 10 equally frequent alleles (heterozygosity .90) at .1-cM intervals across the autosomal genome. As expected, the misclassification rates of MZ twins, full sibs, and parent-offspring pairs remained zero. A zero misclassification rate was also obtained for unrelated pairs, whereas grandparent-grandchild pairs and first cousins had small misclassification rates: .0218 and .0020, respectively. The misclassification rates for half sibs and avuncular pairs also significantly decreased (.2547 and .1005, respectively). Likewise, the classification rates of true grandparent-grandchild, half-sib, and avuncular relationships as 2° relationships increased to .9992, .9915, and .9955, respectively.

Likelihood-Based Methods of Detecting Misspecified Relationships

A variety of methods have been suggested for pairwise relationship estimation. Göring and Ott (1997) used a Bayesian method to identify full sibs, half sibs, and unrelated pairs in the context of affected-sib-pair analysis. They assume prior probabilities for the three relationships in the study population and calculate the posterior probability of a particular relationship, given the ge-

notype data. They used these posterior probabilities to infer the relationship. They also allowed for genotyping of a parent for the testing of a putative sib pair.

As described earlier, McPeek and Sun (2000) use a likelihood-ratio statistic to test the null hypothesis that the putative relationship is correctly specified, with the alternative hypothesis being that the relationship is not correctly specified. Under the alternative, the likelihood is maximized as a function of the probability from the set of other relationships. They determine significance by simulation, since their likelihood-ratio test statistic has a strongly skewed distribution. Simulations indicate that their method yields similar power to distinguish relationships to that of our method. Both Göring and Ott (1997) and McPeek and Sun (2000) calculate their multipoint likelihoods using the same general method as ours. However, they restrict their calculations to autosomal data with no genotyping error.

Allele-Sharing Methods to Detect Misspecified Relationships

For putative full sibs, Ehm and Wagner (1998) proposed a test statistic based on the total number of alleles shared identical by state (IBS) by the pair at a collection of autosomal markers. The authors calculate the mean and variance of this statistic assuming full sibs and use a normal approximation to test for departures from this relationship. Stivers et al. (1996) derived a similar sta-

tistic. McPeek and Sun (2000) also constructed two allele-sharing statistics. The first extends Ehm and Wagner's (1998) IBS test to any relative pair. The second calculates the expected number of alleles shared IBD by the pair at a series of markers. They use a normal approximation to determine whether this statistic deviates from the expected number of alleles shared IBD under the putative relationship. Although computationally simple, these allele-sharing statistics generally have lower power than do multipoint likelihood-based methods (Boehnke and Cox 1997; Ehm and Wagner 1998; McPeek and Sun 2000). These methods also restrict their calculations to autosomal data with no genotyping error (although they are quite robust when error exists), and they fail to infer the actual relationship of a pair if the putative relationship is rejected.

For putative full sibs, Olson (1999) derived an IBD allele-sharing method, based on autosomal data with no genotyping error, that does suggest an alternative relationship when the putative one is rejected. The method is similar to the IBD procedure proposed by McPeek and Sun (2000), but IBD estimates now are calculated at any location along the genome, using existing multipoint methods (Kruglyak et al. 1996; Hauser and Boehnke 1998). The procedure requires calculating critical values for relationship inference that are functions of the genome length and the average marker information content.

Continuous Gamete IBD Methods

Gametes of two related individuals will have regions of IBD sharing and nonsharing along the genome; the lengths and patterns of these regions can distinguish different relationships. Browning (1998) used Monte Carlo procedures to estimate the likelihood of a particular relationship of a pair, given their gamete IBD data, and constructed a likelihood-ratio statistic to test the pair's putative relationship against an alternative one. The method requires observation of IBD status along the chromosomes and is computationally intensive. In similar work, Zhao and Liang (in press) derived a method for exact calculation of the likelihood of a given relationship, given gamete IBD data. This method is computationally efficient, compared with Monte Carlo procedures, and yields results similar to the method of Browning (1998).

Concluding Remarks

We have derived a method for relationship inference that is accurate, computationally fast, and flexible enough to accommodate different genetic phenomena, such as genotyping error and X-linked data. We have implemented our method and extensions in the FOR-TRAN 77 program RELPAIR, version 2.0. The program is freely available on the World Wide Web.

Electronic-Database Information

Accession numbers and URLs for data in this article are as follows:

University of Michigan Center for Statistical Genetics Web page, http://www.sph.umich.edu/statgen/software (for free RELPAIR program)

References

Baum LE (1972) An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes. Inequalities 3:1–8

Boehnke M, Cox NJ (1997) Accurate inference of relationships in sib-pair linkage studies. Am J Hum Genet 61:423–429

Broman KW, Murray JC, Sheffield VC, White RL, Weber JL (1998) Comprehensive human genetic maps: individual and sex-specific variation in recombination. Am J Hum Genet 63:861–869

Broman KW, Weber JL (1998) Estimation of pairwise relationships in the presence of genotyping errors. Am J Hum Genet 63:1563–1564

Browning S (1998) Relationship identification contained in gamete identity by descent data. J Comput Biol 5:323–334

Douglas JA, Boehnke M, Lange K (2000) A multipoint method for detecting genotyping errors and mutations in sibling-pair linkage data. Am J Hum Genet 66:1287–1297

Duren WL, Cox NJ, Hauser ER, Boehnke M, FUSION Study Group (1997) Software for determining most likely relationships in relative pairs. Am J Hum Genet Suppl 57:A273

Ehm MG, Wagner M (1998) A test statistic to detect errors in sib-pair relationships. Am J Hum Genet 62:181–188

Feingold E (1993) Markov processes for modeling and analyzing a new genetic mapping method. J Appl Prob 30:766–779

Göring HHH, Ott J (1997) Relationship estimation in affected sib pair analysis of late-onset diseases. Eur J Hum Genet 5:69–77

Hauser B, Boehnke M (1998) Genetic linkage analysis of complex genetic traits by using affected sibling pairs. Biometrics 54:1238–1246

Kosambi DD (1944) The estimation of map distances from recombination values. Ann Eugenics 12:172–175

Kruglyak L, Daly M, Reeve-Daly M, Lander E (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet 58:1347–1363

Matise TC, Perlin M, Chakravarti A (1994) Automated construction of genetic linkage maps using an expert system

(MultiMap): a human genome linkage map. Nat Genet 6: 384–390

McPeek MS, Sun L (2000) Statistical tests for detection of misspecified relationships by use of genome-screen data. Am J Hum Genet 66:1076–1094

Morton NE (1991) Parameters of the human genome. Proc Natl Acad Sci USA 88:7474–7476

Olson JM (1999) Relationship estimation by Markov-process models in a sib-pair linkage study. Am J Hum Genet 64: 1464–1472

Risch N (1990) Linkage strategies for genetically complex traits. II. The power of affected relative pairs. Am J Hum Genet 46:229–241

Stivers DN, Zhong Y, Hanis CL, Chakraborty R (1996) REL-TYPE: a computer program for determining biological relatedness between individuals based on allele sharing at microsatellite loci. Am J Hum Genet Suppl 59:A190

Thompson EA (1975) The estimation of pairwise relationships. Ann Hum Genet 39:173–188

Valle T, Tuomilehto J, Bergman RN, Ghosh S, Hauser ER, Eriksson J, Nylund SJ, et al (1998) Mapping genes for NIDDM: design of the Finland-United States Investigation of NIDDM Genetics (FUSION) study. Diabetes Care 21: 949–958

Zhao H, Liang F. On relationship inference using gamete identity by descent data. J Comput Biol (in press)