Genotypic probabilities for pairs of inbred relatives

Wenlei Liu[†] and B. S. Weir*

*Program in Statistical Genetics, Department of Statistics, North Carolina State University,
Raleigh, NC 27695-7566, USA*

Expressions for the joint genotypic probabilities of two related individuals are used in many population and quantitative genetic analyses. These expressions, resting on a set of 15 probabilities of patterns of identity by descent among the four alleles at a locus carried by the relatives, are generally well known. There has been recent interest in special cases where the two individuals are both related and inbred, although there have been differences among published results. Here, we return to the original 15-probability treatment and show appropriate reductions for relatives when they are drawn from a population that itself is inbred or when the relatives have parents who are related. These results have application in affected-relative tests for linkage, and in methods for interpreting forensic genetic profiles.

Keywords: inbreeding; relatedness; affected-relative tests; forensic match probabilities

1. INTRODUCTION

Many developments in population and quantitative genetics require knowledge of the probability that two individuals have specified genotypes. This article has been motivated in part by work that tests for linkage between disease susceptibility and marker loci on the basis of the family pedigrees (e.g. Merette and Ott 1996) or of the marker genotypes of affected relatives (e.g. Liu & Weir 2004), and partly by the need to generate matching probabilities when genetic profiles are used in forensic science (e.g. Weir 1994). There has been parallel activity in quantitative genetics concerning the covariances of relatives, especially when they are inbred (e.g. Wright 1987). This article considers the joint genotype probabilities for pairs of individuals, expressed in terms of various alternative formulations of the probabilities of identity by descent (ibd) among the four alleles they carry at a locus. It does not address the interesting problems in estimating the degree of relatedness for pairs of individuals (e.g. Ritland & Travis 2004).

Affected relative tests examine the extent of marker allele sharing for these individuals and indicate linkage between marker and disease loci when this sharing is greater than expected. Génin & Clerget-Darpoux (1996) studied the effects of inbreeding on affected sib pair tests and extended existing tests to apply to sib pairs sampled from a consanguineous population. More recently, Liu & Weir (2004) made adjustments to affected sib pair tests of Lange (1986a,b) for the case when the parents of sib pairs belong to a population in a state of equilibrium between drift and mutation. They

took into account that consequent inbreeding of the sib pair elevates the degree of allele sharing.

The work of Génin & Clerget-Darpoux (1996) was based on their expressions for joint genotype probabilities for sib pairs that involved allele frequencies and the inbreeding coefficient in the parental population. They assumed that all nine condensed ibd relationships (Jacquard 1970) among the four alleles carried by two individuals could be written as functions of the inbreeding coefficient. Although this is possible in some cases, Weeks & Sinsheimer (1998) pointed out that the results of Génin & Clerget-Darpoux did not lead to known results for ibd measures. Weeks & Sinsheimer (1998) presented some affected relative tests for linkage that took account of the identity of alleles between individuals (their relationship) and within individuals (their inbreeding). Cannings (1998) also wrote that the results of Génin & Clerget-Darpoux did not correspond to ‘fully specified population model[s] (except in certain very restricted and uninteresting cases).’ Génin & Clerget-Darpoux (1998a,b) responded to the criticisms of Cannings (1998) and Weeks & Sinsheimer (1998) presented new relationships between the ibd measures and the inbreeding coefficient. However, we believe that these new expressions are still restricted, as indeed are those of Cannings (1998).

Joint genotypic probabilities are needed for the conditional probabilities of one genotype given another. In forensic science, there is a need for the probability that an unknown person (e.g. a perpetrator) has a particular genotype given that a known person (e.g. a suspect) has that genotype. Balding & Nichols (1994) derived conditional probabilities for an equilibrium population, and Weir (1994) gave more general expressions with an approximation for the pure genetic drift case.

In this article, we review known expressions for joint genotype probabilities for two individuals and discuss the special cases of relatedness with and without

* Author for correspondence (weir@stat.ncsu.edu).

[†] Present address: Department of Health Evaluation Sciences, Penn State College of Medicine, A210., Suite 2200, 600 Centerview Dr., Hershey, PA 17033-0855, USA.

One contribution of 16 to a Theme Issue ‘Population genetics, quantitative genetics and animal improvement: papers in honour of William (Bill) Hill’.

inbreeding. We treat the case of sib pairs in equilibrium populations when their parents are either unrelated or are first cousins.

2. BASIC IDENTITY MEASURES

We will retain consistency with previous notation (e.g. Weir 1994) by considering alleles a, b, c, d at an autosomal locus. These letters serve only to identify the alleles and do not specify allelic type. For locus A , the alleles will be written as A_i and the probability that an allele is of type A_i will be written as p_i . It has long been recognized that there are 15 patterns of ibd among the four alleles carried by two individuals, and these and their probabilities are shown in table 1 along with the notation of Cockerham (1971) and of Jacquard (1970). Equivalent treatments have been given by Harris (1964) and Gillios (1965).

The measures in table 1 are the most natural parameters for expressing the joint probabilities of two genotypes. If individuals X and Y carry alleles ab and cd , respectively, then adding over both orders of alleles within heterozygotes, but not over both orders of different genotypes:

$$\left. \begin{aligned} \Pr(A_i A_i, A_i A_i) &= \delta_{abcd} p_i + (\delta_{abc} + \delta_{abd} + \delta_{acd} + \delta_{bcd}) p_i^2 \\ &\quad + (\delta_{ab \cdot cd} + \delta_{ac \cdot bd} + \delta_{ad \cdot bc}) p_i^2 \\ &\quad + (\delta_{ab} + \delta_{ac} + \delta_{ad} + \delta_{bc} + \delta_{bd} + \delta_{cd}) p_i^3 \\ &\quad + \delta_0 p_i^4, \\ \Pr(A_i A_i, A_j A_j) &= \delta_{ab \cdot cd} p_i p_j + \delta_{ab} p_i p_j^2 + \delta_{cd} p_i^2 p_j + \delta_0 p_i^2 p_j^2, \\ \Pr(A_i A_i, A_i A_j) &= (\delta_{abc} + \delta_{abd}) p_i p_j \\ &\quad + (2\delta_{ab} + \delta_{ac} + \delta_{ad} + \delta_{bc} + \delta_{bd}) p_i^2 p_j \\ &\quad + 2\delta_0 p_i^3 p_j, \\ \Pr(A_i A_i, A_j A_k) &= 2\delta_{ab} p_i p_j p_k + 2\delta_0 p_i^2 p_j p_k, \\ \Pr(A_i A_j, A_i A_j) &= 2(\delta_{ac \cdot bd} + \delta_{ad \cdot bc}) p_i p_j \\ &\quad + (\delta_{ac} + \delta_{ad} + \delta_{bc} + \delta_{bd}) p_i p_j \\ &\quad \times (p_i + p_j) + 4\delta_0 p_i^2 p_j^2, \\ \Pr(A_i A_j, A_i A_k) &= (\delta_{ac} + \delta_{ad} + \delta_{bc} + \delta_{bd}) p_i p_j p_k \\ &\quad + 4\delta_0 p_i^2 p_j p_k, \\ \Pr(A_i A_j, A_k A_l) &= 4\delta_0 p_i p_j p_k p_l. \end{aligned} \right\} \quad (2.1)$$

Although it may be easiest to calculate values for all 15 ibd measures, equation (2.1) shows that only nine functions of them are needed for joint genotype probabilities. These nine condensed measures are shown in table 2, again with the mnemonic notation of Cockerham (1971) and the simple notation of Jacquard (1970) and Cannings (1998). Each of the nine condensed ibd states can be described with the notation of Thompson (1974). The numbers a_1, a_2, a_3, a_4 indicate the two alleles in the first individual (ab in our notation) followed by the two alleles in the second individual (cd in our notation). The numbers for ibd alleles are the same, and for the condensed set 1112 indicates the same state as 1121 for example.

Although the measures displayed in tables 1 and 2 are the ones needed for joint genotype probabilities, neither set is the natural one for evaluation for a specified degree of relationship. The more natural set

Table 1. Full set of 15 ibd probabilities.

ibd relationship ^a	probability	
	Cockerham (1971)	Jacquard (1970)
$a \equiv b \equiv c \equiv d$	δ_{abcd}	δ_1
$a \equiv b \equiv c$	δ_{abc}	δ_2
$a \equiv b \equiv d$	δ_{abd}	δ_3
$a \equiv b \equiv c \equiv d$	δ_{acd}	δ_4
$b \equiv c \equiv d$	δ_{bcd}	δ_5
$a \equiv b, c \equiv d$	$\delta_{ab \cdot cd}$	δ_6
$a \equiv c, b \equiv d$	$\delta_{ac \cdot bd}$	δ_9
$a \equiv d, b \equiv c$	$\delta_{ab \cdot bc}$	δ_{12}
$a \equiv b$	δ_{ab}	δ_7
$a \equiv c$	δ_{ac}	δ_{10}
$a \equiv d$	δ_{ad}	δ_{13}
$b \equiv c$	δ_{bc}	δ_{14}
$b \equiv d$	δ_{bd}	δ_{11}
$c \equiv d$	δ_{cd}	δ_8
None	δ_0	δ_{15}

^a Alleles not listed are not ibd to those listed.

Table 2. Nine condensed measures needed for joint genotype probabilities.

state	Cockerham (1971) ^a	Jacquard (1970) ^b	Cannings (1998)
1111	$\delta_{\bar{X}\bar{Y}} = \delta_{abcd}$	Δ_1	π_1
1112	$2\delta_{\bar{X}Y} = \delta_{abc} + \delta_{abd}$	Δ_3	π_4
1222	$2\delta_{X\bar{Y}} = \delta_{acd} + \delta_{bcd}$	Δ_5	π_5
1122	$\delta_{\bar{X} \cdot \bar{Y}} = \delta_{ab \cdot cd}$	Δ_2	π_2
1212	$2\delta_{\bar{X}+Y} = \delta_{ac \cdot bd} + \delta_{ad \cdot bc}$	Δ_7	π_3
1123	$\delta_{\bar{X}\bar{X}} = \delta_{ab}$	Δ_4	π_6
1233	$\delta_{\bar{Y}\bar{Y}} = \delta_{cd}$	Δ_6	π_7
1213	$4\delta_{XY} = \delta_{ac} + \delta_{ad} + \delta_{bc} + \delta_{bd}$	Δ_8	π_8
1234	$\delta_{\bar{X}\bar{Y}} = \delta_0$	Δ_9	π_9

^a Alleles not listed are not ibd to those listed.

^b Also, Génin & Clerget-Darpoux (1996).

lists only the alleles that are ibd to each other. The alleles for which there are ibd relationships can be identified either by how they are arranged within individuals or by the individuals from which they are randomly drawn. Nothing is implied about the ibd status of alleles not identified in these measures, unlike the case in tables 1 and 2 where alleles not identified are assumed not to be ibd to those that are identified. Identifying alleles by their arrangement within individuals $X(ab)$ and $Y(cd)$ yields 14 summary measures, as shown in table 3. Identifying alleles by the individuals in which they are carried, or from which they are drawn, the summary measures can be reduced to a set of eight measures:

$$\delta_{\bar{X}\bar{Y}} = \delta_{abcd},$$

$$\gamma_{\bar{X}Y} = \frac{1}{2}(\gamma_{abc} + \gamma_{abd}) \quad \gamma_{X\bar{Y}} = \frac{1}{2}(\gamma_{acd} + \gamma_{bcd}),$$

$$\Delta_{\bar{X} \cdot \bar{Y}} = \Delta_{ab \cdot cd} \quad \Delta_{\bar{X}+Y} = \frac{1}{2}(\Delta_{ac \cdot bd} + \Delta_{ad \cdot bc}),$$

$$F_X = \theta_{ab} \quad F_Y = \theta_{cd},$$

$$\theta_{XY} = \frac{1}{4}(\theta_{ac} + \theta_{ad} + \theta_{bc} + \theta_{bd}),$$

Table 3. Fourteen summary ibd measures.

$\delta_{abcd} = \delta_{abcd}$	
$\gamma_{abc} = \delta_{abc} + \delta_{abcd}$	$\gamma_{abd} = \delta_{abd} + \delta_{abcd}$
$\gamma_{acd} = \delta_{acd} + \delta_{abcd}$	$\gamma_{bcd} = \delta_{bcd} + \delta_{abcd}$
$\Delta_{ab \cdot cd} = \delta_{ab \cdot cd} + \delta_{abcd}$	
$\Delta_{ac \cdot bd} = \delta_{ac \cdot bd} + \delta_{abcd}$	$\Delta_{ad \cdot bc} = \delta_{ad \cdot bc} + \delta_{abcd}$
$\theta_{ab} = \delta_{ab} + \delta_{abc} + \delta_{abd} + \delta_{ab \cdot cd} + \delta_{abcd}$	$\theta_{ac} = \delta_{ac} + \delta_{abc} + \delta_{acd} + \delta_{ac \cdot bd} + \delta_{abcd}$
$\theta_{ad} = \delta_{ad} + \delta_{abd} + \delta_{acd} + \delta_{ad \cdot bc} + \delta_{abcd}$	$\theta_{bc} = \delta_{bc} + \delta_{abc} + \delta_{bcd} + \delta_{ad \cdot bc} + \delta_{abcd}$
$\theta_{bd} = \delta_{bd} + \delta_{abd} + \delta_{bcd} + \delta_{ac \cdot bd} + \delta_{abcd}$	$\theta_{cd} = \delta_{cd} + \delta_{acd} + \delta_{bcd} + \delta_{ab \cdot cd} + \delta_{abcd}$

and these lead to the nine terms needed for joint genotypic probabilities:

$$\left. \begin{aligned}
 \delta_{abcd} &= \delta_{\bar{X}\bar{Y}}, \\
 \delta_{abc} + \delta_{abd} &= 2(\gamma_{\bar{X}\bar{Y}} - \delta_{\bar{X}\bar{Y}}), \\
 \delta_{acd} + \delta_{bcd} &= 2(\gamma_{X\bar{Y}} - \delta_{\bar{X}\bar{Y}}), \\
 \delta_{ab \cdot cd} &= \Delta_{\bar{X} \cdot \bar{Y}} - \delta_{\bar{X}\bar{Y}}, \\
 \delta_{ac \cdot bd} + \delta_{ad \cdot bc} &= 2(\Delta_{\bar{X} + \bar{Y}} - \delta_{\bar{X}\bar{Y}}), \\
 \delta_{ab} &= F_X - \Delta_{\bar{X} \cdot \bar{Y}} - 2\gamma_{\bar{X}\bar{Y}} + 2\delta_{\bar{X}\bar{Y}}, \\
 \delta_{cd} &= F_Y - \Delta_{X \cdot \bar{Y}} - 2\gamma_{X\bar{Y}} + 2\delta_{\bar{X}\bar{Y}}, \\
 \delta_{ac} + \delta_{ad} + \delta_{bc} + \delta_{bd} \\
 &= 4(\theta_X - \Delta_{\bar{X} + \bar{Y}} - \gamma_{\bar{X}\bar{Y}} - \gamma_{X\bar{Y}} + 2\delta_{\bar{X}\bar{Y}}), \\
 \delta_0 &= 1 - F_X - F_Y - 4\theta_{XY} + \Delta_{\bar{X} \cdot \bar{Y}} + 2\Delta_{\bar{X} + \bar{Y}} \\
 &\quad + 4\gamma_{\bar{X}\bar{Y}} + 4\gamma_{X\bar{Y}} - 6\delta_{\bar{X}\bar{Y}}.
 \end{aligned} \right\} \quad (2.2)$$

An alternative reparameterization was given by Karigl (1981) who identified alleles entirely by the individuals from which they were drawn. In table 4, we show the equivalences between his and Cockerham's measures for alleles drawn from one or two individuals.

3. SPECIAL CASES

(a) No inbreeding or relatedness

When two individuals are not related and not inbred, all four of the alleles they carry at a locus are not ibd, and all descent measures except δ_0 are zero. Joint genotypic probabilities are just the appropriate products of allele probabilities with a factor of two for each heterozygote, and conditional probabilities reduce to single genotype probabilities.

(b) No inbreeding

Non-inbred individuals do not carry ibd alleles at a locus, and the relationship between two individuals can be summarized with the probabilities that they carry zero, one or two pairs of ibd alleles. If the probabilities of these three events are written as P_0 , P_1 , P_2 , respectively:

$$P_2 = \delta_{ac \cdot bd} + \delta_{ad \cdot bc},$$

$$P_1 = \delta_{ac} + \delta_{ad} + \delta_{bc} + \delta_{bd},$$

$$P_0 = \delta_0.$$

Values of P_0 , P_1 , P_2 for common relatives are shown below in table 7, and the joint genotype probabilities in equations (2.1) are shown in table 5. Also shown in table 5 are conditional probabilities of one genotype

Table 4. Equivalent ibd measures of Cockerham and Karigl for individuals $X(ab)$ and $Y(cd)$.

Karigl (1981)	Cockerham (1971)
Φ_{XX}	$(1/2)(1 + F_X)$
Φ_{XY}	θ_{XY}
Φ_{XXX}	$(1/4)(1 + 3F_X)$
Φ_{XXY}	$(1/2)(\theta_{XY} + \gamma_{\bar{X}\bar{Y}})$
Φ_{XXXX}	$(1/8)(1 + 7F_X)$
Φ_{XXXV}	$(1/4)(\theta_{XY} + 3\gamma_{\bar{X}\bar{Y}})$
Φ_{XXYV}	$(1/4)(\theta_{XY} + \gamma_{\bar{X}\bar{Y}} + \gamma_{X\bar{Y}} + \delta_{\bar{X}\bar{Y}})$
$\Phi_{XX,XX}$	$(1/4)(1 + 3F_X)$
$\Phi_{XX,XY}$	$(1/2)(\theta_{XY} + \gamma_{\bar{X}\bar{Y}})$
$\Phi_{XX,YV}$	$(1/4)(1 + F_X + F_Y + \delta_{\bar{X}\bar{Y}})$
$\Phi_{XY,XY}$	$(1/4)(\theta_{XY} + \gamma_{\bar{X}\bar{Y}} + \gamma_{X\bar{Y}} + \Delta_{\bar{X} + \bar{Y}})$

given another, as well as the probability of two genotypes for related individuals divided by the probability of those genotypes if they are from unrelated people. This last ratio is used in the identification of human remains, when genotypes are available from a sample thought to be from a missing person and from a family member of that person (Brenner & Weir 2003).

(c) Random-mating population

In a population mating completely at random, including a random amount of selfing, the ibd relationships among alleles do not depend on the arrangement of those alleles among individuals. Cannings (1998) terms this situation 'permutable'. Instead, it is necessary to identify only the number of ibd alleles: two, three, two pairs or four. The probabilities of these relationships can be written as θ , γ , Δ or δ , respectively. In other words,

$$F_X = F_Y = \theta_{XY} = \theta, \quad \gamma_{\bar{X}\bar{Y}} = \gamma_{X\bar{Y}} = \gamma,$$

$$\Delta_{\bar{X} \cdot \bar{Y}} = \Delta_{\bar{X} + \bar{Y}} = \Delta, \quad \delta_{\bar{X}\bar{Y}} = \delta.$$

The nine condensed measures in this case are shown in table 6.

Cannings (1998) points out that, in general, the inbreeding coefficients F_X , F_Y differ from coancestry coefficients θ_{XY} . The distinction is necessary if there is not completely random union of gametes, but even for dioecious diploids, in which selfing is excluded, the difference will be very small, especially in equilibrium populations. The distinction is not needed in the permutable case, and Canning's parameterization for that case is shown in table 6. He uses three parameters, instead of the four that we use, apparently because he

Table 5. Joint and conditional probabilities for non-inbred relatives X and Y .

X, Y	$\Pr(X, Y)$			$\Pr(X Y)$			$\Pr(X, Y P_0, P_1, P_2)/\Pr(X, Y P_0 = 1)$		
	P_0	P_1	P_2	P_0	P_1	P_2	P_0	P_1	P_2
$A_i A_i, A_i A_i$	p_i^4	p_i^3	p_i^2	p_i^2	p_i	1	1	$1/p_i$	$1/p_i^2$
$A_i A_i, A_j A_j$	$p_i^2 p_j^2$	—	—	p_i^2	—	—	1	—	—
$A_i A_i, A_i A_j$	$2p_i^3 p_j$	$p_i^2 p_j$	—	p_i^2	$p_i/2$	—	1	$1/2p_i$	—
$A_i A_i, A_j A_k$	$2p_i^2 p_j p_k$	—	—	p_i^2	—	—	1	—	—
$A_i A_j, A_i A_j$	$4p_i^2 p_j^2$	$p_i p_j(p_i + p_j)$	$2p_i p_j$	$2p_i p_j$	$(p_i + p_j)/2$	1	1	$(1/4p_i) + (1/4p_j)$	$1/2p_i p_j$
$A_i A_j, A_i A_k$	$4p_i^2 p_j p_k$	$p_i p_j p_k$	—	$2p_i p_j$	$p_j/2$	—	1	$1/4p_i$	—
$A_i A_j, A_k A_l$	$4p_i p_j p_k p_l$	—	—	$2p_i p_j$	—	—	1	—	—

Table 6. Alternative values for condensed ibd measures in an equilibrium population.

State	general	Génin & Clerget-Darpoux (1998b)	Cannings (1998)	Dirichlet ^a
1111	δ	α^3	$\alpha^2\beta$	$6\theta^3$
1112	$2(\gamma - \delta)$	$2\alpha^2(1 - \alpha)$	$\alpha(1 - \alpha)\gamma$	$4\theta^2(1 - \theta)$
1222	$2(\gamma - \delta)$	$2\alpha^2(1 - \alpha)$	$\alpha(1 - \alpha)\gamma$	$4\theta^2(1 - \theta)$
1122	$\Delta - \delta$	$\alpha^2(1 - \alpha)$	$\alpha^2(1 - \beta)$	$\theta^2(1 - \theta)$
1212	$2(\Delta - \delta)$	$2\alpha^2(1 - \alpha)$	$2\alpha^2(1 - \beta)$	$2\theta^2(1 - \theta)$
1123	$\theta - 2\gamma - \Delta + 2\delta$	$\alpha(1 - \alpha)(1 - 2\alpha)$	$\alpha(1 - \alpha)(1 - \gamma)$	$\theta(1 - \theta)^2$
1233	$\theta - 2\gamma - \Delta + 2\delta$	$\alpha(1 - \alpha)(1 - 2\alpha)$	$\alpha(1 - \alpha)(1 - \gamma)$	$\theta(1 - \theta)^2$
1213	$4(\theta - 2\gamma - \Delta + 2\delta)$	$4\alpha(1 - \alpha)(1 - 2\alpha)$	$4\alpha(1 - \alpha)(1 - \gamma)$	$4\theta(1 - \theta)^2$
1234	$1 - 6\theta + 8\gamma + 3\Delta - 6\delta$	$(1 - \alpha)(1 - 2\alpha)(1 - 3\alpha)$	$1 - 6\alpha + 3\alpha^2 + 4\alpha(1 - \alpha)\gamma + 2\alpha^2\beta$	$(1 - \theta)^3$

^a Each term to be divided by $(1 + \theta)(1 + 2\theta)$.

regards the probability of two pairs of ibd alleles as the square of the probability of one pair of ibd alleles. His α , α^2 , $\alpha^2\beta$ are the same as our θ , Δ , δ . His $\alpha(1 - \alpha)\gamma$ is the same as our $2(\gamma - \delta)$. Our Δ is the probability that two pairs of alleles are ibd, whether or not all four alleles are ibd. Therefore, Δ is greater than the probability δ that all four alleles are ibd, and it is not the product of the single pair probabilities.

In the special case when the mutation process does not depend on allelic state, then allele frequencies have a Dirichlet distribution over replicate populations (Balding & Nichols 1994). This provides

$$\gamma = \frac{2\theta^2}{1 + \theta},$$

$$\Delta = \frac{\theta^2(1 + 5\theta)}{(1 + \theta)(1 + 2\theta)},$$

$$\delta = \frac{6\theta^3}{(1 + \theta)(1 + 2\theta)},$$

and the nine condensed ibd probabilities reduce to the values shown in table 6. They are all functions of the single parameter θ . Génin & Clerget-Darpoux (1998b) also presented a single-parameter formulation (table 6). Their α is the same as our θ , and they assign α^2 to both of our γ and Δ , and they assign α^3 to our δ .

The joint genotypic probabilities in the Dirichlet case are

$$\begin{aligned} \Pr(A_i A_i, A_i A_i) &= [3\theta + (1 - \theta)p_i][2\theta + (1 - \theta)p_i] \\ &\quad \times [\theta + (1 - \theta)p_i]p_i/(1 + \theta)(1 + 2\theta), \end{aligned}$$

$$\begin{aligned} \Pr(A_i A_i, A_j A_j) &= (1 - \theta)[\theta + (1 - \theta)p_i] \\ &\quad \times [\theta + (1 - \theta)p_j]p_i p_j/(1 + \theta)(1 + 2\theta), \end{aligned}$$

$$\begin{aligned} \Pr(A_i A_i, A_i A_j) &= 2(1 - \theta)[2\theta + (1 - \theta)p_i] \\ &\quad \times [\theta + (1 - \theta)p_i]p_i p_j/(1 + \theta)(1 + 2\theta), \end{aligned}$$

$$\begin{aligned} \Pr(A_i A_i, A_j A_k) &= 2(1 - \theta)^2[\theta + (1 - \theta)p_i]p_i p_j p_k/(1 + \theta)(1 + 2\theta), \end{aligned}$$

$$\begin{aligned} \Pr(A_i A_j, A_i A_j) &= 4(1 - \theta)[\theta + (1 - \theta)p_i] \\ &\quad \times [\theta + (1 - \theta)p_j]p_i p_j/(1 + \theta)(1 + 2\theta), \end{aligned}$$

$$\begin{aligned} \Pr(A_i A_j, A_i A_k) &= 4(1 - \theta)^2[\theta + (1 - \theta)p_i]p_i p_j p_k/(1 + \theta)(1 + 2\theta), \end{aligned}$$

$$\Pr(A_i A_j, A_k A_l) = 4(1 - \theta)^3 p_i p_j p_k p_l/(1 + \theta)(1 + 2\theta).$$

These equations may be derived directly from Ewens' sampling formula: the probability of allele A_i given that n_i copies of that allele have already been found in a set of n alleles, is

$$\Pr(A_i | n_i \text{ copies of } A_i) = \frac{n_i \theta + (1 - \theta)p_i}{1 + (n - 1)\theta}.$$

The match probability in forensic science is the probability that an unknown person has a particular

Table 7. Summary ibd probabilities for relatives in an equilibrium population.

	parent-child ^a	grandparent-grandchild ^b	full sibs ^c	half sibs ^d	first cousins ^e	unrelated
δ_{abcd}	γ	$(\gamma + \delta)/2$	$(\theta + 2\gamma + \delta)/4$	$(\gamma + \delta)/2$	$(\gamma + 3\delta)/4$	δ
γ_{abc}	θ	$(\theta + \gamma)/2$	$(\theta + \gamma)/2$	$(\theta + \gamma)/2$	$(\theta + 3\gamma)/4$	γ
γ_{abd}	γ	γ	$(\theta + \gamma)/2$	γ	γ	γ
γ_{acd}	$(\theta + \gamma)/2$	$(\theta + 3\gamma)/4$	$(\theta + \gamma)/2$	γ	γ	γ
γ_{bcd}	$(\theta + \gamma)/2$	$(\theta + 3\gamma)/4$	$(\theta + \gamma)/2$	$(\theta + \gamma)/2$	$(\theta + 3\gamma)/4$	γ
$\Delta_{ab \cdot cd}$	γ	$(\gamma + \Delta)/2$	$(\theta + 2\gamma + \Delta)/4$	$(\gamma + \Delta)/2$	$(\gamma + 3\Delta)/4$	Δ
$\Delta_{ac \cdot bd}$	$(\theta + \gamma)/2$	$(\theta + \gamma + 2\Delta)/4$	$(1 + 2\theta + \Delta)/4$	$(\gamma + \Delta)/2$	$(\gamma + 3\Delta)/4$	Δ
$\Delta_{ad \cdot bc}$	$(\theta + \gamma)/2$	$(\theta + \gamma + 2\Delta)/4$	$(\theta + 2\gamma + \Delta)/4$	$(\theta + \Delta)/2$	$(\theta + 3\Delta)/4$	Δ
θ_{ab}	θ	θ	θ	θ	θ	θ
θ_{ac}	$(1 + \theta)/2$	$(1 + 3\theta)/4$	$(1 + \theta)/2$	θ	θ	θ
θ_{ad}	θ	θ	θ	θ	θ	θ
θ_{bc}	$(1 + \theta)/2$	$(1 + 3\theta)/4$	θ	$(1 + \theta)/2$	$(1 + 3\theta)/4$	θ
θ_{bd}	θ	θ	$(1 + \theta)/2$	θ	θ	θ
θ_{cd}	θ	θ	θ	θ	θ	θ
P_0^f	0	1/2	1/2	1/2	3/4	1
P_1^f	1	1/2	1/2	1/2	1/4	0
P_2^f	0	0	1/4	0	0	0

^a Child cd receives allele c from parent ab .^b Grandchild cd receives allele c from grandparent ab .^c Full sibs ab and cd receive alleles a, c from one parent and b, d from the other.^d Half sibs ab and cd receive alleles b, c from common parent.^e Cousins ab and cd receive alleles b, c from full sibs.^f For the case when $\theta = \gamma = \Delta = \delta = 0$.

genotype given that a known person has been found to have that genotype (Balding & Nichols 1994; Weir 1994). In the Dirichlet case, these match probabilities are

$$\Pr(A_i A_i | A_i A_i) = \frac{[3\theta + (1 - \theta)p_i][2\theta + (1 - \theta)p_i]}{(1 + \theta)(1 + 2\theta)},$$

$$\Pr(A_i A_j | A_i A_j) = \frac{2[\theta + (1 - \theta)p_i][\theta + (1 - \theta)p_j]}{(1 + \theta)(1 + 2\theta)}.$$

It is of forensic significance that these match probabilities are greater than the genotype probabilities $\Pr(A_i A_i)$ and $\Pr(A_i A_j)$.

Weir (1994) gave alternative relationships between θ , γ , Δ , δ for the pure drift case. His approximations were based on the transition equations for the four measures over time.

(d) Relatives in a random-mating population

When relatives are taken from a population with a common background level of relatedness, account must be taken of the fact that they will be inbred. The simplest situation is when the population is in equilibrium with respect to ibd status of random sets of alleles. The four alleles in the two relatives in question are traced back to the most recent common ancestors of those individuals, where they may be copies of four or fewer alleles in the ancestors. The relationships among those alleles are governed by the population-wide measures θ , γ , Δ , δ regardless of which generation the ancestors belong to. Because alleles are being traced back from relatives to common ancestors, it is much easier to work with the summary measures. The values are shown in table 7. Weir (1994) showed the more complicated expressions for the original 15 measures.

(e) Sibs from related parents

As a final special case, suppose two full sibs have parents who are related by family membership and by membership of an equilibrium population. For example, suppose two sibs have parents $X(ab)$ and $Y(cd)$ who are cousins. This case was considered by Génin & Clerget-Darpoux (1996, 1998a,b) and by Cannings (1998). The relationships among the four parental alleles are given in table 7, and in particular:

$$\delta_{\bar{X}\bar{Y}} = \frac{1}{4}(\gamma + 3\delta),$$

$$\gamma_{\bar{X}\bar{Y}} = \gamma_{X\bar{Y}} = \frac{1}{8}(\theta + 7\gamma),$$

$$\Delta_{\bar{X} \cdot \bar{Y}} = \frac{1}{4}(\gamma + 3\Delta),$$

$$\Delta_{\bar{X} + \bar{Y}} = \frac{1}{8}(\theta + \gamma + 6\Delta),$$

$$F_X = F_Y = \theta,$$

$$\theta_{XY} = \frac{1}{16}(1 + 15\theta).$$

If the sibs are $G(xy)$ and $H(zw)$, where alleles x, z are from cousin X and alleles y, w are from cousin Y , then the summary measures for these alleles are

$$\begin{aligned} \delta_{xyzw} &= \frac{1}{4}(\theta_{XY} + \gamma_{\bar{X}\bar{Y}} + \gamma_{X\bar{Y}} + \delta_{\bar{X}\bar{Y}}) \\ &= \frac{1}{64}(1 + 19\theta + 32\gamma + 12\delta), \end{aligned}$$

$$\gamma_{xyz} = \gamma_{xzw} = \frac{1}{2}(\theta_{XY} + \gamma_{\bar{X}\bar{Y}}) = \frac{1}{32}(1 + 17\theta + 14\gamma),$$

$$\gamma_{xyw} = \gamma_{yzw} = \frac{1}{2}(\theta_{XY} + \gamma_{X\bar{Y}}) = \frac{1}{32}(1 + 17\theta + 14\gamma),$$

$$\begin{aligned} \Delta_{xy \cdot zw} &= \Delta_{xw \cdot yz} = \frac{1}{4}(\theta_{XY} + \gamma_{\bar{X}\bar{Y}} + \gamma_{X\bar{Y}} + \Delta_{\bar{X} + \bar{Y}}) \\ &= \frac{1}{64}(1 + 21\theta + 30\gamma + 12\Delta), \end{aligned}$$

$$\begin{aligned}\Delta_{xz \cdot yw} &= \frac{1}{4}(1 + F_X + F_Y + \Delta_{\bar{X} \cdot \bar{Y}}) \\ &= \frac{1}{16}(4 + 8\theta + \gamma + 3\Delta), \\ \theta_{xy} &= \theta_{xw} = \theta_{yz} = \theta_{zw} = \theta_{XY} = \frac{1}{16}(1 + 15\theta), \\ \theta_{xz} &= \frac{1}{2}(1 + F_X) = \frac{1}{2}(1 + \theta), \\ \theta_{yw} &= \frac{1}{2}(1 + F_Y) = \frac{1}{2}(1 + \theta).\end{aligned}$$

The nine condensed probabilities needed for the joint genotype probabilities are

$$\begin{aligned}\delta_{xyzw} &= \frac{1}{64}(1 + 19\theta + 32\gamma + 12\Delta), \\ \delta_{xyz} + \delta_{xyw} &= \delta_{xzw} + \delta_{yzw} = \frac{1}{32}(1 + 15\theta - 4\gamma - 12\Delta), \\ \delta_{xy \cdot zw} &= \frac{1}{32}(\theta - \gamma + 6\Delta - 6\delta), \\ \delta_{xz \cdot yw} + \delta_{xw \cdot yz} &= \frac{3}{64}(5 + 5\theta - 10\gamma + 8\Delta - 8\delta), \\ \delta_{xy} &= \delta_{zw} = \frac{1}{64}(1 + 9\theta - 22\gamma - 12\Delta + 24\delta), \\ \delta_{xz} + \delta_{xw} + \delta_{yz} + \delta_{yw} \\ &= \frac{1}{32}(15 - 21\theta - 18\gamma - 24\Delta + 48\delta), \\ \delta_0 &= \frac{1}{16}(3 - 18\theta + 24\gamma + 9\Delta - 18\delta).\end{aligned}$$

These are the same as those of Génin & Clerget-Darpoux (1998a) under their assumption that $\gamma = \Delta = \theta^2$, $\delta = \theta^3$ and, therefore, in the case that $\theta = \gamma = \Delta = \delta = 0$. Otherwise, they are slightly different, and they do not have the pattern that suggested to Génin & Clerget-Darpoux (1996) that some of them were equal and could be pooled.

In their discussion of affected-sib tests, Génin & Clerget-Darpoux (1998a) gave the probabilities that two sibs whose parents were cousins share 0, 1 or two pairs of alleles ibd. These three probabilities, using our results, are

$$\begin{aligned}P_0 &= \delta_{xy \cdot zw} + \delta_{xy} + \delta_{zw} + \delta_0 \\ &= \frac{1}{32}(7 - 26\theta + 25\gamma + 12\Delta - 18\delta), \\ P_1 &= (\delta_{xyz} + \delta_{xyw}) + (\delta_{xzw} + \delta_{yzw}) \\ &\quad + (\delta_{xz} + \delta_{xw} + \delta_{yz} + \delta_{yw}) \\ &= \frac{1}{32}(17 + 9\theta - 26\gamma - 24\Delta + 24\delta), \\ P_2 &= \delta_{xyzw} + (\delta_{xz \cdot yw} + \delta_{xw \cdot yz}) \\ &= \frac{1}{32}(8 + 17\theta + \gamma + 12\Delta - 6\delta).\end{aligned}$$

In the Dirichlet case, these reduce to

$$\begin{aligned}P_0 &= (1 - \theta)(7 + 2\theta)/32(1 + \theta)(1 + 2\theta), \\ P_1 &= (1 - \theta)(17 + 77\theta + 62\theta^2)/32(1 + \theta)(1 + 2\theta), \\ P_2 &= (8 + 41\theta + 81\theta^2 + 62\theta^3)/32(1 + \theta)(1 + 2\theta).\end{aligned}$$

4. DISCUSSION

Génin & Clerget-Darpoux (1996) pointed out that affected sib pair tests of linkage (Blackwelder & Elston 1985) can have higher than expected type I errors when ‘remote consanguinity’ is ignored. They were referring

to the low level of inbreeding and relatedness in actual populations caused by the evolutionary process, primarily drift. In this article, we quantify the degree of inbreeding for relatives taken from such populations, and show how our results differ from those of Génin & Clerget-Darpoux (1996, 1998a,b) and Cannings (1998). The degree of inbreeding of relatives is further increased if their parents were related, first cousins for example, and our treatment allows those cases also to be considered.

Although complete treatments of joint genotype probabilities require knowledge of 15 different patterns of ibd for four alleles, this is too cumbersome an approach in general. Assumptions of random mating and ‘permutability’ can reduce this system down to just four probabilities, but estimation procedures are in place for only the two-allele probability (Weir & Hill 2002). There is great appeal, therefore, in assuming that allele frequencies follow a Dirichlet distribution so that only the two-allele measure is needed. There are sound reasons for adopting this distribution for the infinite-alleles mutation model (Balding & Nichols 1994), but some slight ground for concern for stepwise mutation models (Graham *et al.* 2000). Even for microsatellite markers, however, where the stepwise model may have relevance, some empirical evidence on matching probabilities suggests that results based on the Dirichlet distribution can be used with confidence. As we have already done (Liu & Weir 2004) with the affected sib pair tests of Lange (1986a,b), we suggest that tests of the type proposed by Génin & Clerget-Darpoux (1996) be developed for inbred affected relatives and that these tests be constructed on the basis of joint genotype probabilities expressed in terms of allele frequencies and the coancestry coefficient θ .

This work was supported in part by NIH grant GM 45344. Helpful comments were made by Dr Gary Chase.

This work is dedicated to Bill Hill in recognition of 25 years of friendship and collaboration.

REFERENCES

- Balding, D. J. & Nichols, R. A. 1994 DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. *Forensic Sci. Int.* **64**, 125–140.
- Blackwelder, W. C. & Elston, R. C. 1985 A comparison of sib-pair linkage tests for disease susceptibility loci. *Genet. Epidemiol.* **2**, 85–97.
- Brenner, C. H. & Weir, B. S. 2003 Issues and strategies in the DNA identification of World Trade Center victims. *Theor. Popul. Biol.* **63**, 173–178.
- Cannings, C. 1998 On the probabilities of identity states in permutable populations. *Am. J. Hum. Genet.* **62**, 698–702.
- Cockerham, C. C. 1971 Higher order probability functions of identity of alleles by descent. *Genetics* **69**, 235–246.
- Génin, E. & Clerget-Darpoux, F. 1996 Consanguinity and sib-pair method: an approach using identity by descent between and within individuals. *Am. J. Hum. Genet.* **59**, 1149–1162.
- Génin, E. & Clerget-Darpoux, F. 1998a Reply to Weeks and Sinsheimer. *Am. J. Hum. Genet.* **62**, 731–736.
- Génin, E. & Clerget-Darpoux, F. 1998b On the probability of identity states in permutable populations: reply to Cannings. *Am. J. Hum. Genet.* **62**, 726–727.

- Gillios, M. 1965 Relation d'identité en génétique. I. Postulats et axiomes mendéliens. II. Corrélation génétique dans le cas de dominance. *Ann. Inst. Henri Poincaré (A) Phys. Théor.* **2**, 1–94.
- Graham, J. J., Curran, J. & Weir, B. S. 2000 Conditional genotypic probabilities for microsatellite loci. *Genetics* **155**, 1973–1980.
- Harris, D. L. 1964 Genotypic covariances between inbred relatives. *Genetics* **50**, 1319–1348.
- Jacquard, A. 1970 *Structures Génétiques des Populations*. Paris: Masson & Cie. [*Genetics of human populations*, Springer-Verlag, New York, 1974]
- Karigl, G. 1981 A recursive algorithm for the calculation of identity coefficients. *Ann. Hum. Genet.* **45**, 299–305.
- Lange, K. 1986a A test statistic for the affected-sib-set method. *Ann. Hum. Genet.* **50**, 283–290.
- Lange, K. 1986b The affected sib-pair method using identity by descent relations. *Am. J. Hum. Genet.* **39**, 148–150.
- Liu, W. & Weir, B. S. 2004 Affected sib pair tests in inbred populations. *Ann. Hum. Genet.* **68**, 606–619.
- Merette, C. & Ott, J. 1996 Estimating parental relationship in linkage analysis of recessive traits. *Am. J. Med. Genet.* **63**, 386–391.
- Ritland, K. & Travis, S. 2004 Inferences involving individual coefficients of relatedness and inbreeding in natural populations of *Abies*. *For. Ecol. Manage.* **197**, 171–180.
- Thompson, E. A. 1974 Gene identities and multiple relationships. *Biometrics* **30**, 667–680.
- Weeks, D. E. & Sinsheimer, J. S. 1998 Consanguinity and relative-pair methods for linkage analysis. *Am. J. Hum. Genet.* **62**, 728–731.
- Weir, B. S. 1994 The effects of inbreeding on forensic calculations. *Annu. Rev. Genet.* **28**, 597–621.
- Weir, B. S. & Hill, W. G. 2002 Estimating *F*-statistics. *Annu. Rev. Genet.* **36**, 721–750.
- Wright, A. J. 1987 Covariance of inbred relatives with special reference to selfing. *Genetics* **115**, 545–552.