Variation in actual relationship among descendants of inbred individuals

Genet. Res., Camb. (2012), **94**, *pp.* 267–274. © Cambridge University Press 2012
doi:10.1017/S0016672312000468

Variation in actual relationship among descendants of inbred individuals

W. G. HILL[1]* AND B. S. WEIR[2]
[1] *Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, UK*
[2] *Department of Biostatistics, University of Washington, P.O. Box 357232, Seattle, WA 98195-7232, USA*

(*Received 1 June 2012; revised 6 July 2012; accepted 25 August 2012*)

Summary

In previous analyses, the variation in actual, or realized, relationship has been derived as a function of map length of chromosomes and type of relationship, the variation being greater the shorter the total chromosome length and the coefficient of variation being greater the more distant the relationship. Here, the results are extended to allow for the relatives' ancestor being inbred. Inbreeding of a parent reduces variation in actual relationship among its offspring, by an amount that depends on the inbreeding level and the type of mating that led to that level. For descendants of full-sibs, the variation is reduced in later generations, but for descendants of half-sibs, it is increased.

1. Introduction

Measures of relationship specify the probabilities that relatives share alleles identical by descent (ibd), with the actual or realized identity at individual loci binomially distributed due to Mendelian segregation. At individual loci, the actual identity by descent is binomially distributed, but because of the linkage, there are covariances in this quantity among loci; therefore, there is still variation in the proportion of alleles-shared ibd and hence in the actual or realized relationship, even assuming infinitely many genomic sites. In previous papers, formulae for this variance have been obtained (Stam, 1980; Hill, 1993*a*, *b*; Guo, 1995; Visscher, 2009) and have recently been generalized to cover all relationships (Hill & Weir, 2011, subsequently HW11). In the previous analyses, ancestors were assumed not to be inbred; although formulae for variation in the actual inbreeding have been obtained by adapting those for variation in relationship (HW11).

The magnitude of the variation in actual relationship is important in several contexts, discussed further by HW11. These include the need to allow for relationship in genomic data cleaning and in association studies (Laurie *et al.*, 2010) and the ability to assess the pedigree relationship using genome sharing rather than just genotypes at individual loci, thereby incorporating the correlation structure induced by linkage. In quantitative genetic applications, the accuracy of prediction of breeding values in genomic selection programmes (Meuwissen *et al.*, 2001) and of estimation of quantitative genetic parameters from variation within families (Visscher *et al.*, 2006) depend on the variation in actual relationship.

Partially inbred individuals are found in all populations, arising from matings of close relatives such as full-sibs, more distant ones such as second cousins, and innumerable complex situations. Data from dense SNP markers and sequencing enable shared identity of genomic regions of individuals to be established (Weir *et al.*, 2006). For example, inbred individuals are found in some of the GENEVA consortium data being used in human genome-wide association studies (Cornelis *et al.*, 2010), from which variation in actual relationship has been demonstrated (Laurie *et al.*, 2010; HW11). Among pairs of individuals with the same pedigrees, there can be considerable variation in the estimates of the proportions of loci at which they share zero, one or two pairs of alleles ibd. In addition to the non-zero levels of inbreeding found in natural populations, deliberate inbreeding is undertaken in some breeding programmes. We now extend the results on variation in identity states

* Corresponding author: Institute of Evolutionary Biology, School of Biological sciences, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, UK. E-mail: w.g.hill@ed.ac.uk

Table 1. *Two-locus coancestry*† $\theta^*_{A;A}$ (c) *of individual A with itself as a function of the one- and two-locus inbreeding coefficients* F_A *and* F^*_A (c) *of* A. *Individual* A *has genotype* $m_i m_j/p_i p_j$ *at loci* i, j.

		Second gamete from A			
	$\theta^*_{A;A}(c)$	$\Pr(m_i m_j)=\frac{1-c}{2}$	$\Pr(m_i p_j)=\frac{c}{2}$	$\Pr(p_i m_j)=\frac{c}{2}$	$\Pr(p_i p_j)=\frac{1-c}{2}$
Frist gamete from A	$\Pr(m_i m_j)=\frac{1-c}{2}$	1	F_A	F_A	$F^*_A(c)$
	$\Pr(m_i p_j)=\frac{c}{2}$	F_A	1	$F^*_A(c)$	F_A
	$\Pr(p_i m_j)=\frac{c}{2}$	F_A	$F^*_A(c)$	1	F_A
	$\Pr(p_i m_j)=\frac{1-c}{2}$	$F^*_A(c)$	F_A	F_A	1

†The probability that two gametes from A carry identical by descent (ibd) alleles at both loci.

obtained for non-inbred ancestors to those where the common ancestors of relatives are inbred.

The notation and methodology used here are based heavily on that used previously (HW11). Basically, the probability that descendants each carry identical alleles at a pair of linked loci is computed dependent on the relationship among the parents. The excess of this probability over that assuming loci are unlinked provides an estimate of the covariance that single sites carry identical alleles, and integrating this covariance over all pairs of sites provides the variance of actual identity. The analysis is extended here to include the probability that the parent or parents share alleles at pairs of linked loci as a consequence of their relatedness and the inbreeding of their common ancestors.

2. Measures of identity by descent

The inbreeding coefficient F_X, the probability of ibd alleles at a locus, of an individual X in a pedigree is known to follow from the path-counting equation $\sum(1/2)^t\theta_{A;A}$ where t is the number of individuals in a pedigree loop linking the individual's parents to their common ancestor A, and $\theta_{A;A}$ is the coancestry of A with itself: the probability that two alleles transmitted by that individual are ibd. This coancestry is given by $\theta_{A;A}=(1+F_A)/2$, where F_A is the inbreeding coefficient of A. The count t includes the two parents but excludes the common ancestor, the factor 1/2 is for the passage of an allele through each individual in the pedigree loop, and the sum is over all distinct loops to A and over all common ancestors A.

For two loci, with recombination rate c between them, the path-counting equation for the probability of X receiving alleles ibd at each locus, through transmission of the ibd segments including both loci, is $[(1-c)/2]^t\theta^*_{A;A}(c)$ where $\theta^*_{A;A}(c)$ is the two-locus coancestry for A with itself. This has value

$$\theta^*_{A;A}(c)=F_A+\beta[1-2F_A+F^*_A(c)], \quad (1)$$

where $\beta=[(1-c)^2+c^2]/2$, as shown in Table 1 (Weir & Cockerham, 1969). Here, $F^*_A(c)$ is the two-locus inbreeding coefficient, or the probability that A has ibd alleles at both loci. Note that when the loci are completely linked, $c=0$ ($\beta=1/2$), $F^*_A(0)=F_A$ and $\theta^*_{A;A}(0)=\theta_{A;A}$. When the loci segregate independently, $c=1/2$ ($\beta=1/4$), $F^*_A(1/2)=F^2_A$ and $\theta^*_{A;A}(1/2)=\theta^2_{A;A}$.

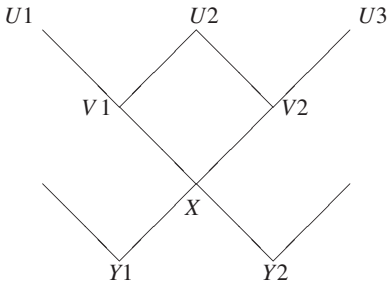

Fig. 1. Pedigree for HS offspring $Y1, Y2$ of individual X, the offspring of HS parents.

The inbreeding coefficient of an individual is also the coancestry of its parents, so if X has parents $V1$, $V2$ (e.g. Figs 1 and 2) then $F_X=\theta_{V1;V2}$. Although these two quantities are equal, they have different reference points: the coancestries θ, θ^* are for alleles on gametes transmitted by individuals, whereas the inbreeding coefficients F, F^* are for alleles on gametes received by an individual, i.e. on gametes within an individual. There is need for this last perspective for more than one individual: $\psi_{Y1;Y2}$ or $\psi^*_{Y1;Y2}(c)$ are the probabilities of ibd for alleles at one or two loci on gametes received by individuals $Y1$ and $Y2$. Clearly, $F_X=\psi_{X;X}$. The same path-counting equations hold for $\psi_{Y1;Y2}$ as for $\theta_{Y1;Y2}$, but the count t then excludes $Y1$ and $Y2$.

(i) *Inbred individual examples*

Consider an inbred individual X, the offspring of a mating of half-sibs $V1$ and $V2$ who have common parent $U2$ (Fig. 1). The probability for alleles at any locus of X being ibd is the inbreeding coefficient $F_X=1/8$, and the variance in actual inbreeding

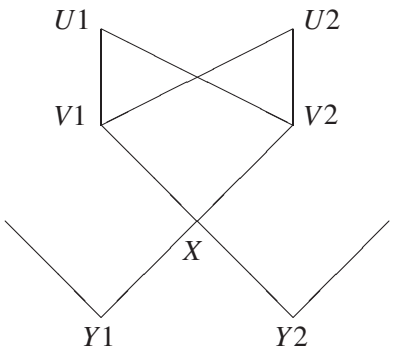

Fig. 2. Pedigree for HS offspring $Y1, Y2$ of individual X, the offspring of FS parents.

among independent loci is $F_X(1-F_X)=7/64$. For a recombination fraction between these sites of c, the two-locus inbreeding coefficient of X is given by $F^*_X(c)=(1-c)^2\beta/4$ (Table 2), which reduces to 1/8 when $c=0$ and to 1/64 when $c=1/2$, i.e. F_X and F^2_X, respectively. This argument is fairly easy to see because the probability of ibd for an individual at both sites is the same as the probability that two random haplotypes, sampled one from each parent, are ibd. Therefore, in this case where parents $V1$ and $V2$ are half-sibs, it is the probability that a pair of half-cousins, one with parent $V1$ and one with parent $V2$, share identical alleles at the two loci (HW11, Table 2). Weir & Cockerham (1969) presented a general algorithm for finding the probability of identity for alleles a, a' and b, b' at loci **A** and **B**, as shown in Appendix A.

Alternatively, consider an inbred individual X, the offspring of a mating of full-sibs $V1$ and $V2$ who have parents $U1$ and $U2$ (Fig. 2). The one- and two-locus inbreeding coefficients are $F_X=1/4$ and $F^*_X(c)=(1-c)^2\beta/2+c^2/8$ (Table 2, Appendix A). The two-locus value reduces to 1/4 when $c=0$ and to 1/16 when $c=1/2$, i.e. F_X and F^2_X, respectively. These results also follow as the probabilities of identity for alleles carried by two first cousins, one with parent $V1$ and one with parent $V2$ (HW11, Table 2).

3. Descendants of half-sibs

For unilineal relatives $V1$, $V2$ (e.g. Fig. 1), the inbreeding coefficient F_X of their offspring X is the probability k_1 they share and transmit a pair of alleles ibd, and the path-counting equation is for identity resulting from that pair of alleles descending from common ancestor $U2$. The actual state of identity can be indicated by the variable $\check{k}_1$ that takes the value 1 for identical alleles and 0 for non-identity. Taking expectations over all loci $\mathcal{E}(\check{k}_{1i})=k_1$ and $\mathrm{Var}(\check{k}_{1i})=k_1(1-k_1)$. At two loci, i, j, the actual inbreeding coefficient is $\check{F}^*_X(c)=\check{k}_{1i}\check{k}_{1j}$ and this has expectation $F^*_X(c)=\mathcal{E}(\check{k}_{1i}\check{k}_{1j})=F^2_X+\mathrm{Cov}(\check{k}_{1i}, \check{k}_{1j})$. The variance in the actual inbreeding of X averaged over the genome involves the sum of the variances at individual sites and the covariances at pairs of sites. With a large number of sites, it is the contribution of the covariances that dominates.

The relatedness of unilineal relatives also depends only on the measure k_1. If $\check{k}_{1i}$ indicates actual ibd status at locus i for the half-sibs $Y1, Y2$ with common parent X (Fig. 1)

$$\mathcal{E}(\check{k}_{1i})=\theta_{X;X}=\frac{1}{2}(1+F_X),$$

$$\mathcal{E}(\check{k}_{1i}\,\check{k}_{1j})=\theta^*_{X;X}(c)=F_X+\beta[1-2F_X+F^*_X(c)],$$

$$\mathrm{Cov}(\check{k}_{1i}, \check{k}_{1j})=\theta^*_{X;X}(c)-\theta^2_{X;X}=\theta^*_{X;X}(c)-\theta^*_{X;X}(1/2).$$

To predict the sharing of ibd pairs of alleles by individuals who are descendants of $Y1$ and $Y2$ but are otherwise unrelated, note that the probability of a gametic pair of alleles is transmitted from parent to offspring is $(1-c)/2$ and to t-th generation descendants is $[(1-c)/2]^t$. For example, $t=1$ is for half-uncle nephew (e.g. $Y1$ and the offspring of $Y2$) and $t=2$ is for half-cousins (e.g offspring of $Y1$ and $Y2$) or half-great uncle-great nephew (e.g. $Y1$ with a grandson of $Y2$). For descendants $Z1, Z2$ of $Y1$, $Y2$ such that there are t individuals (excluding $Z1, Z2, X$) in the loop from $Z1$ to X to $Z2$, $\mathcal{E}(\check{k}_{1i}, \check{k}_{1j})=\psi^*_{Z1;Z2}(c)$ and

$$\mathcal{E}(\check{k}_{1i}\,\check{k}_{1j})=\left(\frac{1-c}{2}\right)^t\theta^*_{X;X}(c). \qquad (2)$$

To facilitate calculations over multiple generations, and to integrate over the chromosomes, we adopt methods used previously (HW11). Details are given in Appendix B. Letting $b=(1-c)/2$, we can write the right-hand side of eqn (2) as $\sum_n \alpha_n b^n$, and recognizing that setting $c=1/2$, $b=1/4$ (independent loci) gives the product of expected values $\mathcal{E}(\check{k}_{1i})$, $\mathcal{E}(\check{k}_{1j})$:

$$\mathrm{Cov}(\check{k}_{1i}, \check{k}_{1j})=\sum_n \alpha_n\left[b^n-\left(\frac{1}{4}\right)^n\right].$$

The range of values of n, and the values of α_n, depend on the pedigree of the common ancestor X and we give common examples of $\theta^*_{X;X}(c)$ in Table 2 (essentially for $t=1$).

Assuming Haldane's mapping function, for a chromosome of length l Morgans, and computing the variance of actual relationship as the mean covariance over all pairs of loci,

$$\mathrm{Var}(\check{k}_1, l)=\textstyle\sum_n \alpha_n\phi_n(l),$$

where (Appendix B)

$$\phi_n(l)=\begin{cases}\dfrac{1}{2l^2}\left(\dfrac{1}{4}\right)^n\displaystyle\sum_{r=1}^{n}\binom{n}{r}\frac{2rl-1+e^{-2rl}}{r^2}, & n\geqslant 1,\\ 0, & n\leqslant 0.\end{cases} \qquad (3)$$

Table 2. *Correspondence between relationship and identity coefficients for common ancestor* X *at linked loci as a function of recombination rate* c, $\beta=[(1-c)^2+c^2]/2$ *and of* $b=(1-c)/2$.

Parents of X	Relationship-equivalent offspring	F_X	$F^*_X(c)$	$\theta^*_{X;X}(c)=F_X+\beta[1-2F_X+F^*_X(c)]$
One parent (selfing)	HS	$\frac{1}{2}$	β	$16b^4-16b^3+8b^2-2b+\frac{3}{4}$
Parent and offspring	Half-uncle nephew	$\frac{1}{4}$	$\frac{1}{2}(1-c)\beta$	$16b^5-16b^4+8b^3-\frac{3}{4}b+\frac{1}{2}$
FS	First cousins	$\frac{1}{4}$	$\frac{1}{2}(1-c)^2\beta+\frac{1}{8}c^2$	$32b^6-32b^5+18b^4-7b^3+\frac{17}{4}b^2-\frac{3}{2}b+\frac{9}{16}$
HS	Half cousins	$\frac{1}{8}$	$\frac{1}{4}(1-c)^2\beta$	$16b^6-16b^5+8b^4-2b^3+\frac{13}{4}b^2-\frac{3}{2}b+\frac{1}{2}$
Uncle–niece	Cousins once removed	$\frac{1}{8}$	$\frac{1}{4}(1-c)^3\beta+\frac{1}{16}(1-c)c^2$	$32b^7-32b^6+18b^5-7b^4+\frac{9}{4}b^3+\frac{5}{2}b^2-\frac{23}{16}b+\frac{1}{2}$

For the genome as a whole, letting l_i be the map length of chromosome i and $\sum_i l_i=L$, the variance is $\sum_i l_i^2\mathrm{Var}(\check{k}_1, l_i)/L^2$.

If X is the result of a parent-offspring (PO) mating or a full-sib (FS) mating, for example, $F_X=1/4$; but we show in Table 2 that the $\theta^*_{X;X}(c)$ values are different unless $c=0$ or $c=1/2$. This leads to different variances of the actual identities for half-sib progeny of X and pairs of their descendants:

$$\mathrm{Var}(\check{k}_1, l) = \begin{cases} \text{PO: } 16\phi_{t+5}(l)-16\phi_{t+4}(l) \\ \qquad +8\phi_{t+3}(l)-\frac{3}{4}\phi_{t+1}(l)+\frac{1}{2}\phi_t(l), \\ \text{FS: } 32\phi_{t+6}(l)-32\phi_{t+5}(l)+18\phi_{t+4}(l) \\ \qquad -7\phi_{t+3}(l)+\frac{17}{4}\phi_{t+2}(l)-\frac{3}{2}\phi_{t+1}(l)+\frac{9}{16}\phi_t(l). \end{cases} \tag{4}$$

The above results give the variance of $\check{k}_1$. As $Y1$ and $Y2$ and their descendants cannot share both genes at a locus (i.e. $k_2=0$), the variation in actual relationship $2\check{\theta}=\check{k}_2+\check{k}_1/2$ is given by $\mathrm{Var}(\check{k}_1, l)/4$ and in actual coancestry $\check{\theta}=\check{k}_2/2+\check{k}_1/4$ by $\mathrm{Var}(\check{k}_1, l)/16$.

4. Descendants of full-sibs

We now consider the case of matings between female $X1$ and male $X2$, unrelated to each other but with inbreeding coefficients F_{X1} and F_{X2}, respectively, and evaluate the variance in actual relationship among their full-sib progeny $Y1$ and $Y2$ and descendants of these such as first cousins.

Full-sibs can share 0, 1 or 2 alleles at each locus. As haplotypes are transmitted independently by the two parents, the variance in relationship among full-sibs is simply the sum of the components from paternal and maternal half-sibs with relevant inbreeding coefficients.

The actual state for $Y1$ and $Y2$ sharing pairs of alleles at each of two loci, i and j, is $\check{k}_{2i}\check{k}_{2j}=\check{k}_{1i}{}^m\check{k}_{1i}{}^p\check{k}_{1j}{}^m\check{k}_{1j}{}^p$ where m and p denote maternally and paternally derived alleles. Hence, from the definition of the two-locus coancestry,

$$\mathcal{E}(\check{k}_{2i}\,\check{k}_{2j})=\mathcal{E}(\check{k}_{1i}^m\,\check{k}_{1j}^m)\mathcal{E}(\check{k}_{1i}^p\,\check{k}_{1j}^p)=\theta^*_{X1;X1}(c)\theta^*_{X2;X2}(c),$$

which reduces to $\theta_{X1;X1}\theta_{X2;X2}=(1+F_{X1})(1+F_{X2})/4$ if $c=0, \beta=1/2$ and to the square of that if $c=1/2$, $\beta=1/4$. Evaluation depends on the pedigrees of $X1$ and $X2$, but is straightforward by expansion in terms of coefficients b as above and in Table 2.

The sharing of single copies among descendants of the full-sibs can be evaluated extending the methods for descendants of half-sibs. Suppose that parents $X1$, $X2$ have full-sib offspring $Y1$, $Y2$ and $Y2$ has offspring $Z2$. Then $Y1$ and $Z2$ are uncle and nephew and they can have only one ibd allele at each locus. Either $X1$ or $X2$ can transmit an entire haplotype to both $Y1$ and $Y2$ and the latter haplotype can be transmitted to $Z2$. This probability of the event is $[\theta^*_{X1;X1}(c)+\theta^*_{X2;X2}(c)](1-c)/2$ and it results in $Y1$ and $Z2$ sharing the haplotype. Alternatively, $X1$ can transmit ibd alleles at one locus and $X2$ can transmit ibd alleles at the other locus so $Y1, Y2$ share two pairs of ibd alleles: if $Y2$ then transmits these ibd alleles to $Z2$ then uncle and nephew again share ibd alleles at both loci. The probability of this event is $c\theta_{X1;X1}\theta_{X2;X2}$ so

$$\mathcal{E}(\check{k}_{1i}\,\check{k}_{1j})=\frac{1}{2}(1-c)[\theta^*_{X1;X1}(c)+\theta^*_{X2;X2}(c)]+c\theta_{X1;X1}\theta_{X2;X2}, \tag{5}$$

which reduces to $(\theta_{X1;X1}+\theta_{X2;X2})/2=(2+F_{X1}+F_{X2})/4$ if $c=0$, and to the square of that if $c=1/2$. For great uncle-great nephew and more distant uncle–nephew relationships, the probabilities are obtained as products of terms in eqn (5) by powers of $(1-c)/2$.

Similarly, for cousins $Z1, Z2$, the offspring of $Y1, Y2$ and the grand-offspring of $X1, X2$

$$\mathcal{E}(\check{k}_{1i}\,\check{k}_{1j})=\frac{1}{2}\left(\frac{1-c}{2}\right)^2\times[\theta^*_{X1;X1}(c)+\theta^*_{X2;X2}(c)] + \frac{1}{2}c^2\theta_{X1;X1}\theta_{X2;X2}. \tag{6}$$

Values for later descendants are obtained by scaling eqn (6), for example, by $(1-c)/2$ for cousins once removed and by $[(1-c)/2]^2$ for second cousins or cousins twice removed. All expressions for $\mathcal{E}(\check{k}_{1i}\check{k}_{1j})$ can be written as polynomials in b and evaluated accordingly.

5. Mapping functions, map length and physical genome length

In the analysis undertaken in this paper and in previous analyses of variation in genome sharing (HW11 and references cited therein) and indeed in other studies on other statistics such as distribution of lengths of shared regions (Stam, 1980; Donnelly, 1983), the Haldane mapping function (Haldane, 1919), $c=(1-e^{-2l})/2$, has been used. Not least, this is mathematically tractable, and explicit integration of the formulae-relating recombination fraction to map length is feasible, as in eqn (3). Haldane's function does not allow for interference, however, and various others have been constructed to incorporate interference. The importance of this assumption in variances of genome sharing, whether or not parents are inbred, has not been checked.

In mammalian studies, the Kosambi mapping function (Kosambi, 1944), $c=(1-e^{-4l})/[2(1+e^{-4l})]$ is most widely used, including in the published human linkage map (Matise *et al.*, 2007). For both functions $c\to l$ as $l\to 0$ and $c\to 0{\cdot}5$ as $l\to\infty$ but, for intermediate values of l, c is relatively larger for the Kosambi function: for example, for $l=0{\cdot}5$, where the absolute difference is near its maximum, for Haldane $c=0{\cdot}316$ and for Kosambi $c=0{\cdot}381$. To assess the dependence of the variation in genome sharing on the mapping function, numerical integration was used to evaluate Appendix eqn (B2), replacing the term $(1+e^{-2(x-y)})/4$ for $b=(1-c)/2$ by $[2-(1-e^{-4(x-y)})/(1+e^{-4(x-y)})]/4$. Numerical integration using bivariate Simpson's rule was used, and precision was checked by concurrent numerical integration of the Haldane function.

The variance of actual relationship is smaller with the Kosambi than the Haldane mapping function (Appendix C), as would be expected because the recombination fraction is, for given map length, larger with the former. The disparity increases the longer the chromosome, but it already differs little between l of 2 and 3M. Although the degree of relationship and type of relationship, for example, lineal or collateral, have some effect, it is rather small. Hence, as an approximate conclusion, the SD of relationship for l of 0·5, 1, 2 and 3M is about 4, 7, 10 and 11% smaller, respectively, with the Kosambi function incorporating interference (Appendix C). Although these are clear differences, the qualitative impact is rather small, and likely to be a little under 10% for the human genome as a whole.

Observations of genomic identity between chromosomes at the molecular level are initially likely to be in terms of the physical length, measured in Megabases not map lengths. Most or all calculations in this and other work on prediction of lengths of genome sharing are at the level of map distance. The conversion from one to the other then depends on the correspondence of the physical and linkage maps. This varies among chromosomes and species, around the typical mammalian figure of 1 cm/Mb, depending *inter alia* on positions of centromeres and repetitive regions, and the ratio of Morgans to Mb depends on chromosome length and differs among chromosomes; for example, the chicken has a very high M/Mb ratio relative to mammals and indeed relative to the zebra finch, but for both species of birds, the recombination rates on the microchromosomes are relatively high (Stapley *et al.*, 2008). For human chromosomes, although the linkage map is not far from linearly related to the physical map for the longer metacentric chromosomes, the relationship is somewhat sigmoidal; whereas for the shortest acrocentric chromosomes, no recombination are seen for over 25% of the centromeric end (Matise *et al.*, 2007 and http://compgen.rutgers.edu/RutgersMap/MapBrowser.aspx). Generalizations are therefore difficult, but it does imply a need to convert the initially observed lengths of shared regions into map distances before drawing inferences from analyses such as that presented here.

6. Discussion

The methodology given here fills a small lacuna in the analysis of variation in actual relationship, but to our knowledge has not been analysed previously. The formulae may be complicated, but the algorithms are easy to use.

As an example, consider the case of variance, expressed as SD, in actual relationship of half-sibs when the common parent of these sibs has undergone inbreeding (Fig. 3*a*) by one of several routes. The SD is not greatly reduced by the parental inbreeding, even in the case of selfing ($F=0{\cdot}5$), but the coefficient of variation CV (Fig. 3*b*) is reduced substantially more, because the expected relationship increases with F. The values differ only very slightly according to the mode of inbreeding for given F for example, by an offspring-parent compared with a full-sib mating.

Also consider comparisons between different levels of relationship according to the degree of inbreeding of the parent. For a single locus, or completely linked loci, from eqn (2) setting $c=0$ and hence $F(X)=F_X^*(c)$, $\mathrm{Var}(\check{k}_1,0)=(1/2)^{t+1}(1+F_X)[1-(1/2)^{t+1}(1+F_X)]$. Thus, for half-sib offspring ($t=0$), the variance is

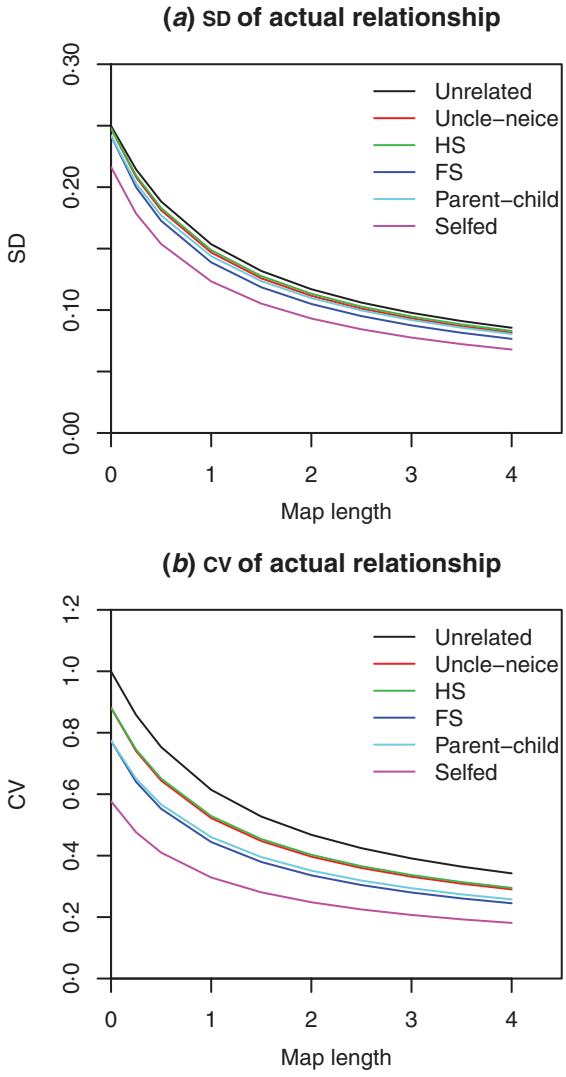

Fig. 3. (*a*) SD and (*b*) CV of actual relationship as a function of map length (*l*) of HS offspring of individuals whose parents were unrelated ($F=0$), or obtained by uncle-niece or HS mating ($F=1/8$), or obtained by FS or PO mating ($F=1/4$), or by selfing ($F=1/2$). Expected relationships for these values of F are 0·25, 0·281, 0·312 and 0·375, respectively.

highest when $F_X=0$; but for $t>0$, it is highest when $F_X=1$. Examples shown in Fig. 4 comparing variances for half-sibs and half-cousins as a function of map length and degree of inbreeding of the parent indicate that, as map length increases, the ranking of variances remains the same, i.e. decreasing with F_X for half-sibs and increasing with F_X for half-cousins. The (likely) explanation is that all half-sib offspring inherit a haplotype from their parent, which are therefore increasingly similar the more inbred is the parent. In contrast, a grandoffspring has a 50% chance of inheriting no haplotype from the inbred parent, and the similarity of these is more than outweighed by the divergence between the inbred and non-inbred parent.

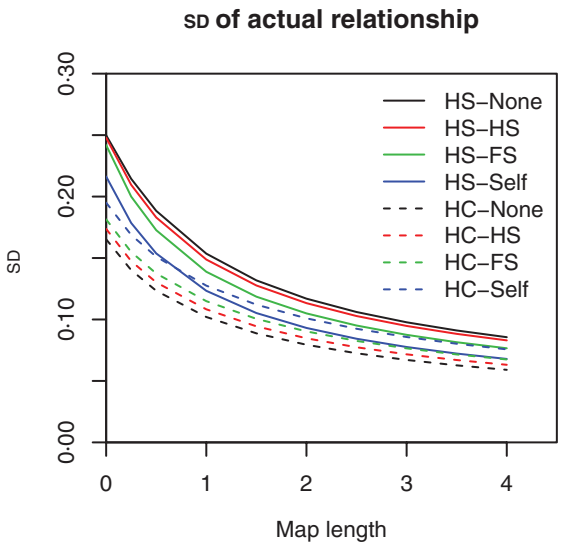

Fig. 4. SD of actual relationship as a function of map length (*l*) of HS offspring and half-cousin (HC) descendants of individuals whose parents were unrelated (none, $F=0$), or got by HS mating ($F=1/8$), or got by FS mating ($F=1/4$), or by selfing (self, $F=1/2$).

For offspring of full-sib matings, however, where for individual loci or no recombination $\mathrm{Var}(\check{k}_1, 0)\propto(2+F_{X1}+F_{X2})[1-(2+F_{X1}+F_{X2})/4]$ (from eqns (5) or (6)), the variance of relationship falls as inbreeding of either parent rises. This is as would be expected from the preceding argument on half-sibs because the grandoffspring must inherit from one or other grandparent.

This work was supported in part by NIH grant (GM 075091) and the Leverhulme Trust. The authors thank Ian White for helpful comments.

Appendix A. Derivation of two-locus descent measures

Weir & Cockerham (1969) presented a general algorithm for finding the probability of identity by descent for alleles a, a' and b, b' at loci **A** and **B**, respectively. Depending on whether these four alleles are transmitted on two gametes (ab from one individual $U1$ and $a'b'$ from another individual $U2$), or three gametes (ab from one individual $U1$, a' from a second individual $U2$ and b' from a third individual $U3$), or four gametes (a, b, a', b' from different individuals $U1$, $U2$, $U3$, $U4$) the probabilities are written as $\theta^*_{U1;U2}(c)$, $\gamma^*_{U1;U2,U3}(c)$ or $\delta^*_{U1,U2;U3,U4}(c)$, respectively. Calculation of any of these probabilities proceeds by tracing alleles back to founding individuals in a pedigree, taking recombination into account when necessary.

For individual X in Fig. 1, the offspring of half-sib parents, the two pairs of alleles $ab, a'b'$ are from individuals $V1$, $V2$ and may all have descended

from $U2$ so

$$\begin{aligned}F_X^*(c) &= \theta^*_{V1;V2}(c)\\ &= \left(\frac{1-c}{2}\right)^2[\theta^*_{U1;U2}(c)+\theta^*_{U1;U3}(c)+\theta^*_{U2;U2}(c)\\ &\quad+\theta^*_{U2;U3}(c)]+\left(\frac{1-c}{2}\right)\left(\frac{c}{2}\right)[\gamma^*_{U1;U2,\,U3}(c)\\ &\quad+\gamma^*_{U1;U3,\,U2}(c)+\gamma^*_{U2;U2,\,U3}(c)+\gamma^*_{U2;U3,\,U2}(c)\\ &\quad+\gamma^*_{U2;U1,\,U2}(c)+\gamma^*_{U2;U2,\,U1}(c)+\gamma^*_{U3;U1,\,U2}(c)\\ &\quad+\gamma^*_{U3;U2,\,U1}(c)]+\left(\frac{c}{2}\right)^2[\delta^*_{U1,\,U2;U2,\,U3}(c)\\ &\quad+\delta^*_{U1,\,U2;U3,\,U2}(c)+\delta^*_{U2,\,U1;U2,\,U3}(c)\\ &\quad+\delta^*_{U2,\,U1;U3,\,U2}(c)].\end{aligned}$$

Ignoring the terms that are zero (those with alleles at the same site coming from different ancestors), and using eqn (1) with $F_{U2}=F^*_{U2}(c)=0$

$$\begin{aligned}F_X^*(c) &= \left(\frac{1-c}{2}\right)^2\theta^*_{U2;U2}(c)\\ &= \frac{1}{4}\beta(1-c)^2,\end{aligned}$$

where $\beta=[(1-c)^2+c^2]/2$. Setting $c=0$ gives the one-locus result $F_X=1/8$, and setting $c=1/2$ gives the square of that.

For individual X in Fig. 2, the offspring of full-sib parents, the two pairs of alleles ab, $a'b'$ are from individuals $V1$, $V2$ and then from $U1$ and $U2$ so

$$\begin{aligned}F_X^*(c) &= \theta^*_{V1;V2}(c)\\ &= \left(\frac{1-c}{2}\right)^2[\theta^*_{U1;U1}(c)+\theta^*_{U1;U2}(c)+\theta^*_{U2;U1}(c)\\ &\quad+\theta^*_{U2;U2}(c)]+2\left(\frac{1-c}{2}\right)\left(\frac{c}{2}\right)[\gamma^*_{U1;U1,\,U2}(c)\\ &\quad+\gamma^*_{U1;U2,\,U1}(c)+\gamma^*_{U2;U1,\,U2}(c)+\gamma^*_{U2;U2,\,U1}(c)]\\ &\quad+\left(\frac{c}{2}\right)^2[\delta^*_{U1,\,U2;U1,\,U2}(c)+\delta^*_{U1,\,U2;U2,\,U1}(c)\\ &\quad+\delta^*_{U2,\,U1;U1,\,U2}(c)+\delta^*_{U2,\,U1;U2,\,U1}(c)].\end{aligned}$$

Ignoring the terms that are zero, and using eq (1) with $F_{U1}=F^*_{U1}(c)=F_{U2}=F^*_{U2}(c)=0$.

$$\begin{aligned}F_X^*(c) &= \left(\frac{1-c}{2}\right)^2[\theta^*_{U1;U1}(c)+\theta^*_{U2;U2}(c)]\\ &\quad+\left(\frac{c}{2}\right)^2[\delta^*_{U1,\,U2;U1,\,U2}(c)+\delta^*_{U2,\,U1;U2,\,U1}(c)]\\ &= \left(\frac{1-c}{2}\right)^2[(1-c)^2+(c)^2]+\left(\frac{c}{2}\right)^2 2\left(\frac{1}{2}\right)^2\\ &= \frac{1}{2}\beta(1-c)^2+\frac{1}{8}c^2.\end{aligned}$$

Setting $c=0$ gives the one-locus result $F_X=1/4$, and setting $c=1/2$ gives the square of that.

Appendix B. Evaluation of covariances (based on HW11)

Let $b=(1-c)/2$, the probability a pair of loci are jointly transmitted between generations, and express powers of c as polynomials in b:

$$c^n=\sum_{i=0}^{n}\binom{n}{i}(-2b)^i.$$

Writing $\theta^*_{X;X}(c)=\mathcal{E}(\check{k}_{i1}\check{k}_{1j})$ as a polynomial (examples in Table 2)

$$\mathcal{E}(\check{k}_{i1}\,\check{k}_{1j})=\sum_{n=0}^{N}\alpha_n b^n,$$

and noting that the covariance is zero for unlinked loci ($b=1/4$),

$$\mathrm{Cov}(\check{k}_{i1}\,,\,\check{k}_{1j})=\sum_{n=0}^{N}\alpha_n\left[b^n-\left(\frac{1}{4}\right)^n\right]. \tag{B1}$$

Assuming Haldane's mapping function, $b=(1-c)/2=(1+e^{-2d})/4$ where d is the map distance between the loci, so $b^n-(1/4)^n=(1/4)^n[(1+e^{-2d})^n-1]$. Integrating over all pairs of loci, we define

$$\phi_n(l)=\frac{2}{l^2}\left(\frac{1}{4}\right)^n\int_{x=0}^{l}\int_{y=0}^{x}[(1+e^{-2(x-y)})^n-1]\,dy\,dx. \tag{B2}$$

Integration of eqn (B2) gives eqn (3) in the text.

Appendix C. Effect of mapping function

Appendix Table C1 *SD of actual relationship computed using Kosambi mapping function divided by SD of actual relationship computed using the Haldane mapping function for different map lengths and pedigree relationships, including cases where the common ancestor is inbred.*

	Map length (M)			
Pedigree relationship*	0·5	1·0	2·0	3·0
GP–GO	0·972	0·931	0·884	0·865
GGGP–GGGO	0·965	0·928	0·892	0·878
G4P–G4O	0·960	0·929	0·902	0·892
HS	0·959	0·929	0·899	0·892
Half-cousins	0·962	0·929	0·898	0·885
HS, parent by selfing	0·958	0·926	0·906	0·899
HS, parent by PO mating	0·959	0·926	0·902	0·893

*P, parent; O, offspring; GP, grandparent; GGGP, great great grandparent.

References

Cornelis, M. C., Agrawal, A., Cole, J. W., Hansel, N. H., Barnes, K. C., Beaty, T. H., Bennett, S. N., Bierut, L. J.,

Boerwinkle, E., Doheny, K. F., Feenstra, B., Feingold, E., Fornage, M., Haiman, C. A., Harris, E. L., Hayes, M. G., Heit, J. A., Hu, F. B., Kang, J. H., Laurie, C. C., Ling, H., Teri, A., Manolio, T. A., Marazita, M. L., Mathias, R. A., Mirel, D. B., Paschall, J., Pasquale, L. R., Pugh, E. W., Rice, J. P., Udren, J., van Dam, R. M., Wang, X., Wiggs, J. L., Williams, K. & Yu, K. (2010). The gene, environment association studies consortium (GENEVA): maximizing the knowledge obtained from GWAS by collaboration across studies of multiple conditions. *Genetic Epidemiology* **34**, 364–372.

Donnelly, K. P. (1983). The probability that related individuals share some section of the genome identical by descent. *Theoretical Population Biology* **23**, 34–64.

Guo, S-W. (1995). Proportion of genome shared identical by descent by relatives: concept, computation, and applications. *American Journal of Human Genetics* **56**, 1468–1476.

Haldane, J. B. S. (1919). The combination of linkage values and the calculation of distances between the loci of linked factors. *Journal of Genetics* **8**, 299–309.

Hill, W. G. (1993*a*). Variation in genetic composition in backcrossing programs. *Journal of Heredity* **84**, 212–213.

Hill, W. G. (1993*b*). Variation in genetic identity within kinships. *Heredity* **71**, 652–653.

Hill, W. G. & Weir, B. S. (2011). Variation in actual relationship as a consequence of Mendelian sampling and linkage. *Genetics Research* **93**, 47–74.

Kosambi, D. D. (1944). The estimation of map distances from recombination values. *Annals of Eugenics* **12**, 172–175.

Laurie, C. C., Doheny, K. F., Mirel, D. B., Pugh, E. W., Bierut, L. J., Bhangale, T., Boehm, F., Caporaso, N. E., Edenberg, H. J., Gabriel, S. B., Harris, E. L., Hu, F. B., Jacobs, K. B., Kraft, P., Landi, M. T., Lumley, T., Manolio, T., McHugh, C., Painter, I., Paschall, J., Rice, J. P., Rice, K. M., Zheng, X. & Weir, B. S., for the GENEVA Investigators. (2010). Quality control and quality assurance in genotypic data for genome-wide association studies. *Genetic Epidemiology* **34**, 591–602.

Matise, T. C., Chen, F., Chen, W., De la Vega, F. M., Hansen, M., He, C., Hyland, F. C. L., Kennedy, G. C., Kong, X., Murray, S. S., Ziegle, J. S., Stewart, W. C. L., & Buyske, S. (2007). A second-generation combined linkage-physical map of the human genome. *Genome Research* **17**, 1783–1786.

Meuwissen, T. H. E., Hayes, B. J. & Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**, 1819–1829.

Stam, P. (1980). The distribution of the fraction of the genome identical by descent in finite populations. *Genetical Research* **35**, 131–155.

Stapley, J., Birkhead, T. R., Burke, T. & Slate, J. (2008). A linkage map of the zebra finch *Taeniopygia guttata* provides new insights into avian genome evolution. *Genetics* **179**, 651–667.

Visscher, P. M., Medland, S. E., Ferreira, M. A. R., Morley, K. I., Zhu, G., Cornes, B. K., Montgomery, G. W. & Martin, N. G. (2006). Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS Genetics* **2**, e41. doi:10.1371/journal.pgen.0020041

Visscher, P. M. (2009). Whole genome approaches to quantitative genetics. *Genetica* **136**, 351–358.

Weir, B. S., Anderson, A. D. & Hepler, A. B. (2006). Genetic relatedness analysis: modern data and new challenges. *Nature Reviews Genetics* **7**, 771–780.

Weir, B. S. & Cockerham, C. C. (1969). Pedigree mating with two linked loci. *Genetics* **61**, 923–940.