



Effect of heterogeneity in recombination rate on variation in realised relationship

Ian M. S. White¹ · William G. Hill¹

Received: 22 March 2019 / Revised: 21 May 2019 / Accepted: 27 May 2019 / Published online: 20 June 2019
© The Author(s), under exclusive licence to The Genetics Society 2019

Abstract

Individuals of a specified pedigree relationship vary in the proportion of the genome they share identical by descent, i.e. in their realised or actual relationship. Predictions of the variance in realised relationship have previously been based solely on the proportion of the map length shared, which requires the implicit assumption that both recombination rate and genetic information are uniformly distributed along the genome. This ignores the possible existence of recombination hotspots, and fails to distinguish between coding and non-coding sequences. In this paper, we therefore quantify the effects of heterogeneity in recombination rate at broad and fine-scale levels on the variation in realised relationship. Variance is usually greater on a chromosome with a non-uniform recombination rate than on a chromosome with the same map length and uniform recombination rate, especially if recombination rates are higher towards chromosome ends. Reductions in variance can also be obtained, however, and the overall pattern of change is quite complex. In general, local (fine-scale) variation in recombination rate, e.g. hotspots, has a small influence on the variance in realised relationship. Differences in rates across longer regions and between chromosome ends can increase or decrease the variance in a realised relationship, depending on the genomic architecture.

Introduction

Pedigree relationship (e.g. uncle–nephew) of diploid organisms determines the expected proportion of the genome that relatives share identical by descent (ibd). At any locus or site in the genome, there is a given probability that two relatives have a copy of the same parental locus. If all sites segregated independently, the variance in the actual proportion shared would be negligible. However, linkage produces a positive correlation of identity among sites and thereby reduces the effective number of segregating sites on each chromosome. As a result, individual pairs having the same pedigree relationship can vary substantially in their actual or realised relationship.

The magnitude of the variation depends on the closeness of the relationship, and more generally on the pedigree,

differing for example between that for offspring–grandparent and half-sib pairs, though the expected proportion of ibd sharing is 0.25 for both. As already noted, the variance also depends on the degree of linkage and therefore on the number and map lengths of chromosomes. While the variance of relationship becomes smaller the more distant the relationship, its coefficient of variation increases.

The variance in realised relationship is an important parameter which has to be taken into account when constructing pedigrees in natural or domesticated populations using data on genetic markers (Wang 2016) and in inferences about, for example, risks of inbreeding. It is also the basis of a method for estimating genetic variance free from confounding environmental effects (Visscher 2009; Odegaard and Meuwissen 2012).

Formulae and evaluations of the variance in actual or realised relationship have been developed over a number of years (e.g. Franklin 1977; Stam and Zeven 1981; Hill 1993; Guo 1995), and a comprehensive analysis for a broad range of relationships was given by Hill and Weir (2011). Visscher (2009) provides some approximate summary statistics for the human genome and describes how these can be utilised to estimate genetic variation in quantitative traits in non-pedigreed populations. Variation in realised inbreeding

✉ Ian M. S. White
i.m.s.white@ed.ac.uk

¹ Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Charlotte Auerbach Road, Edinburgh EH9 3FL, UK

of the offspring of related parents (Hill and Weir 2011) and variation in relationship among partially inbred individuals (Hill and Weir 2012) can be computed similarly.

In the previous work cited above, the assumption has been that the recombination rate is uniform along the genome. However, average recombination rates (expressed as centimorgans per megabase, cM/Mb) usually differ both among and within chromosomes. A typical mammalian value is 1 cM/Mb, depending inter alia on positions of centromeres and repetitive regions. The range of sex-averaged values among human chromosomes is from 0.96 to 2.11 cM/Mb (Kong et al. 2002), and greater variation is seen at finer scales of observation. For the longer human metacentric chromosomes, the relationship between linkage map and physical map is not far from linear, although it is somewhat sigmoidal. For the shortest acrocentric chromosome, over 25% of the centromeric end shows no recombination (Matise et al. 2007). Birds typically have many small chromosomes with high and diverse recombination rates (Groenen et al. 2009). For example, the chicken has a very high cM/Mb ratio compared with mammals, and recombination rates on the microchromosomes are particularly high (Stapley et al. 2017). Generally, there can be both broad-scale and small-scale differences within chromosomes, notably at recombination hotspots (McVean et al. 2004; Stapley et al. 2017).

Observations of genomic identity between chromosomes at the molecular level are initially likely to be in terms of physical length rather than map length. Here, we examine the effect on the variance of ibd sharing of measuring lengths of shared and unshared segments on the physical scale (Mb) rather than on the map scale (cM), extending some results of Hill and Weir (2011).

Materials and methods

First, we re-express some methods and results of Hill and Weir (2011) in a simpler form as a basis for further analysis. The pedigree relationships mentioned are described in Table 1.

Simplification and generalisation of basic formulae

We assume that the genome consists of a number of independently segregating chromosomes, and focus on just one of these.

We begin by considering a unilineal relationship, i.e. a pair of individuals who share exactly one pair of alleles ibd at any given locus with probability k , and otherwise share zero pairs of alleles, with probability $1-k$. Sharing derives from either the maternal or paternal lineage, but not both. The probability k corresponds to k_1 in Hill and Weir (2011).

Table 1 Summary of relationships

GPO	Grandparent and grand-offspring
G2PO	Great-grandparent and great grand-offspring, etc.
HS	Half-sibs
HUN	Half-uncle and nephew
HC	Half-cousins
HC1R	Half-cousins once removed ^a
UN	Uncle and nephew
FC	First cousins
GUGN	Great-uncle and great-nephew
C1R	First cousins once removed ^a
2C	Second cousins ^b
2C1R	Second cousins once removed ^{a,b}
Bilineal sharing	
FS	Full sibs
DHC	Double half-cousins ^c
DFC	Double first cousins ^c
HSFC	Half-sibs/first cousins ^d
HSHC	Half-sibs/half cousins ^d

^aThe relationship ‘X once removed’ is that between A and an offspring of B where X is the relationship between A and B

^bIf A and B are second cousins, one of the parents of A is the first cousin to one of the parents of B

^cFor DFC, the relationship is between the offspring of two brothers married to two sisters, or for DHC, between the offspring of two half-brothers married to two half-sisters

^dHSFC and HSHC denote the relationship between two offspring of the same father, whose mothers are sisters (for HSFC) or half-sisters (for HSHC)

Wright’s (1922) coefficient of relationship is then $k/2$, and $k/4$ is the coancestry coefficient.

For a range of different relationships, Hill and Weir (2011) calculate the probability of ibd sharing at two linked sites with recombination fraction c as a polynomial in $1-c$. Here, we follow the same approach, but express the probability as a polynomial in the linkage parameter $1-2c$, which leads to equivalent but simpler coefficients. Following Hill and Weir (2011), we assume that Haldane’s (1919) mapping function applies, so that the linkage parameter is $\exp(-2d)$, where $d > 0$ is the map distance between sites. Let I and J be 0/1 indicator variables for sharing at the two sites. Then $E(I)$ and $E(J)$ are both equal to k , and the probability of ibd sharing at both loci is

$$E(IJ) = k^2[a_0 + a_1 \exp(-2d) + a_2 \exp(-4d) + \dots], \quad (1)$$

where a_0, a_1 , etc. are the polynomial coefficients, all of which are positive or zero. Values of the coefficients for a range of relationships are given in Table 2. The limit $d \rightarrow \infty$, when the two sites segregate independently, shows that

Table 2 Coefficients of the variance of identity required for Eq. (1), for a range of relationships

	a_0	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	Total
GPO	1	1	0	0	0	0	0	0	0	2
HS	1	0	1	0	0	0	0	0	0	2
UN	1	0	0.5	0.5	0	0	0	0	0	2
G2PO	1	2	1	0	0	0	0	0	0	4
HUN	1	1	1	1	0	0	0	0	0	4
GUGN	1	1	0.5	1	0.5	0	0	0	0	4
FC	1	0	1.5	1	0.5	0	0	0	0	4
G3PO	1	3	3	1	0	0	0	0	0	8
HC	1	2	2	2	1	0	0	0	0	8
C1R	1	1	1.5	2.5	1.5	0.5	0	0	0	8
G4PO	1	4	6	4	1	0	0	0	0	16
HC1R	1	3	4	4	3	1	0	0	0	16
2C	1	2	2.5	4	4	2	0.5	0	0	16
G5PO	1	5	10	10	5	1	0	0	0	32
2C1R	1	3	4.5	6.5	8	6	2.5	0.5	0	32
Bilineal sharing										
FS	1	0	2	0	1	0	0	0	0	4
FSFC	1	0	2.5	1	2	1	0.5	0	0	8
SHHC	1	2	3	4	3	2	1	0	0	16
DFC	1	0	3	2	3.25	3	2.5	1	0.25	16
DHC	1	4	8	12	14	12	8	4	1	64

The scale factor is k^2 , where $k = 1/\text{total}$.

$a_0 = 1$. The case $d = 0$ (when the two sites segregate as one) shows that the coefficients sum to $1/k$.

The covariance of identity for two sites d Morgans apart $E(IJ) - E(I)E(J)$ is obtained by subtracting k^2 from expression (1):

$$k^2[a_1 \exp(-2d) + a_2 \exp(-4d) + \dots] \quad (2)$$

For example, for great-uncle and great-nephew, the covariance is

$$(1/16)[\exp(-2d) + 0.5\exp(-4d) + \exp(-6d) + 0.5\exp(-8d)].$$

Expression (2) shows how, for a given relationship, covariance is attenuated by recombination as the map distance between sites increases. The size and number of coefficients are related to the number and complexity of coancestry paths linking the two relatives (Wright 1922).

Variation as a function of physical length

In order to compute the variance in relationship as a function of physical length, we weight each DNA base equally. Hence, let X and Y denote physical positions on the chromosome, measured as proportions of its total physical

length, and let $\psi(X)$ and $\psi(Y)$ denote the corresponding map positions, expressed as proportions of the total map length ℓ . We refer to ψ as the ‘Marey’ map, relating map length to physical length (Chakravarti 1991). The derivative of ψ is the local recombination rate, i.e. the rate at which the proportion of map length increases with increasing proportion of physical length. Scaling by the ratio of map length of the chromosome in cM to physical length of the chromosome in Mb produces the conventional units cM/Mb. Following Hill and Weir (2011), we obtain the variance of ibd for a chromosome by averaging over all pairs of loci. Specifically, in the s^{th} term of expression (2), we replace d by $\ell(\psi(X) - \psi(Y))$ and integrate $\exp[-2s\ell(\psi(X) - \psi(Y))]$ over the triangular region $0 < Y < X < 1$ to obtain the average

$$\varphi(s\ell) = 2 \iint \exp[-2s\ell(\psi(X) - \psi(Y))] dY dX, \quad (3)$$

where we have assumed that the number of genomic sites is large enough that the sum of terms can be approximated by Eq. (3). Then, from Eq. (2), variance of ibd sharing on the chromosome is

$$v(\ell) = k^2[a_1 \varphi(\ell) + a_2 \varphi(2\ell) + \dots] \quad (4)$$

For the grandparent–grand-offspring relationship (GPO), the variance is $k^2 \varphi(\ell)$, and Eq. (4) extends the result to other relationships, taking coefficients a_1, a_2 , etc. from Table 2. For half-sibs (HS), for example, the variance is $k^2 \varphi(2\ell)$, so that the variance of sharing is the same for a HS relationship with map length ℓ and a GPO relationship with map length 2ℓ .

For the special case of uniform recombination rate (indicated by suffix *),

$$\varphi_*(s\ell) = (2s\ell - 1 + \exp(-2s\ell)) / (2s^2 \ell^2) \quad (5)$$

(Hill and Weir 2011). The corresponding variance of relationship is

$$v_*(\ell) = k^2[a_1 \varphi_*(\ell) + a_2 \varphi_*(2\ell) + \dots]$$

and the ratio

$$\frac{v(\ell)}{v_*(\ell)} = \frac{a_1 \varphi(\ell) + a_2 \varphi(2\ell) + \dots}{a_1 \varphi_*(\ell) + a_2 \varphi_*(2\ell) + \dots} \quad (6)$$

is a measure of the effect of non-linearity in the Marey map, which we refer to subsequently as the discrepancy ratio, with values far from unity indicating a strong effect.

For bilinear relationships (e.g. full-sibs, double first cousins), ibd sharing at a given site is possible through both

paternal and maternal lineages, and the overall proportion of ibd sharing is the average of two values, each of which derives from a unilineal relationship (maternal or paternal). Because the lineages segregate independently, the covariance of identity is the sum of two terms like Eq. (2), one for each lineage. In the simplest cases, e.g. FS and DFC, when the chromosomal relationship is the same in both lineages, the bilinear variance is simply twice that for the corresponding unilineal relationship. Variation in ‘double’ sharing (simultaneous sharing in both lineages) can be calculated in a similar way to variance in single sharing, based on products of terms like Eq. (1) for each lineage (Table 2).

Broken-stick models

In practice, physical and map positions will be measured at a discrete set of markers. We assume that these are closely spaced, and that recombination rate is uniform between markers. The Marey map then takes the form of a continuous, piecewise linear function, resembling a broken stick. The variation in physical distance shared can then be calculated by summation, rather than numerical integration.

Therefore, we consider a single chromosome divided into segments, within each of which the recombination rate is constant. Denote the map length of the i^{th} segment by ℓ_i and its physical length, expressed as a proportion of the total physical length of the chromosome, by p_i , with $\sum p_i = 1$. First, we consider the case of grandparent and grand-offspring, for which the only non-zero coefficient in Eq. (4) is $a_1 = 1$ (Table 2).

Let S_i be the proportion of physical (and map) length shared in the i^{th} segment. The calculation of $\text{var}(S_i)$ is the same as that for a whole chromosome with constant recombination rate: from Eq. (5)

$$\text{var}(S_i) = k^2(2\ell_i - 1 + \exp(-2\ell_i))/(2\ell_i^2), \quad (7)$$

and the covariance of sharing between two segments is

$$\text{cov}(S_i, S_j) = k^2 c_i c_j \exp(-2\ell_{ij})/(\ell_i \ell_j), \quad (8)$$

where $c_i = (1 - \exp(-2\ell_i))/2$ is the recombination probability for the i^{th} segment, and ℓ_{ij} is the total map length of segments lying between the i^{th} and j^{th} segments, with $\ell_{ij} = 0$ when they are adjacent.

For the whole chromosome, the proportion of physical length shared $\sum p_i S_i$ has variance

$$\sum p_i^2 \text{var}(S_i) + 2 \sum_{i < j} p_i p_j \text{cov}(S_i, S_j). \quad (9)$$

For other relationships in Table 2, the variance is calculated from Eq. (9), with Eq. (7) modified to

$$\text{var}(S_i) = k^2 \sum_s a_s (2s\ell_i - 1 + \exp(-2s\ell_i))/(2s^2 \ell_i^2),$$

and with a similar modification to Eq. (8).

Application

We investigated variance of identity for two very different genomes, human and chicken. For human data, we used web table E from Kong et al. (2002), which provided physical and map distances (separately for males and females) for segments between 5136 markers on the 22 autosomes. For the chicken genome, we used physical and map distances for segments between the 9268 markers in the published linkage map for chromosomes 1–28 (Groenen et al. 2009, Supplemental Table 1). We omitted chromosomes 16, 22 and 25, for which we did not have sufficient data to produce reliable estimates. For each data set, theoretical variances for a GPO relationship were calculated on map and physical scales, using a broken-stick model for the latter, and the results combined to produce the variance for the genome-wide proportion of ibd sharing on each scale. This was calculated as $\sum p_i^2 v_i / (\sum p_i)^2$, where p_i was the length and v_i the variance of sharing for the i^{th} chromosome, with v_i and p_i calculated on the same scale (map or physical).

The effect of interference

Our results are based on Haldane’s mapping function, which assumes no interference in the crossover process. We briefly consider the effect of replacing the Haldane mapping function by Kosambi’s (1944) mapping function, which allows for interference.

Results

An illustrative example

For illustration, we consider the three Marey maps shown in Fig. 1. In one (afferent, ‘bringing inwards’), recombination is concentrated at the centre of the chromosome, and in another (efferent, ‘conveying outwards’), there are high recombination rates near the chromosome ends. Midway between these two extremes, the Marey map is linear, corresponding to a constant recombination rate.

The chromosome of Fig. 1 comprises two segments of equal physical length. Let S_1 and S_2 be the proportions

shared in each segment, so that the proportion of the whole chromosome shared is the average of S_1 and S_2 , with variance $V(1 + \rho)/2$, where $V = \text{var}(S_1) = \text{var}(S_2)$ is the variance within each segment, and ρ is the correlation between S_1 and S_2 . For the GPO relationship, the effect of changing the Marey map from efferent (recombination mainly at ends) to afferent (recombination mainly at the centre) is that the segment variance changes very little, while the correlation steadily diminishes (Table 3).

In this example, variance of sharing for the whole chromosome is largely determined by the correlation between the amounts of sharing in the two segments, which in turn is dependent on the degree to which recombination is concentrated at the junction between the segments. A low recombination rate at the centre creates high correlation,

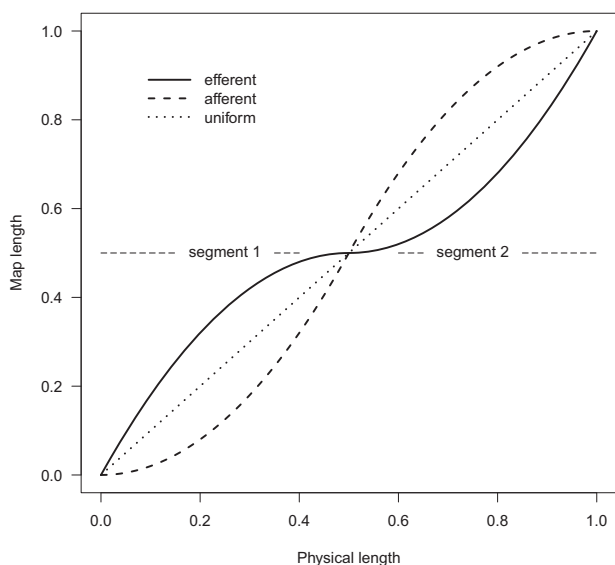


Fig. 1 Three examples of Marey maps. Recombination is distributed away from the chromosome centre for the ‘efferent’ map, and concentrated at the centre for the ‘afferent’ map. The horizontal dashed lines indicate the two segments mentioned in the text

Table 3 Variance (V) of physical length shared within two half-chromosomes, and the correlation (ρ) between sharing in one and in the other, for a range of Marey maps (the three examples of Fig. 1 and two intermediate cases)

Marey map	V	ρ	$V(1 + \rho)/2$
Strongly efferent	0.739	0.755	0.649
Weakly efferent	0.737	0.639	0.604
Uniform	0.736	0.543	0.568
Weakly afferent	0.737	0.461	0.538
Strongly afferent	0.739	0.392	0.514

Each segment has map length 0.5 M. Variance of sharing for the whole chromosome is $V(1 + \rho)/2$. All values are scaled by the variance of shared map length for a 1 M chromosome. The assumed relationship is GPO.

whereas a high recombination rate there reduces the correlation.

Recombination hotspots

Consider a chromosome with a uniform recombination rate, except for one or more idealised recombination hotspots, each with finite map length but assumed zero physical length. The hotspots divide the chromosome into segments, within each of which the recombination rate is assumed to be constant. The standard broken-stick calculation applies, except that in Eq. (8), the exponential terms for intervening segments include any hotspots therein.

When considering the effect of a hotspot, the baseline for comparison can be either the result of removing the hotspot, leaving just the uniform background, or the result of merging the hotspot map length uniformly with the background. With the first approach, the effect of a single hotspot is always a reduction in variance, with maximum effect when centrally situated. With the second approach, which we consider more appropriate and therefore use below, the effect can be either an increase or decrease in variance, and the maximum effect could be either at a central position or at a chromosome end.

Figure 2 shows the effect of two hotspots at different positions on the chromosome. The upper curve has an

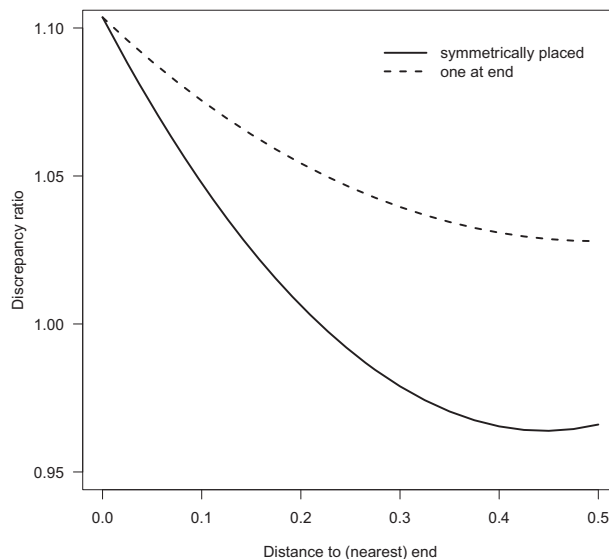


Fig. 2 Effect on the discrepancy ratio (Eq. 6) of different positions of two hotspots on the chromosome. A background of uniform recombination rate is assumed. The relationship is GPO, chromosome map length is 1 M and the map length of each hotspot is 0.1 M. Upper curve: one hotspot is placed at a chromosome end. Lower curve: hotspots are symmetrically placed about the chromosome centre. The baseline for comparison is the uniform case (i.e. hotspot map lengths merged uniformly with the background). The upper curve can also be viewed as the effect of a single 0.1 M hotspot on a 0.8 M uniform background

alternative interpretation as the effect of a single hotspot. In each case, the GPO relationship is assumed. The variance obtained with a single central hotspot is generally less than if the recombination rate is constant along the chromosome. Thus, a single hotspot reduces variance when it is centrally placed, and increases variance when it is near a chromosome end. In the central position, the hotspot breaks up the positive correlation between adjacent segments in a similar way to the afferent map of Fig. 1. The effect (increase or decrease) is generally small, however. The maximum effect of two hotspots occurs when both hotspots are at a chromosome end (either with one at each end or with both at the same end).

Further analysis shows that similar results to those illustrated in Fig. 2 are obtained with a wide range of relationships, map lengths and hotspot intensities. Generally, the effect on ibd variance is small, unless hotspot map lengths are large compared with the background and the hotspots are positioned near chromosome ends.

We now consider briefly the more realistic situation, where numerous hotspots are located throughout the chromosome, with varying intensities (crossover probabilities). In this case, the Marey map consists of a mixture of straight line segments (between hotspots) and vertical jumps (at the hotspots). Fixing the total map length of the chromosome constrains the individual jumps to be small. As a result, the Marey map, viewed on a broad scale, is reasonably smooth and only departs from a straight line if there is a systematic trend in either hotspot or background intensity. In the absence of such broad-scale changes, the Marey map will be nearly linear and the discrepancy ratio will be close to 1. Thus, the effect of multiple hotspots changes from fine scale to broad scale as the number of hotspots increases.

Broad-scale variation in recombination rate

A broken-stick model with three segments of equal physical length, but different map lengths, was found sufficiently flexible to describe many types of broad-scale variation in recombination rate.

For the GPO relationship, and a chromosome with map length $2M$, Fig. 3 shows the effect on the discrepancy ratio of different ways of dividing the total map length between the outside segments and the middle segment, (i) when the outside segments have equal map length, and (ii) when one of the two outside segments has zero map length. The maximum disparity occurs when all recombination is concentrated in one end segment. This is in agreement with our results for hotspots, but now the magnitude of the effect is much greater.

The results for other relationships and map lengths are expressed in terms of the three extreme points labelled **A**, **B** and **C** in Fig. 3. For different map lengths and relationships,

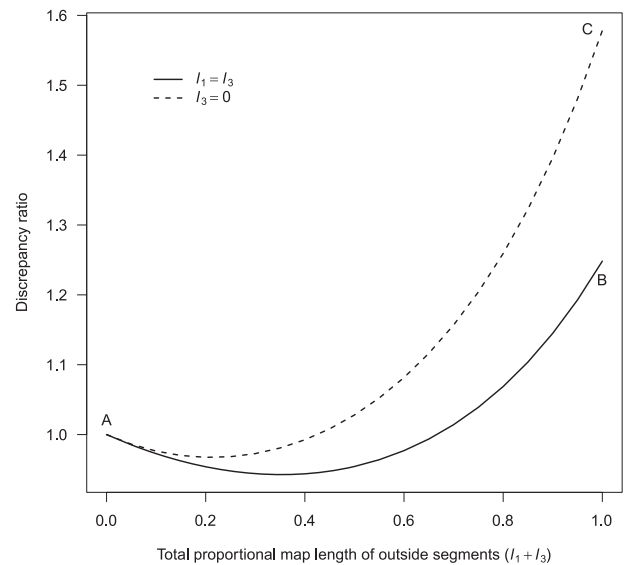


Fig. 3 With l_1 , l_2 , and l_3 , the proportional map lengths of the left, middle and right segments of a three-segment broken stick ($l_1 + l_2 + l_3 = 1$), the figure shows the effect on the discrepancy ratio (Eq. 6) of changing $l_1 + l_3$, and of different ways of partitioning $l_1 + l_3$ between the outside segments. GPO relationship, map length $2M$. Lower curve: the effect of changing $l_1 + l_3$ while $l_1 = l_3$; upper curve, the same, but with no recombination in one of the two outside segments (say $l_3 = 0$). At point **A**, all recombination is confined to the middle segment, with end segments either completely sharing or completely non-sharing. At point **B**, there is no recombination in the middle segment, with map length equally divided between the end segments. At point **C**, all recombination is confined to one of the end segments

A, **B**, and **C** change position, but the area bounded by the three extreme points retains the same basic shape (Fig. 4). The largest discrepancy ratio (at **C**) increases as the relationship becomes more remote. Changes in **A**, **B** and particularly **C** become more pronounced as map length is increased. Overall, the potential for an increase in variance is greatest with more remote relationships and longer chromosome lengths. Similar trends with map length and bilinear relationship are obtained in the variance of double sharing (Fig. 5).

Contrasting results for human and chicken genomes

Figure 6 shows discrepancy ratios for the GPO relationship, separately for chickens, human males and human females. With the human data, both male and female, an 'efferent' pattern for every chromosome consistently produced discrepancy ratios >1 . In females, the discrepancy ratio was almost constant, never rising above 1.1, but there was a clear positive correlation with chromosome length. In males, variances were greater than those for females on both physical and map scales. Discrepancy ratios were also higher, but with no discernible relationship with chromosome length, apart from relatively small values for the two

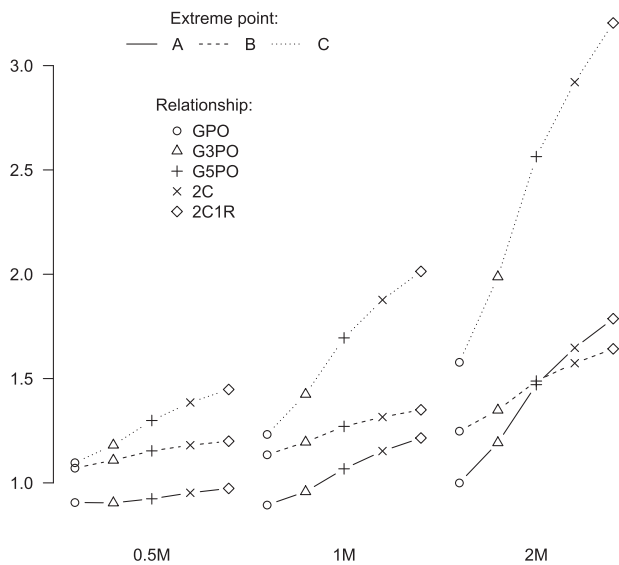


Fig. 4 Discrepancy ratios (Eq. 6) for Marey maps represented by points **A**, **B** and **C** of Fig. 3, for map lengths 0.5, 1 and 2 M, and a subset of unilinear relationships from Table 1

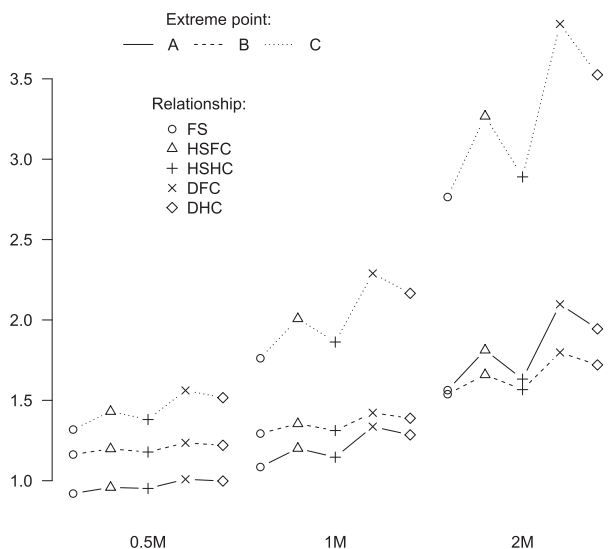


Fig. 5 Discrepancy ratios (Eq. 6) for double sharing for Marey maps represented by points **A**, **B**, and **C** of Fig. 3, using map lengths 0.5, 1 and 2 M, and the bilinear relationships of Table 1

shortest chromosomes. For the chicken genome, there was a strong positive correlation between the discrepancy ratio and chromosome length, with values exceeding 1.0 for the longer chromosomes, and <1.0 for some of the shorter chromosomes. Similar results were obtained with more remote relationships.

For the 25 chicken chromosomes, although the discrepancy ratio for individual chromosomes never exceeded a 15% increase (Fig. 6), the variance of the genome-wide proportion of physical length shared was 41% greater than that on the

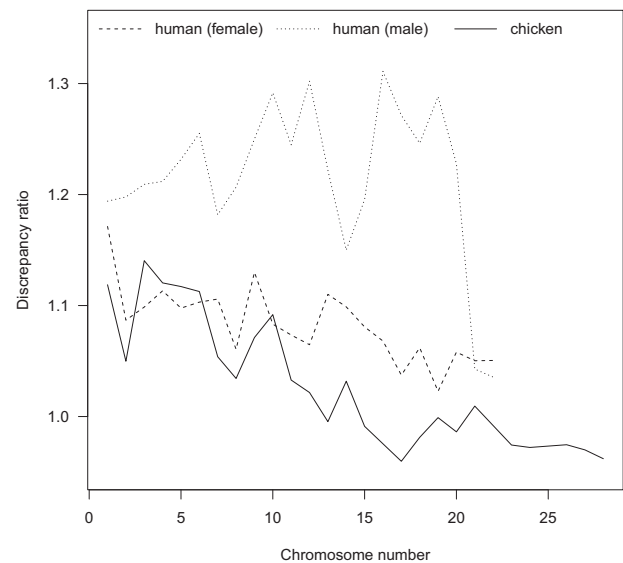


Fig. 6 Discrepancy ratios (Eq. 6) for human chromosomes 1–22 and chicken chromosomes 1–15, 17–21, 23, 24 and 26–28. Assumed relationship GPO

map scale. The corresponding genome-wide values for the 22 human autosomes were 16% (females) and 33% (males).

Allowing for interference

The linkage parameter associated with a segment of map length ℓ is $\exp(-2\ell)$ according to Haldane's function and $1-\tanh(2\ell)$ according to Kosambi's function. Plotting these curves shows that the linkage parameter for the Kosambi map diminishes to zero more rapidly than that for the Haldane map, such that the Kosambi linkage parameter at map length ℓ is about the same as the Haldane linkage parameter evaluated at 1.3ℓ . Using this approximation, we estimate that the effect on standard deviation of ibd sharing of changing from the Haldane to the Kosambi map is a reduction of 4%, 7%, 9% and 10% for map lengths of 0.5, 1, 2 and 3 M (based on a constant recombination rate). These values are in close agreement with those found by Hill and Weir (2012). For the human genome, Caballero et al. (2019) estimated an average 11% reduction in SD of sharing due to interference over a range of relationships from full-sibs to second cousins. For the case of non-uniform recombination rate, inspection of Figs. 4 and 5 shows that increasing map length by 30% would cause a small increase in the discrepancy ratio under scenarios **A** and **B**, and a large increase with remote relationships under scenario **C**.

Discussion

Variation in recombination rate within and between chromosomes is a well-recognised feature of the genome.

Stapley et al. (2017) reviewed contributing factors, and Ritz et al. (2017) discussed the extent to which variation in recombination rate is itself adaptive. Since the *PRDM9* gene was identified, analysis and discussion have been published on its evolution, its role in determining sequence-wide hotspots and (possibly) speciation events (Schwartz et al. 2014). Factors that determine observed chromosome numbers, average recombination rates and distribution over chromosomes have been discussed in the evolutionary literature. It is likely that there are many factors involved, including, for example, epistasis and reproductive isolation. Charlesworth and Charlesworth (2010, pp. 546–561) provide an extensive discussion.

The chicken genome provides an example of substantial heterogeneity in recombination rate both between and within chromosomes (Groenen et al. 2009). The longest chromosomes have a nearly constant rate, except for elevated values at the chromosome ends, where the influence of the higher recombination rate is likely to be small, because a small amount of genome is involved. The shorter chromosomes tend to have higher mean recombination rates than do the long chromosomes and contribute rather little to the overall variance in relationship.

More generally, we have shown that, when recombination rate varies along the chromosome, variation in relationship as measured in terms of the physical length of the genome can differ quite substantially from that measured in map units. The effect is usually an increase in variance, although a reduction is also possible. Among the main influences on the difference are recombination hotspots and cases where much of the recombination generated by meiosis is found towards one chromosome end. The discrepancy ratio is similar over a wide range of relationships, but tends to increase as relationships become more distant (and variances decrease). The effect of the degree of relationship on the discrepancy ratio is the same order of magnitude as the effect of map length or non-linearity in the Marey map. For example, the difference in discrepancy ratio between GPO and 2C1R is similar to the difference between map lengths of 0.5 and 2 M, or the difference between having recombination concentrated at the chromosome centre (scenario A in Fig. 4) or at one chromosome end (scenario C).

We find that the pattern of large-scale variation in recombination rate is more important than small-scale variation in its effect on variance of sharing. For this reason, the broken-stick calculation, which works well with high-density linkage map data, should also give reasonable results even with markers which are more sparsely distributed. However, it is not appropriate for the type of cytological data that has been used to define crossover distributions (e.g. based on MLH1 foci).

Our observations raise the question of which is the more appropriate quantity to analyse, physical or map distance.

Map distance does not necessarily reflect parts of the genome with high gene density but low recombination rate, for example. In some applications, map distance may be the appropriate quantity, because it is potentially the best indicator of local linkage disequilibrium. This is relevant when using genomic data to predict genetic merit in livestock and crops (Meuwissen et al. 2001) or complex human diseases (Maier et al. 2018). Here, the issue is how much information about the genotype at trait loci comes from neighbouring markers, and marker linkage is likely to be the most suitable quantity with which to weight observations.

Although in theory the differences between physical and map distances can have a substantial impact on variance of sharing, in practice, the impact seems to be limited. Compared with the vast range of recombination rates among different orders of animals and even species, the effects on variation in relationship are relatively small. Only in rather specific circumstances (e.g. distant relationship combined with a long chromosome in which most of the recombination occurs near one end) is the discrepancy ratio likely to be large enough to be of practical importance.

Data availability

The chicken genomic data (Groenen et al. 2009, Supp. Table S1) is available at <https://genome.cshlp.org/content/19/3/510/suppl/DC1>.

Acknowledgements The authors thank the referees for their comments and suggestions, and USS for financial support during the course of this work.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Caballero M, Seidman DN, Dyer TD, Lehmann DM, Curran JE, Duggirala R, et al. (2019) Surprising impacts of crossover interference and sex-specific genetic maps on identical by descent distributions. <https://www.biorxiv.org/content/10.1101/527655v1>
- Chakravarti A (1991) A graphical representation of genetic and physical maps: the Marey map. *Genomics* 11:219–222
- Charlesworth B, Charlesworth D (2010) *Elements of Evolutionary Genetics*. Robertson & Co., Colorado
- Franklin IR (1977) The distribution of the proportion of the genome which is homozygous by descent in inbred individuals. *Theor Popul Biol* 11:60–80
- Groenen MAM, Wahlberg P, Foglio M, Cheng HH, Megens H-J et al. (2009) A high-density SNP-based linkage map of the chicken

- genome reveals sequence features correlated with recombination rate. *Genome Res* 19:510–519
- Guo S-W (1995) Proportion of genome shared identical by descent by relatives: concept, computation, and applications. *Am J Hum Genet* 56:1468–1476
- Haldane JBS (1919) The combination of linkage values, and the calculation of distance between linked factors. *J Genet* 8:299–309
- Hill WG (1993) Variation in genetic composition in backcrossing programs. *J Hered* 84(3):212–223
- Hill WG, Weir BS (2011) Variation in actual relationship as a consequence of Mendelian sampling and linkage. *Genet Res* 93:47–74
- Hill WG, Weir BS (2012) Variation in actual relationship among descendants of inbred individuals. *Genet Res* 94:267–274
- Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B et al. (2002) A high-resolution recombination map of the human genome. *Nat Genet* 31:241–247
- Kosambi DD (1944) The estimation of map distances from recombination values. *Ann Eugen* 12:172–175
- Maier RM, Zhu Z, Lee SH, Trzaskowski M, Ruderfor DM, Stahl EA et al. (2018) Improving genetic prediction by leveraging genetic correlations among human diseases and traits. *Nat Commun* 9:989. <https://doi.org/10.1038/s41467-017-02769-6>
- Matise TC, Chen F, Chen W, De La Vega FM, Hansen M et al. (2007) A second-generation combined linkage-physical map of the human genome. *Genome Res* 17(12):1783–1786
- McVean GAT, Myers SR, Hunt S, Deloukas P, Bentley DR et al. (2004) The fine-scale structure of recombination rate variation in the human genome. *Science* 304:581–584
- Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829
- Odegaard J, Meuwissen THE (2012) Estimation of heritability from limited family data using genome-wide identity-by-descent sharing. *Genet Sel Evol* 44:16
- Ritz KR, Noor MAF, Singh ND (2017) Variation in recombination rate: adaptive or not? *Trends Genet* 33:364–374
- Schwartz JJ, Roach DJ, Thomas JH, Shendure J (2014) Primate evolution of the recombination regulator *PRDM9*. *Nat Comm* 5:4370
- Stam P, Zeven AC (1981) The theoretical proportion of the genome identical by descent in finite populations. *Genet Res* 35:131–155
- Stapley J, Feulner PGD, Johnston SE, Santure AW, Smadja CM (2017) Variation in recombination frequency and distribution across eukaryotes: patterns and processes. *Philos Trans R Soc B* 372:20160455
- Visscher PM (2009) Whole genome approaches to quantitative genetics. *Genetica* 136:351–358
- Wang J (2016) Pedigree or markers: which are better in estimating relatedness and inbreeding coefficient? *Theor Popul Biol* 107:4–13
- Wright S (1922) Coefficients of inbreeding and relationship. *Am Nat* 56:330–338