Variation in genetic relatedness is determined by the aggregate recombination process

Supporting Material

Carl Veller[*,1], **Nathaniel B. Edelman**[*], **Pavitra Muralidhar**[*] **and Martin A. Nowak**[*,†]
[*]Department of Organismic and Evolutionary Biology, Harvard University, [†]Department of Mathematics, Harvard University

---

[1]Corresponding author: 2320 Storer Hall, Department of Evolution and Ecology, University of California, Davis, One Shields Ave., Davis, CA 95616. E-mail: carl.veller@gmail.com

S1. General case for direct descent

Label the starting generation 0, so that the offspring generation is 1, the grand-offspring generation 2, etc. For an individual in generation 0, one of its descendants in generation t, and a locus k, let $P_k^{(t)}$ be a random variable that takes the value 1 if an allele carried by the generation-t descendant at locus k was inherited from the generation-0 individual, and takes the value 0 otherwise. When segregation follows Mendel's first law (as we assume throughout),

$$\text{Prob}\left(P_k^{(t)}=1\right)=\frac{1}{2^{t-1}} \tag{S.1}$$

and because $P_k^{(t)}$ takes only the values 0 or 1, $\mathbb{E}[P_k^{(t)}]=\text{Prob}\left(P_k^{(t)}=1\right)=1/2^{t-1}$. Then

$$\text{Var}\left(P_k^{(t)}\right)=\mathbb{E}\left[\left(P_k^{(t)}\right)^2\right]-\left(\mathbb{E}\left[P_k^{(t)}\right]\right)^2=\text{Prob}\left(P_k^{(t)}=1\right)-\left(\frac{1}{2^{t-1}}\right)^2=\frac{1}{2^{t-1}}-\left(\frac{1}{2^{t-1}}\right)^2. \tag{S.2}$$

Case 1: No sex differences in recombination. Consider two loci, i and j. For the alleles at these loci in the generation-t descendant both to have been inherited from the generation-0 ancestor requires that the loci were not recombinant in any of the gametes, from the gamete produced by the generation-1 descendant through to that produced by the generation-$[t-1]$ descendant, that link the generation-t descendant and the specified generation-0 ancestor (probability $1-r_{ij}$ for each of the $t-1$ relevant gametes) and, conditional on this, that the appropriate alleles co-segregated to the gamete in each relevant meiosis (probability $1/2$ each time). Therefore,

$$\text{Prob}\left(P_i^{(t)}=P_j^{(t)}=1\right)=\frac{1}{2^{t-1}}(1-r_{ij})^{t-1}, \tag{S.3}$$

so that

$$\begin{aligned}\text{Cov}\left(P_i^{(t)},P_j^{(t)}\right)&=\mathbb{E}\left[P_i^{(t)}P_j^{(t)}\right]-\mathbb{E}\left[P_i^{(t)}\right]\mathbb{E}\left[P_j^{(t)}\right]\\&=\text{Prob}\left(P_i^{(t)}=P_j^{(t)}=1\right)-\left(\frac{1}{2^{t-1}}\right)^2\\&=\frac{1}{2^{t-1}}(1-r_{ij})^{t-1}-\left(\frac{1}{2^{t-1}}\right)^2\\&=\frac{1}{2^{t-1}}\left((1-r_{ij})^{t-1}-\frac{1}{2^{t-1}}\right).\end{aligned} \tag{S.4}$$

Assume that there are L loci in total, with L very large, and let $P^{(t)}$ be the proportion of the generation-t descendant's genome inherited from the generation-0 ancestor: $P^{(t)}=\frac{1}{L}\sum_{k=1}^{L}P_k^{(t)}$. Then

$$\mathbb{E}[P^{(t)}]=\frac{1}{L}\sum_{k=1}^{L}\mathbb{E}[P_k^{(t)}]=\frac{1}{L}\sum_{k=1}^{L}\frac{1}{2^{t-1}}=\frac{1}{2^{t-1}}, \tag{S.5}$$

and

$$\begin{aligned}\text{Var}\left(P^{(t)}\right)&=\text{Var}\left(\frac{1}{L}\sum_{k=1}^{L}P_k^{(t)}\right)=\frac{1}{L^2}\sum_{k=1}^{L}\text{Var}\left(P_k^{(t)}\right)+\frac{1}{L^2}\sum_{i\neq j}\text{Cov}\left(P_i^{(t)},P_j^{(t)}\right)\\&=\frac{1}{L^2}L\left(\frac{1}{2^{t-1}}-\left(\frac{1}{2^{t-1}}\right)^2\right)+\frac{1}{L^2}\sum_{i\neq j}\frac{1}{2^{t-1}}\left((1-r_{ij})^{t-1}-\frac{1}{2^{t-1}}\right)\\&=\frac{1}{L}\left(\frac{1}{2^{t-1}}-\left(\frac{1}{2^{t-1}}\right)^2\right)+\frac{L(L-1)}{L^2}\frac{1}{L(L-1)}\sum_{i\neq j}\frac{1}{2^{t-1}}\left((1-r_{ij})^{t-1}-\frac{1}{2^{t-1}}\right)\\&\xrightarrow[L\to\infty]{}\frac{1}{2^{t-1}}\left(\overline{(1-r)^{t-1}}-\frac{1}{2^{t-1}}\right) \qquad \text{(S.6)}\\&=\frac{1}{2^{t-1}}\left(1-\frac{1}{2^{t-1}}+\sum_{\tau=1}^{t-1}(-1)^{\tau}\binom{t-1}{\tau}\overline{r^{\tau}}\right) \qquad \text{(S.7)}\end{aligned}$$

where a bar represents the average taken with respect to all locus pairs, and $\binom{t-1}{\tau}=\frac{(t-1)!}{\tau!(t-1-\tau)!}$. The limit in Eq. (S.6) follows from the fact that $1/L\to 0$, $L(L-1)/L^2\to 1$, and the number of pairs (i,j) such that $i\neq j$ is $L(L-1)$. The series in Eq. (S.7) derives from the binomial expansion of $(1-r)^{t-1}$ in Eq. (S.6).

Finally, let $IBD^{(t)}$ be the fraction of the generation-t individual's (diploid) genome that is inherited identically by descent from the generation-0 ancestor. $IBD^{(t)} = P^{(t)}/2$, so

$$\mathbb{E}[IBD^{(t)}] = \frac{1}{2^t} \tag{S.8}$$

and

$$\mathrm{Var}(IBD^{(t)}) \xrightarrow[L\to\infty]{} \frac{1}{2^{t+1}}\left(\overline{(1-r)^{t-1}} - \frac{1}{2^{t-1}}\right) \tag{S.9}$$

$$= \frac{1}{2^{t+1}}\left(1 - \frac{1}{2^{t-1}} + \sum_{\tau=1}^{t-1}(-1)^{\tau}\binom{t-1}{\tau}\overline{r^{\tau}}\right). \tag{S.10}$$

In the special case of the descendant being a grand-offspring ($t = 2$), Eq. (S.9) becomes

$$\mathrm{Var}(IBD_{\mathrm{grand}}) = \mathrm{Var}(IBD^{(2)}) \xrightarrow[L\to\infty]{} \frac{1}{8}\left(\overline{(1-r)} - \frac{1}{2}\right) = \frac{1}{8}\left(\frac{1}{2} - \bar{r}\right), \tag{S.11}$$

which is Eq. (2) in the Main Text.

Case 2: Sex differences in recombination. Let $r_{ij}^{\female}$ and $r_{ij}^{\male}$ be the sex-specific recombination rates between loci i and j. If, among the $t-1$ individuals in the lineage between the generation-0 ancestor and the focal generation-t descendant, there are f females and $m = t-1-f$ males, then

$$\mathrm{Cov}\left(P_i^{(t)}, P_j^{(t)}\right) = \frac{1}{2^{t-1}}\left((1-r_{ij}^{\female})^f(1-r_{ij}^{\male})^m - \frac{1}{2^{t-1}}\right), \tag{S.12}$$

so that, by a similar calculation to Eq. (S.9) above,

$$\mathrm{Var}\left(IBD^{(t)}\right) = \frac{1}{2^{t+1}}\left(\overline{(1-r^{\female})^f(1-r^{\male})^m} - \frac{1}{2^{t-1}}\right). \tag{S.13}$$

If the number of females in the lineage is not known, it can be taken to be binomially distributed with parameter $1/2$, in which case the average in Eq. (S.13) is calculated across all locus pairs and all possible numbers of females $f = 0, 1, \ldots, t-1$ (with associated probabilities $\binom{t-1}{f}/2^{t-1}$).

S2. General case for indirect relationships

Relationships with one common ancestor

Consider an individual (generation 0) and two of its descendants (generation t_1 and t_2) who have no more recent common ancestor than the generation-0 individual, and no other recent common ancestor that is not also an ancestor of their shared generation-0 ancestor. The two generation-1 ancestors of the focal descendants (which could be the focal descendants themselves if $t_1 = t_2 = 1$) are half-sibs. Let $P_k^{(t_1,t_2)}$ be a random variable that takes the value 1 if both focal descendants carry, at locus k, an allele inherited identically from their common generation-0 ancestor. Assuming Mendelian segregation,

$$\text{Prob}\left(P_k^{(t_1,t_2)} = 1\right) = \frac{1}{2^{t_1+t_2-1}}, \tag{S.14}$$

so that $\mathbb{E}\left[P_k^{(t_1,t_2)}\right] = \text{Prob}\left(P_k^{(t_1,t_2)} = 1\right) = 1/2^{t_1+t_2-1}$ and

$$\text{Var}\left(P_k^{(t_1,t_2)}\right) = \mathbb{E}\left[\left(P_k^{(t_1,t_2)}\right)^2\right] - \left(\mathbb{E}\left[P_k^{(t_1,t_2)}\right]\right)^2 = 1/2^{t_1+t_2-1} - \left(1/2^{t_1+t_2-1}\right)^2.$$

Now consider two loci, i and j. For alleles at both loci in both descendants to have been inherited identically from their common ancestor in generation 0 requires that:

(i) The two generation-1 ancestors carry alleles IBD at both loci. This requires that the two loci be recombinant in both or neither of the generation-0 ancestor's two gametes that produced the generation-1 ancestors [probability $(1-r_{ij})^2 + r_{ij}^2 = 1 - 2r_{ij}(1-r_{ij})$], and, given this, that the same alleles segregated to the two gametes (probability $1/2$).
(ii) Given (i), the two loci are then not recombinant in any subsequent gamete leading to the focal generation-t_1 and generation-t_2 descendants, which occurs with probability $(1-r_{ij})^{t_1+t_2-2}$.
(iii) Given (i) and (ii), the ancestor's alleles always segregate into the gametes leading to the focal descendants, which occurs with probability $1/2^{t_1+t_2-2}$.

Therefore,

$$\text{Prob}\left(P_i^{(t_1,t_2)} = P_j^{(t_1,t_2)} = 1\right) = \left(\frac{1}{2} - r_{ij}(1-r_{ij})\right)(1-r_{ij})^{t_1+t_2-2}\frac{1}{2^{t_1+t_2-2}}, \tag{S.15}$$

so that

$$\begin{aligned}
\text{Cov}\left(P_i^{(t_1,t_2)}, P_j^{(t_1,t_2)}\right) &= \mathbb{E}\left[P_i^{(t_1,t_2)}P_j^{(t_1,t_2)}\right] - \mathbb{E}\left[P_i^{(t_1,t_2)}\right]\mathbb{E}\left[P_j^{(t_1,t_2)}\right] \\
&= \text{Prob}\left(P_i^{(t_1,t_2)} = P_j^{(t_1,t_2)} = 1\right) - \left(\frac{1}{2^{t_1+t_2-1}}\right)^2 \\
&= \left(\frac{1}{2} - r_{ij}(1-r_{ij})\right)(1-r_{ij})^{t_1+t_2-2}\frac{1}{2^{t_1+t_2-2}} - \left(\frac{1}{2^{t_1+t_2-1}}\right)^2 \\
&= \frac{1}{2^{t_1+t_2-1}}\left(\left[1-2r_{ij}(1-r_{ij})\right](1-r_{ij})^{t_1+t_2-2} - \frac{1}{2^{t_1+t_2-1}}\right).
\end{aligned} \tag{S.16}$$

Now let $IBD^{(t_1,t_2)}$ be the fraction of the genome that both the focal descendants have inherited from their common generation-0 ancestor: $IBD^{(t_1,t_2)} = \frac{1}{2L}\sum_{k=1}^{L} P_k^{(t_1,t_2)}$. Then

$$\mathbb{E}\left[IBD^{(t_1,t_2)}\right] = \frac{1}{2L}\sum_{k=1}^{L}\mathbb{E}\left[P_k^{(t_1,t_2)}\right] = \frac{1}{2L}\sum_{k=1}^{L}\frac{1}{2^{t_1+t_2-1}} = \frac{1}{2^{t_1+t_2}}, \tag{S.17}$$

while

$$\begin{aligned}
\text{Var}\left(IBD^{(t_1,t_2)}\right) &= \text{Var}\left(\frac{1}{2L}\sum_{k=1}^{L}P_k^{(t_1,t_2)}\right) \\
&= \frac{1}{4L^2}\sum_{k=1}^{L}\text{Var}\left(P_k^{(t_1,t_2)}\right) + \frac{1}{4L^2}\sum_{i\neq j}\text{Cov}\left(P_i^{(t_1,t_2)}, P_j^{(t_1,t_2)}\right) \\
&= \frac{1}{4L^2}L\frac{1}{2^{t_1+t_2-1}}\left(1 - \frac{1}{2^{t_1+t_2-1}}\right) \\
&\quad + \frac{1}{4L^2}\sum_{i\neq j}\frac{1}{2^{t_1+t_2-1}}\left(\left[1-2r_{ij}(1-r_{ij})\right](1-r_{ij})^{t_1+t_2-2} - \frac{1}{2^{t_1+t_2-1}}\right) \\
&\xrightarrow[L\to\infty]{} \frac{1}{2^{t_1+t_2+1}}\left(\overline{\left[1-2r_{ij}(1-r_{ij})\right](1-r_{ij})^{t_1+t_2-2}} - \frac{1}{2^{t_1+t_2-1}}\right).
\end{aligned} \tag{S.18}$$

In the special case of the focal descendants being half-sibs ($t_1 = t_2 = 1$), Eq. (S.18) becomes

$$\mathrm{Var}(IBD_{\text{h-sib}}) = \mathrm{Var}\left(IBD^{(1,1)}\right) \xrightarrow[L\to\infty]{} \frac{1}{8}\left(\frac{1}{2} - 2\overline{r_{ij}(1-r_{ij})}\right) = \frac{1}{8}\left(\frac{1}{2} - \bar{r}_{(2)}\right), \tag{S.19}$$

which is Eq. (3) in the Main Text. Here, $2r_{ij}(1-r_{ij})$ is the probability that i and j are recombinant in exactly one of two gametes, and $\bar{r}_{(2)}$ is the average value of $\bar{r}$ calculated from the pooled crossovers of two independent meioses of the common parent.

Relationships with two common ancestors

Consider an unrelated mating pair (generation 0) and two of its descendants (generation t_1 and t_2) who have no more recent common ancestors than the generation-0 mating pair, and no other recent common ancestor that is not an ancestor of one member of the generation-0 mating pair. If $t_1 = t_2 = 1$, the focal descendants are full-siblings.Note that we have restricted attention to two-ancestor pedigrees where the two ancestors were a mating pair; i.e., we have excluded from attention pedigrees such as that depicted in Fig. S1B below.

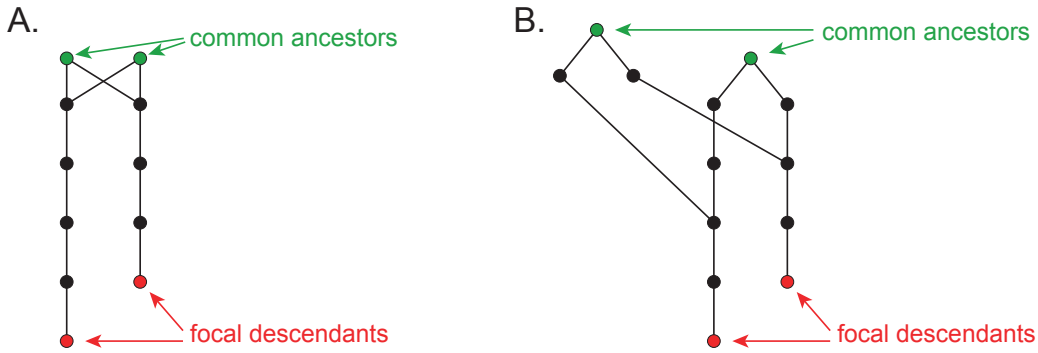

Figure S1 Examples of the general kinds of two-ancestor pedigrees we do consider (A) and do not consider (B). In pedigree A, $t_1 = 5$ and $t_2 = 4$.

Let $IBD^{(t_1,t_2)}$ be the proportion of the focal descendants' genomes that they share IBD. Label the members of the mating pair 1 and 2 (female and male, respectively), and let $P_{k,m}^{(t_1,t_2)}$ be a random variable that takes the value 1 if both focal descendants carry, at locus k, an allele inherited from member $m \in (1,2)$ of the mating pair. Then $IBD^{(t_1,t_2)} = \frac{1}{2L}\sum_{k=1}^{L}\left(P_{k,1}^{(t_1,t_2)} + P_{k,2}^{(t_1,t_2)}\right)$. Notice that, if $t_1, t_2 > 1$, then $P_{k,1}^{(t_1,t_2)} = 1 \;\Rightarrow\; P_{k,2}^{(t_1,t_2)} = 0$. Therefore, we consider three separate cases:

Case 1: $t_1 = t_2 = 1$ (full-sibs). In the case of full-sibs, because they each inherit their maternal and paternal genomes independently, the random variables $P_{i,1}^{(1,1)}$ and $P_{j,2}^{(1,1)}$ are independent for all (i,j), and have the same distribution as $P_k^{(1,1)}$ defined in the subsection 'Relationships with one common ancestor' above. Moreover, the random variables $\frac{1}{2L}\sum_{k=1}^{L} P_{k,1}^{(1,1)}$ and $\frac{1}{2L}\sum_{k=1}^{L} P_{k,2}^{(1,1)}$ have the same distribution as $\frac{1}{2L}\sum_{k=1}^{L} P_k^{(1,1)}$, with the appropriate recombination process—female and male, respectively—substituted in each case. Therefore,

$$\mathbb{E}\left[IBD_{\text{sib}}\right] = \mathbb{E}\left[IBD^{(1,1)}\right] = \mathbb{E}\left[\frac{1}{2L}\sum_{k=1}^{L}\left(P_{k,1}^{(1,1)} + P_{k,2}^{(1,1)}\right)\right] = \frac{1}{L}\sum_{k=1}^{L}\mathbb{E}\left[P_k^{(1,1)}\right] = \frac{1}{2},$$

and

$$\begin{aligned}
\mathrm{Var}\left(IBD_{\text{sib}}\right) &= \mathrm{Var}\left(IBD^{(1,1)}\right) = \mathrm{Var}\left(\frac{1}{2L}\sum_{k=1}^{L}\left(P_{k,1}^{(1,1)} + P_{k,2}^{(1,1)}\right)\right) \\
&= \mathrm{Var}\left(\frac{1}{2L}\sum_{k=1}^{L} P_{k,1}^{(1,1)}\right) + \mathrm{Var}\left(\frac{1}{2L}\sum_{k=1}^{L} P_{k,2}^{(1,1)}\right) \\
&\xrightarrow[L\to\infty]{} \frac{1}{8}\left(\frac{1}{2} - \bar{r}_{(2)}^{\,\varphi}\right) + \frac{1}{8}\left(\frac{1}{2} - \bar{r}_{(2)}^{\,\sigma}\right) \qquad \text{[from Eq. (S.19)]} \\
&= \frac{1}{8}\left(1 - \bar{r}_{(2)}^{\,\varphi} - \bar{r}_{(2)}^{\,\sigma}\right),
\end{aligned}$$

which is Eq. (4) in the Main Text.

Case 2: $t_1 = 1, t_2 > 1$. Without loss of generality, let the first focal individual be an offspring of the ancestral mating pair, and the second focal individual a more distant descendant. Then the second individual carries one haploid copy of the genome inherited

through the pedigree, while the first individual carries two copies. Let the random variable $P_k^{(1,t_2)}$ take the value 1 if the focal descendants carry, at locus k, an allele inherited IBD. $P_k^{(1,t_2)} = P_{k,1}^{(1,t_2)} + P_{k,2}^{(1,t_2)}$, so that

$$\mathbb{E}\left[P_k^{(1,t_2)}\right] = \mathbb{E}\left[P_{k,1}^{(1,t_2)}\right] + \mathbb{E}\left[P_{k,2}^{(1,t_2)}\right] = 2 \times 1/2^{t_2} = 1/2^{t_2-1}$$

and $IBD^{(1,t_2)} = \frac{1}{2L}\sum_{k=1}^{L} P_k^{(1,t_2)}$. Now consider two loci, i and j. For $P_i^{(1,t_2)} = P_j^{(1,t_2)} = 1$, we require one of the following mutually exclusive events to occur:

- $P_{i,1}^{(1,t_2)} = P_{j,1}^{(1,t_2)} = 1$. This requires that the generation-1 siblings inherited the same maternal alleles at i and j (probability $[(1-r_{ij}^{\varphi})^2 + r_{ij}^{\varphi 2}]/2$), and that this allele pair was thereafter transmitted faithfully to the focal generation-t_2 descendant (probability $(1-r_{ij})^{t_2-1}/2^{t_2-1}$).

  Total probability: $\left[(1-r_{ij}^{\varphi})^2 + r_{ij}^{\varphi 2}\right](1-r_{ij})^{t_2-1}/2^{t_2}$.

- $P_{i,2}^{(1,t_2)} = P_{j,2}^{(1,t_2)} = 1$. This requires that the generation-1 siblings inherited the same paternal alleles at i and j (probability $[(1-r_{ij}^{\sigma})^2 + r_{ij}^{\sigma 2}]/2$), and that this allele pair was thereafter transmitted faithfully to the focal generation-t_2 descendant (probability $(1-r_{ij})^{t_2-1}/2^{t_2-1}$).

  Total probability: $\left[(1-r_{ij}^{\sigma})^2 + r_{ij}^{\sigma 2}\right](1-r_{ij})^{t_2-1}/2^{t_2}$.

- $P_{i,1}^{(1,t_2)} = P_{j,2}^{(1,t_2)} = 1$. This requires that the generation-1 siblings inherited the same maternal allele at i (probability $1/2$) and the same paternal allele at j (probability $1/2$), that these two alleles were transmitted together to the generation-2 individual in the lineage of the focal generation-t_2 descendant (probability $r_{ij}/2$), and that these two alleles were then transmitted together to the focal generation-t_2 descendant (probability $(1-r_{ij})^{t_2-2}/2^{t_2-2}$).

  Total probability: $r_{ij}(1-r_{ij})^{t_2-2}/2^{t_2+1}$.

- $P_{i,2}^{(1,t_2)} = P_{j,1}^{(1,t_2)} = 1$. This requires that the generation-1 siblings inherited the same paternal allele at i (probability $1/2$) and the same maternal allele at j (probability $1/2$), that these two alleles were transmitted together to the generation-2 individual in the lineage of the focal generation-t_2 descendant (probability $r_{ij}/2$), and that these two alleles were then transmitted together to the focal generation-t_2 descendant (probability $(1-r_{ij})^{t_2-2}/2^{t_2-2}$).

  Total probability: $r_{ij}(1-r_{ij})^{t_2-2}/2^{t_2+1}$.

Therefore,

$$\begin{aligned}
\text{Prob}\left(P_i^{(1,t_2)} = P_j^{(1,t_2)} = 1\right) &= \left[(1-r_{ij}^{\varphi})^2 + r_{ij}^{\varphi 2}\right](1-r_{ij})^{t_2-1}/2^{t_2} \\
&\quad + \left[(1-r_{ij}^{\sigma})^2 + r_{ij}^{\sigma 2}\right](1-r_{ij})^{t_2-1}/2^{t_2} \\
&\quad + 2r_{ij}(1-r_{ij})^{t_2-2}/2^{t_2+1} \\
&= \frac{1}{2^{t_2}}\left((1-r_{ij})^{t_2-1}\left[(1-r_{ij}^{\varphi})^2 + r_{ij}^{\varphi 2} + (1-r_{ij}^{\sigma})^2 + r_{ij}^{\sigma 2} + r_{ij}(1-r_{ij})^{-1}\right]\right).
\end{aligned}$$

If the sex of members of the pedigree other than the ancestral mating pair is known, then it is clear where in the above calculations to substitute specific male or female values of r_{ij}.

From the calculations above,

$$\begin{aligned}
\text{Cov}\left(P_i^{(1,t_2)}, P_j^{(1,t_2)}\right) &= \mathbb{E}\left[P_i^{(1,t_2)} P_j^{(1,t_2)}\right] - \mathbb{E}\left[P_i^{(1,t_2)}\right]\mathbb{E}\left[P_j^{(1,t_2)}\right] \\
&= \text{Prob}\left(P_i^{(t_1,t_2)} = P_j^{(t_1,t_2)} = 1\right) - \left(\frac{1}{2^{t_2-1}}\right)^2 \\
&= \frac{1}{2^{t_2}}\left((1-r_{ij})^{t_2-1}\left[(1-r_{ij}^{\varphi})^2 + r_{ij}^{\varphi 2} + (1-r_{ij}^{\sigma})^2 + r_{ij}^{\sigma 2} + r_{ij}(1-r_{ij})^{-1}\right] - \frac{1}{2^{t_2-2}}\right). \qquad \text{(S.20)}
\end{aligned}$$

Using this result, and from calculations similar to those throughout this paper,

$$\mathbb{E}\left[IBD^{(1,t_2)}\right] = \frac{1}{2L}\sum_{k=1}^{L}\mathbb{E}\left[P_k^{(1,t_2)}\right] = \frac{1}{2L}\sum_{k=1}^{L}\frac{1}{2^{t_2-1}} = \frac{1}{2^{t_2}} \qquad \text{(S.21)}$$

is the coefficient of relationship (e.g., $1/4$ for aunt-nephew [$t_2 = 2$]), while

$$\begin{aligned}\mathrm{Var}\left(IBD^{(1,t_2)}\right) &\xrightarrow[L\to\infty]{} \frac{1}{4}\overline{\mathrm{Cov}\left(P_i^{(1,t_2)}, P_j^{(1,t_2)}\right)} \\ &= \frac{1}{2^{t_2+2}}\overline{(1-r_{ij})^{t_2-1}\left[(1-r_{ij}^{\varphi})^2 + r_{ij}^{\varphi 2} + (1-r_{ij}^{\sigma})^2 + r_{ij}^{\sigma 2} + r_{ij}(1-r_{ij})^{-1}\right]} - \frac{1}{2^{2t_2}}. \end{aligned} \tag{S.22}$$

Case 3: $t_1, t_2 > 1$. Let the random variable $P_k^{(t_1,t_2)}$ take the value 1 if the focal descendants have inherited, within the focal pedigree, the same allele at locus k. Then

$$\mathbb{E}\left[P_k^{(t_1,t_2)}\right] = \mathbb{E}\left[P_{k,1}^{(t_1,t_2)}\right] + \mathbb{E}\left[P_{k,2}^{(t_1,t_2)}\right] = 2 \times 1/2^{t_1+t_2-1} = 1/2^{t_1+t_2-2}$$

and $IBD^{(t_1,t_2)} = \frac{1}{2L}\sum_{k=1}^{L} P_k^{(t_1,t_2)}$. Now consider two loci, i and j. For $P_i^{(t_1,t_2)} = P_j^{(t_1,t_2)} = 1$, we require one of the following mutually exclusive events to occur:

- $P_{i,1}^{(t_1,t_2)} = P_{j,1}^{(t_1,t_2)} = 1$. This requires that the generation-1 siblings inherited the same maternal alleles at i and j (probability $[(1-r_{ij}^{\varphi})^2 + r_{ij}^{\varphi 2}]/2$), and that this allele pair was thereafter transmitted faithfully to both focal descendants (probability $(1-r_{ij})^{t_1+t_2-2}/2^{t_1+t_2-2}$).

  Total probability: $\left[(1-r_{ij}^{\varphi})^2 + r_{ij}^{\varphi 2}\right](1-r_{ij})^{t_1+t_2-2}/2^{t_1+t_2-1}$.

- $P_{i,2}^{(t_1,t_2)} = P_{j,2}^{(t_1,t_2)} = 1$. This requires that the generation-1 siblings inherited the same paternal alleles at i and j (probability $[(1-r_{ij}^{\sigma})^2 + r_{ij}^{\sigma 2}]/2$), and that this allele pair was thereafter transmitted faithfully to both focal descendants (probability $(1-r_{ij})^{t_1+t_2-2}/2^{t_1+t_2-2}$).

  Total probability: $\left[(1-r_{ij}^{\sigma})^2 + r_{ij}^{\sigma 2}\right](1-r_{ij})^{t_1+t_2-2}/2^{t_1+t_2-1}$.

- $P_{i,1}^{(t_1,t_2)} = P_{j,2}^{(t_1,t_2)} = 1$. This requires that the generation-1 siblings inherited the same maternal allele at i (probability $1/2$) and the same paternal allele at j (probability $1/2$), that both generation-1 siblings transmitted these two alleles in producing the generation-2 cousins (probability $r_{ij}^2/4$), and that the allele pair was then transmitted faithfully to both focal descendants (probability $(1-r_{ij})^{t_1+t_2-4}/2^{t_1+t_2-4}$).

  Total probability: $r_{ij}^2(1-r_{ij})^{t_1+t_2-4}/2^{t_1+t_2}$.

- $P_{i,2}^{(t_1,t_2)} = P_{j,1}^{(t_1,t_2)} = 1$. This requires that the generation-1 siblings inherited the same paternal allele at i (probability $1/2$) and the same maternal allele at j (probability $1/2$), that both generation-1 siblings transmitted these two alleles in producing the generation-2 cousins (probability $r_{ij}^2/4$), and that the allele pair was then transmitted faithfully to both focal descendants (probability $(1-r_{ij})^{t_1+t_2-4}/2^{t_1+t_2-4}$).

  Total probability: $r_{ij}^2(1-r_{ij})^{t_1+t_2-4}/2^{t_1+t_2}$.

Therefore,

$$\begin{aligned}\mathrm{Prob}\left(P_i^{(t_1,t_2)} = P_j^{(t_1,t_2)} = 1\right) &= \left[(1-r_{ij}^{\varphi})^2 + r_{ij}^{\varphi 2}\right](1-r_{ij})^{t_1+t_2-2}/2^{t_1+t_2-1} \\ &\quad + \left[(1-r_{ij}^{\sigma})^2 + r_{ij}^{\sigma 2}\right](1-r_{ij})^{t_1+t_2-2}/2^{t_1+t_2-1} \\ &\quad + 2r_{ij}^2(1-r_{ij})^{t_1+t_2-4}/2^{t_1+t_2} \\ &= \frac{1}{2^{t_1+t_2-1}}\left((1-r_{ij})^{t_1+t_2-2}\left[(1-r_{ij}^{\varphi})^2 + r_{ij}^{\varphi 2} + (1-r_{ij}^{\sigma})^2 + r_{ij}^{\sigma 2} + r_{ij}^2(1-r_{ij})^{-2}\right]\right).\end{aligned}$$

Again, if the sex of members of the pedigree other than the ancestral mating pair is known, then it is clear where in the above calculations to substitute specific male or female values of r_{ij}.

From the calculations above,

$$\begin{aligned}\mathrm{Cov}\left(P_i^{(t_1,t_2)}, P_j^{(t_1,t_2)}\right) &= \mathbb{E}\left[P_i^{(t_1,t_2)}P_j^{(t_1,t_2)}\right] - \mathbb{E}\left[P_i^{(t_1,t_2)}\right]\mathbb{E}\left[P_j^{(t_1,t_2)}\right] \\ &= \mathrm{Prob}\left(P_i^{(t_1,t_2)} = P_j^{(t_1,t_2)} = 1\right) - \left(\frac{1}{2^{t_1+t_2-2}}\right)^2 \\ &= \frac{1}{2^{t_1+t_2-1}}\left((1-r_{ij})^{t_1+t_2-2}\left[(1-r_{ij}^{\varphi})^2 + r_{ij}^{\varphi 2} + (1-r_{ij}^{\sigma})^2 + r_{ij}^{\sigma 2} + r_{ij}^2(1-r_{ij})^{-2}\right] - \frac{1}{2^{t_1+t_2-3}}\right).\end{aligned} \tag{S.23}$$

Using this result, and from calculations similar to those throughout this paper,

$$\mathbb{E}\left[IBD^{(t_1,t_2)}\right] = \frac{1}{2L}\sum_{k=1}^{L}\mathbb{E}\left[P_k^{(t_1,t_2)}\right] = \frac{1}{2L}\sum_{k=1}^{L}\frac{1}{2^{t_1+t_2-2}} = \frac{1}{2^{t_1+t_2-1}} \tag{S.24}$$

is the coefficient of relationship (e.g., $1/8$ for full cousins [$t_1 = t_2 = 2$]), while

$$\begin{aligned}\mathrm{Var}\left(IBD^{(t_1,t_2)}\right) &\xrightarrow[L\to\infty]{} \frac{1}{4}\overline{\mathrm{Cov}\left(P_i^{(t_1,t_2)}, P_j^{(t_1,t_2)}\right)} \\ &= \frac{1}{2^{t_1+t_2+1}}\overline{(1-r_{ij})^{t_1+t_2-2}\left[(1-r_{ij}^{\varphi})^2 + r_{ij}^{\varphi 2} + (1-r_{ij}^{\sigma})^2 + r_{ij}^{\sigma 2} + r_{ij}^2(1-r_{ij})^{-2}\right]} - \frac{1}{2^{2t_1+2t_2-2}}.\end{aligned} \tag{S.25}$$

S3. Mathematical differences between calculations of the variance of bp versus cM genetic relatedness.

The calculations in Sections S1 and S2 define genetic relatedness as the proportion of total physical genome length (in bp) shared by relatives IBD. An alternative definition of genetic relatedness is the proportion of total genetic length (in cM) shared IBD. cM genetic relatedness suffers from several fundamental problems that bp genetic relatedness does not suffer from, as discussed in the Main Text, but here we focus on technical differences between calculations of the variance of bp versus cM genetic relatedness. These differences are not obvious at first—in both sets of calculations, the variance of genetic relatedness boils down to a sum, across all distinct pairs of loci i and j, of the covariance of identity by descent at the two loci:

$$\mathrm{Var}(IBD) \propto \sum_{i \neq j} \mathrm{Cov}_{ij}. \tag{S.26}$$

[Compare, for example, the calculations in Sections S1 and S2 with calculations of the variance of cM genetic relatedness in Hill and Weir (2011).] The covariance of IBD state between loci i and j is always some function of the recombination fraction between i and j (see Sections S1 and S2). In calculations of the variance of cM genetic relatedness, it is typically further assumed that crossover interference is absent, so that recombination fractions translate to a simple exponential function of genetic map distances, according to Haldane's map function.

The fundamental difference between the two sets of calculations lies only in the fact that calculations of the variance of cM genetic relatedness implicitly assume that the loci indexed by i and j are uniformly spaced along the genetic map, i.e., that there is a constant cM distance between each pair of adjacent loci (Franklin 1977). Our calculations, on the other hand, implicitly assume that the loci are uniformly spaced along the physical map—as would be the case, for example, if we took our set of loci to be the set of all base pairs in the genome. That calculations of the variance of cM and bp genetic relatedness assume even spacing of loci along the genetic and physical maps, respectively, can be seen in the fact that, in both sets of calculations, Cov_{ij} is given the same weight for each locus pair (i, j).

The fact that the loci are assumed to be evenly spaced along the genetic and physical maps in the two sets of calculations implies that the number of loci on a given chromosome is proportional to that chromosome's genetic length (for cM genetic relatedness) or its physical length (for bp genetic relatedness). This observation allows us to compare the effect of independent assortment of chromosomes on the variance of bp versus cM genetic relatedness. Suppose that the length of chromosome k (measured in bp or cM, depending on the definition of genetic relatedness being employed) is l_k, and that the total genome length is $L = \sum_{k=1}^{n} l_k$. The proportion of loci that lie on chromosome k is therefore l_k/L, and the proportion of locus pairs such that both loci lie on chromosome k is $(l_k/L)^2$. Now, when loci i and j lie on different chromosomes (and therefore recombine freely), the covariance of identity by descent at the two loci is zero for every pedigree relationship [see, e.g., Eqs. (S.4), (S.12), (S.16), (S.20), and (S.23) above]. Therefore, for every pair of unlinked loci i and j, the term Cov_{ij} falls out of the sum in Eq. (S.26). The greater the fraction of locus pairs (i, j) for which Cov_{ij} falls out of Eq. (S.26), the greater the negative effect of independent assortment of chromosomes on the variance of genetic relatedness. This fraction of locus pairs is given by

$$1 - \sum_{k=1}^{n} (l_k/L)^2, \tag{S.27}$$

which is the Gini-Simpson index for chromosome lengths, and is also proportional (by a factor of $1/2$) to the contribution of independent assortment of chromosomes to $\bar{r}$, $\bar{r}_{(2)}$, and analogs (Veller *et al.* 2019). For a fixed number of chromosomes, the Gini-Simpson index measures the homogeneity of chromosome lengths—the more homogenous they are, the larger is the Gini-Simpson index, and thus the greater is the fraction of locus pairs that fall out of Eq. (S.26), causing independent assortment of chromosomes to have a stronger negative effect on the variance of genetic relatedness. Therefore, to compare the effect of independent assortment of chromosomes on bp versus cM genetic relatedness, we calculate how homogenous chromosome lengths are in the two cases according to the Gini-Simpson index. In humans, for both males and females, it turns out that chromosome lengths are more homogenous when measured in cM than in bp: the Gini-Simpson index for autosomal chromosome lengths measured in bp is 0.9460; for autosomal chromosome lengths measured in cM, the Gini-Simpson index is 0.9488 in males and 0.9479 in females. Thus, in humans, independent assortment of chromosomes has a weaker negative effect on the genome-wide variance of bp genetic relatedness than on the genome-wide variance of cM genetic relatedness, as we have noted in the Main Text.

S4. Variance calculations for relatedness to grandparents and to siblings

Table S1 Standard deviations of genetic relatedness to paternal and maternal grandparents, and to siblings, in humans.

	Grandparent							Sibling			
	Paternal				Maternal						
Chrom.	bp, cytolog.	bp, linkage (Kosambi)	bp, linkage (Haldane)	cM	bp, linkage (Kosambi)	bp, linkage (Haldane)	cM	bp, cytolog. + link. (Kos.)	bp, linkage (Kosambi)	bp, linkage (Haldane)	cM
1	0.145	0.149	0.166	0.156	0.115	0.133	0.126	0.146	0.147	0.163	0.147
2	0.152	0.155	0.171	0.161	0.117	0.135	0.130	0.152	0.151	0.167	0.152
3	0.164	0.166	0.180	0.168	0.126	0.144	0.138	0.165	0.164	0.181	0.162
4	0.171	0.174	0.187	0.174	0.127	0.145	0.139	0.170	0.171	0.187	0.166
5	0.173	0.175	0.188	0.176	0.128	0.146	0.142	0.173	0.172	0.188	0.169
6	0.183	0.183	0.195	0.181	0.133	0.151	0.145	0.184	0.183	0.198	0.175
7	0.179	0.177	0.190	0.178	0.135	0.153	0.148	0.181	0.178	0.194	0.176
8	0.187	0.192	0.201	0.184	0.139	0.157	0.152	0.192	0.194	0.209	0.183
9	0.186	0.193	0.202	0.186	0.149	0.166	0.157	0.199	0.203	0.217	0.188
10	0.184	0.189	0.199	0.182	0.139	0.157	0.152	0.188	0.191	0.206	0.181
11	0.187	0.191	0.201	0.189	0.142	0.160	0.157	0.193	0.195	0.210	0.190
12	0.179	0.185	0.196	0.182	0.140	0.158	0.154	0.183	0.188	0.204	0.183
13	0.177	0.184	0.196	0.192	0.152	0.169	0.170	0.192	0.196	0.211	0.204
14	0.175	0.182	0.194	0.195	0.160	0.176	0.178	0.195	0.198	0.214	0.214
15	0.180	0.185	0.196	0.197	0.154	0.170	0.172	0.196	0.197	0.213	0.209
16	0.190	0.201	0.209	0.194	0.155	0.171	0.169	0.207	0.214	0.228	0.204
17	0.193	0.196	0.205	0.195	0.150	0.167	0.168	0.204	0.204	0.218	0.204
18	0.198	0.202	0.209	0.201	0.158	0.173	0.173	0.213	0.214	0.228	0.213
19	0.199	0.223	0.226	0.203	0.178	0.191	0.182	0.226	0.250	0.261	0.223
20	0.195	0.211	0.217	0.211	0.167	0.181	0.184	0.213	0.229	0.242	0.231
21	0.198	0.214	0.219	0.219	0.189	0.199	0.205	0.235	0.249	0.261	0.260
22	0.201	0.207	0.214	0.218	0.188	0.199	0.205	0.236	0.241	0.254	0.259
Genome	0.040	0.041	0.044	0.040	0.031	0.035	0.034	0.041	0.041	0.045	0.040

Key: bp: Genetic relatedness defined as the proportion of total physical genome length shared IBD.

cM: Genetic relatedness defined as the proportion of the total genetic map length shared IBD. For relatedness to paternal grandparent, this is the male map; for relatedness to maternal grandparent, it is the female map; for relatedness of siblings, it is the sex-averaged map.

Kosambi: Linkage map distances d converted to recombination fractions r using Kosambi's map function, $r = \frac{1}{2}\tanh(2d)$.

Haldane: Linkage map distances d converted to recombination fractions r using Haldane's map function, $r = \frac{1}{2}(1 - e^{-2d})$.

bp, cytolog. + link. (Kos.): Cytological data used for male meiosis; linkage data used (with Kosambi's map function) for female meiosis.

S5. Calculating the variance across instances of a given pedigree relationship

The calculations in Sections S1 and S2 deal with variation in genetic relatedness in a given instance of some pedigree relationship, generated by the randomness of recombination and segregation in the meiotic processes of the individuals involved in the pedigree. Often, however, we will be interested in the variance of genetic relatedness across instances of a given pedigree relationship. For example, Visscher *et al.* (2006) use the fact that sibling pairs vary in how genetically related they are to develop a method for estimating the heritability of traits. Since there can be systematic differences across individuals in their recombination processes, this must be taken into account in calculating the cross-pedigree variance, or 'population variance', of genetic relatedness.

Consider a specified pedigree relationship, and let R represent all information about the meioses involved in determining identity-by-descent sharing for a given instance of this relationship. E.g., if the relationship is grandoffspring to paternal grandparent, then, for a given instance of this relationship, R carries all information about the recombination process of the father in question, e.g., his recombination fraction r_{ij} for every pair of loci (i, j), his degree of crossover covariation across chromosomes, etc. In general, for our purposes, R can be thought of as a probability distribution over all possible crossover configurations in the relevant meioses. R itself varies across instances of the specified pedigree relationship, because the recombination process can vary systematically across individuals of the same sex.

The calculations in Sections S1 and S2 implicitly condition on a particular R, and reveal that, conditional on R, the variance of genetic relatedness can be written in the form

$$\mathrm{Var}(IBD|R) \propto \sum_{i,j} \mathrm{Cov}(IBD_i, IBD_j|R) = \sum_{i,j} f(r_{ij}),$$

where IBD_k is the identity-by-descent state at locus k, and the r_{ij} are the recombination fractions associated with the particular R.

To calculate the population variance of genetic relatedness, variation in R must be taken into account. From the law of total variance,

$$\mathrm{Var}(IBD) = \mathbb{E}_R\left[\mathrm{Var}(IBD|R)\right] + \mathrm{Var}_R\left(\mathbb{E}[IBD|R]\right),$$

where $\mathbb{E}_R$ and Var_R are the expectation and variance taken with respect to the distribution of R. $\mathbb{E}[IBD|R]$ is a constant (1/2 for siblings, etc.), so $\mathrm{Var}_R\left(\mathbb{E}[IBD|R]\right) = 0$, and

$$\mathrm{Var}(IBD) = \mathbb{E}_R\left[\mathrm{Var}(IBD|R)\right] \propto \mathbb{E}_R\left[\sum_{i,j} f(r_{ij})\right] = \sum_{i,j} \mathbb{E}_R[f(r_{ij})].$$

Therefore, in calculating the unconditional/population variance $\mathrm{Var}(IBD)$, the population average of $f(r_{ij})$ can be calculated separately for each pair of loci i and j, with the results then summed across all locus pairs. An immediate implication is that, because each locus pair can be treated separately in this calculation, the covariation of crossovers across chromosomes within individual gametes does not affect $\mathrm{Var}(IBD)$, just as it does not affect $\mathrm{Var}(IBD|R)$. Similarly, covariation across individuals of recombination rates on different chromosomes (caused, for example, by inter-individual differences in the average lengths of the chromosome axes at meiotic prophase I) does not affect $\mathrm{Var}(IBD)$.

In general, therefore, calculation of the population variance of genetic relatedness is similar in form to calculation of the variance in a given instance of the specified pedigree relationship—in the latter case, the average of $f(r_{ij})$ is taken over all locus pairs, while in the former case, the average of $f(r_{ij})$ is taken over all locus pairs and recombination processes R. An important practical point arises when f is a nonlinear function of r_{ij}, as it is for all pedigree relationships except grandoffspring-grandparent. In such cases, because $\mathbb{E}_R\left[f(r_{ij})\right] \neq f\left(\mathbb{E}_R[r_{ij}]\right)$, it is technically invalid to use population-averaged recombination fractions $\mathbb{E}_R[r_{ij}]$ (as would be obtained, for example, from linkage maps) in the calculation of $\mathrm{Var}(IBD)$. Instead, $f(r_{ij})$ itself should be estimated for each individual instance of the pedigree relationship, with the results then averaged across instances. Therefore, in general, it is preferable to use disaggregated data of crossover positions in individual nuclei (meiocytes or gametes) to calculate the population variance of genetic relatedness.

We note that this is not an issue in the particular relationship of grandoffspring-grandparent, because, in this case, $f(r_{ij}) = \frac{1}{8}\left(\frac{1}{2} - r_{ij}\right)$, so that $\mathbb{E}_R\left[f(r_{ij})\right] = \mathbb{E}_R\left[\frac{1}{8}\left(\frac{1}{2} - r_{ij}\right)\right] = \frac{1}{8}\left(\frac{1}{2} - \mathbb{E}_R[r_{ij}]\right) = f\left(\mathbb{E}_R[r_{ij}]\right)$. Therefore, in this case, it is valid to use population-averaged recombination rates to calculate the population variance of genetic relatedness.

Consider the other focal relationship in our Main Text, that of siblings (we'll concentrate on half-siblings, for simplicity). In this case, $f(r_{ij}) = \frac{1}{8}\left(\frac{1}{2} - 2r_{ij}(1 - r_{ij})\right)$. Because f is nonlinear, $\mathbb{E}_R\left[f(r_{ij})\right] \neq f\left(\mathbb{E}_R[r_{ij}]\right)$; in fact, because f is strictly concave in r_{ij}, $\mathbb{E}_R\left[f(r_{ij})\right] > f\left(\mathbb{E}_R[r_{ij}]\right)$ as long as there is variation in R. So, using population-averaged recombination rates will systematically lead to underestimation of the true population variance of genetic relatedness of siblings (or half-siblings). The size of the error introduced by using population-averaged recombination rates can be calculated precisely. The true population variance of the genetic relatedness of half-siblings is given by $V = \mathbb{E}_R\left[\frac{1}{8}\left(\frac{1}{2} - \overline{2r_{ij}(1 - r_{ij})}\right)\right]$, where the overbar represents an average across all locus pairs. The estimate

obtained using population-averaged recombination rates is $V' = \frac{1}{8}\left(\frac{1}{2} - \overline{2\mathbb{E}_R[r_{ij}](1-\mathbb{E}_R[r_{ij}])}\right)$. The amount by which V exceeds V' is

$$\begin{aligned}
V - V' &= \mathbb{E}_R\left[\frac{1}{8}\left(\frac{1}{2} - \overline{2r_{ij}(1-r_{ij})}\right)\right] - \frac{1}{8}\left(\frac{1}{2} - \overline{2\mathbb{E}_R[r_{ij}](1-\mathbb{E}_R[r_{ij}])}\right) \\
&= \frac{1}{4}\left(-\mathbb{E}_R[\overline{r_{ij}(1-r_{ij})}] + \overline{\mathbb{E}_R[r_{ij}](1-\mathbb{E}_R[r_{ij}])}\right) \\
&= \frac{1}{4}\left(\overline{-\mathbb{E}_R[r_{ij}(1-r_{ij})]} + \overline{\mathbb{E}_R[r_{ij}](1-\mathbb{E}_R[r_{ij}])}\right) \\
&= \frac{1}{4}\left(\overline{-\mathbb{E}_R[r_{ij}] + \mathbb{E}_R[r_{ij}^2] + \mathbb{E}_R[r_{ij}] - (\mathbb{E}_R[r_{ij}])^2}\right) \\
&= \frac{1}{4}\left(\overline{\mathbb{E}_R[r_{ij}^2] - (\mathbb{E}_R[r_{ij}])^2}\right) \\
&= \frac{1}{4}\overline{\mathrm{Var}_R(r_{ij})}.
\end{aligned}$$

Therefore, the error introduced by using population-averaged recombination rates to estimate the population variance of the relatedness of half-siblings (and siblings) is simply proportional to the average, taken across locus pairs, of the population variance of pairwise recombination rates.

To see how large this error can be in practice, we made use of what—to our knowledge—is the richest available dataset of crossover positions in individual nuclei, generated by Bell *et al.* (2020) by single-cell sequencing of large numbers of sperm from 20 human male donors (∼1000–2000 sperm per donor). We set 100 loci evenly spaced along the genomic map of each chromosome, and calculate, for each individual and each pair of linked loci i and j, the proportion of gametes in which loci i and j were recombinant. Given these recombination fractions r_{ij} for each individual (and assuming that $r_{ij} = 1/2$ when i and j are on different chromosomes), we calculate $\bar{r}_{(2)} = \overline{2r_{ij}(1-r_{ij})}$ for each individual. Averaging these individual values of $\bar{r}_{(2)}$ across the 20 donors yields an estimate of the population-averaged value of $\bar{r}_{(2)}$, which is the relevant value for calculating the population variance of genetic relatedness of paternal half-siblings in human.

At the same time, by first averaging the r_{ij} across individuals, we calculate a pooled-recombination-rate value of $\bar{r}_{(2)}$, which, the above considerations predict, should be larger than the value of $\bar{r}_{(2)}$ averaged across individuals (leading to a lower estimate of the variance of genetic relatedness of half-siblings). The calculations just described were performed for the whole genome and for each chromosome (using chromosome-specific values of $\bar{r}_{(2)}$), allowing us to compute the variance of genetic relatedness of paternal half-siblings using the correct and incorrect measures both genome-wide and per-chromosome.

We find that, as predicted, using a value of $\bar{r}_{(2)}$ calculated from pooled recombination fractions leads to a smaller estimate of the variance of genetic relatedness of paternal half-siblings than if we calculate this variance using a value of $\bar{r}_{(2)}$ averaged across individuals. However, the size of the error is small. The estimate of the genome-wide variance, based on the correct averaging of $\bar{r}_{(2)}$, is 1.187×10^{-3}, while the estimate based on the incorrect averaging is 1.184×10^{-3}, a difference of only 0.25%. The associated standard deviations differ by only 0.13% The chromosome-specific proportional errors are of the same order, and in fact are typically smaller than the genome-wide error.

Therefore, in the case of human siblings, we conclude that, although it is technically invalid to use population-averaged recombination rates to calculate the variance of genetic relatedness, in practice the errors introduced are likely to be small.

Literature Cited

Bell, A. D., C. J. Mello, J. Nemesh, S. A. Brumbaugh, A. Wysoker, *et al.*, 2020 Insights into variation in meiosis from 31,228 human sperm genomes. Nature **583**: 259–264.

Franklin, I. R., 1977 The distribution of the proportion of the genome which is homozygous by descent in inbred individuals. Theoretical Population Biology **11**: 60–80.

Hill, W. G. and B. S. Weir, 2011 Variation in actual relationship as a consequence of Mendelian sampling and linkage. Genetics Research **93**: 47–64.

Veller, C., N. Kleckner, and M. A. Nowak, 2019 A rigorous measure of genome-wide genetic shuffling that takes into account crossover positions and Mendel's second law. Proceedings of the National Academy of Sciences **116**: 1659–1668.

Visscher, P. M., S. E. Medland, M. A. R. Ferreira, K. I. Morley, G. Zhu, *et al.*, 2006 Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. PLoS Genetics **2**: e41.