The effect of consanguinity on between-individual identity-by-descent sharing

Alissa L. Severson*
Shai Carmi†
Noah A. Rosenberg‡

March 27, 2019

Abstract

Consanguineous unions increase the rate at which identical genomic segments are paired within individuals to produce runs of homozygosity (ROH). The extent to which such unions affect identity-by-descent (IBD) genomic sharing between rather than within individuals in a population, however, is not immediately evident from within-individual ROH levels. Using the fact that the time to the most recent common ancestor (T_MRCA) for a pair of genomes at a specific locus is inversely related to the extent of IBD sharing between the genomes in the neighborhood of the locus, we study IBD sharing for a pair of genomes sampled either within the same individual or in different individuals. We develop a coalescent model for a set of mating pairs in a diploid population, treating the fraction of consanguineous unions as a parameter. Considering mating models that include unions between sibs, first cousins, and nth cousins, we determine the effect of the consanguinity rate on the mean T_MRCA for pairs of lineages sampled either within the same individual or in different individuals. The results indicate that consanguinity not only increases ROH sharing between the two genomes within an individual, it also increases IBD sharing between individuals in the population, the magnitude of the effect increasing with the kinship coefficient of the type of consanguineous union. Considering computations of ROH and between-individual IBD in Jewish populations whose consanguinity rates have been estimated from demographic data, we find that in accord with the theoretical results, increases in consanguinity and ROH levels inflate levels of IBD sharing between individuals in a population. The results contribute more generally to the interpretation of runs of homozygosity, identity-by-descent sharing between individuals, and the relationship between ROH and IBD.

Keywords: coalescent; consanguinity; identity by descent; runs of homozygosity; time to the most recent common ancestor

Introduction

Consanguineous unions, which mating pairs have a close genetic relationship, produce offspring whose two genomic copies have higher levels of identity-by-descent sharing than is seen for corresponding offspring of non-consanguineous unions. The offspring of consanguineous unions can inherit two copies of a segment of the genome from the same recent ancestor—a shared close relative of both mother and father—through separate maternal and paternal lines of descent. Because this ancestor is recent, little time has been available for recombination to break the segment, so that the two copies can be identical over a long distance (Figure 1A).

Long runs of homozygosity (ROH), regions in which the two homologous chromosomes of an individual are identical over long distances, have been observed to co-occur with known high rates of consanguinity (WOODS et al., 2006; HUNTER-ZINCK et al., 2010; SCOTT et al., 2016; CEBALLOS et al., 2018). In humans, many populations with a high rate for consanguineous unions have been seen to be among the populations with the largest fractions of their genomes residing in long runs of homozygosity (KIRIN et al., 2010; PEMBERTON et al., 2012; KARAFET et al., 2015; KANG et al., 2016).

Measurement of identity-by-descent (IBD) sharing between genomes in distinct individuals has emerged as a powerful method for analysis of population relationships and demographic history (BROWNING and BROWNING, 2012; PALAMARA et al., 2012; HARRIS and NIELSEN, 2013; RALPH and COOP, 2013; THOMPSON, 2013). IBD sharing is computed for pairs of genomes in individuals at different geographic scales or in comparisons of pairs from the same or different populations. The pattern of sharing is then used to infer demographic histories.

Informally, high levels of IBD sharing between individuals within populations have been seen in some of the same populations that possess high ROH levels (KANG et al., 2016). However, it is not clear from a theoretical understanding of the determinants of IBD sharing that ROH levels, measured within individuals, and levels of between-individual IBD sharing would have a direct relationship. Consanguinity increases the probability that the same genomic segment appears in two copies in the same individual; the way in which consanguinity relates...
verify the prediction of the model by examining ROH and thereby increasing between-individual IBD sharing. We consider the effect of consanguinity on genomic sharing between genomes in two “unrelated” individuals (blue).

It is possible that the increased IBD sharing within individuals that is produced by consanguinity increases IBD sharing between individuals, as an enlarged inbreeding coefficient decreases effective population size and hence might increase genomic sharing between all pairs of individuals. On the other hand, it is possible that the increased genomic sharing within offspring resulting from consanguinity has little or no effect on sharing between pairs of genomes in individuals from different families; increased IBD sharing for individuals within a family that has many consanguineous unions might be counteracted by decreased IBD sharing for individuals from different families that are not closely related.

A difficulty in evaluating the effect of relationships among consanguinity, ROH, and IBD sharing between individuals is that the phenomena of interest concern properties of a diploid population pedigree. Unlike in many problems in population genetics, in which a diploid population of size $2N$ exchangeable individuals can be approximated by a model of a corresponding haploid population of size $4N$ (Wakeley, 2009, Chapter 6.1), for the study of consanguinity, it is important to consider mating pairs of diploid individuals, and to account for the possibility that individuals might have many consanguineous mating pairs in their ancestry.

Here, adapting a model of $N$ diploid mating pairs, each of which can represent a consanguineous pair or a non-consanguineous pair, we study the effect of consanguinity on the mean time to the most recent common ancestor ($T_{MRCA}$) for pairs of gene lineages sampled either as the two genomic copies within an individual or as two copies from different individuals. In the model, not only does consanguinity decrease $E[T_{MRCA}]$ for pairs of genomic copies within an individual, thereby increasing ROH levels, it also decreases $E[T_{MRCA}]$ for pairs of genomic copies in separate individuals in a population, thereby increasing between-individual IBD sharing. We verify the prediction of the model by examining ROH and IBD sharing in data from human populations.

**Model**

$T_{MRCA}$, ROH, and IBD: Our goal is to study the relationship between ROH within individuals and IBD sharing between individuals. To do so, we examine a model of a genetic locus in a population, in which we can consider two random variables: $T$, the $T_{MRCA}$ for the two gene lineages sampled from the same individual chosen at random in the population, and $V$, or $T_{MRCA}$ for a pair of lineages from two individuals chosen at random in separate mating pairs. The choice to study $T_{MRCA}$ arises from the fact that the length of genome shared around a focal site is closely related to $T_{MRCA}$ at the site (Palamara et al., 2012; Carmi et al., 2014; Browning and Browning, 2015). Thus, lower values of $T$ lead to longer homozygous segments within individuals, and lower values of $V$ lead to longer IBD segments in pairs of individuals. The relationship between $T$ and $V$ and its dependence on model parameters then provide insight into the relationship between ROH and IBD.

**Diploid mating model:** We study a diploid discrete-generation model with sib mating that was introduced by Campbell (2015), extending it to permit other forms of consanguinity. The model of Campbell (2015) considers a constant-sized diploid population with $N \geq 2$ monogamous mating pairs, $2N$ individuals, and $4N$ allelic copies at a locus. Some of the mating pairs are consanguineous, and the others are non-consanguineous. In particular, in each generation, a constant fraction ($c_0$) of the pairs represent sib matings (Figure 2A). Although in principle, $c_0$ can be viewed as a probability of consanguinity that ranges from 0 to 1, in our model, to ensure that the number of sib mating pairs is an integer, $c_0$ must be a multiple of $1/N$.

One generation back in time, for each of the $c_0N$ sib mating pairs, a single parental mating pair is chosen uniformly at random with replacement to represent the parents of the mating sibs (Figure 2B). For each of the remaining $(1 - c_0)N$ non-consanguineous mating pairs, two parental mating pairs in the previous generation are chosen uniformly at random from the $\binom{2N}{2}$ possibilities,
alleles coalesce, giving a coalescence time of 1 generation.

In this model, two allelic copies at a locus have three possible states (Figure 3). They can be the two alleles of the same individual (state 1). Alternatively, they can be in the two individuals of a mating pair, one in each member of the pair (state 2). Finally, they can be in two individuals in separate mating pairs (state 3). We define three random variables corresponding to these three states: $T$ is $T_{MRCA}$ for two alleles in the same individual, $U$ is $T_{MRCA}$ for two alleles in two individuals in a mating pair, and $V$ is $T_{MRCA}$ for two alleles in two individuals in separate mating pairs (Figure 3).

CAMPBELL (2015) derived the mean coalescence time $E[T]$ for two alleles in an individual as a function of the population size $N$ and the fraction of sib mating pairs $c_0$. We begin by recapitulating the results of CAMPBELL (2015) in the diploid model with sib mating, also examining $E[U]$ and $E[V]$. Next, we extend the model to consider $E[T]$, $E[U]$, and $E[V]$ in other consanguinity regimes: first cousin mating, $n$th cousin mating, and a superposition of multiple degrees of cousin mating. We find that in all regimes, consanguinity decreases $E[T]$, $E[U]$, and $E[V]$, thereby predicting that consanguinity increases both ROH lengths within individuals and IBD sharing between individuals. A single result unifies the consanguinity regimes in terms of the kinship coefficient of the pairs of individuals in consanguineous unions.

Sib mating

Following CAMPBELL (2015), we first rederive $E[T]$, $E[U]$, and $E[V]$ in units of generations by setting up recursions using a first-step analysis. For $E[T]$, if two alleles are present within one individual (state 1), then they must have been present in two individuals in a mating pair in the previous generation (state 2). Hence, $T = U + 1$ and

$$E[T] = E[U] + 1. \quad (1)$$

For $E[U]$, if two alleles are in the two individuals of a mating pair (state 2), then with probability $c_0$, the pair is a sib mating pair. Three cases are possible for the previous generation. With probability $\frac{1}{4}$, the two alleles coalesce, giving a coalescence time of 1 generation. With probability $\frac{1}{4}$, they are the two alleles of the same individual (state 1), giving a mean coalescence time of $E[T] + 1$. With probability $\frac{1}{2}$, they are two alleles in the two individuals of a mating pair (state 2), generating mean coalescence time $E[U] + 1$.

Chance sib mating is forbidden among the pairs that are not among the $c_0N$ sib mating pairs. If the two individuals in the mating pair are not sibs, then in the previous generation, the alleles trace to two separate mating pairs (state 3), giving mean coalescence time $E[V] + 1$. Combining the various cases, we have

$$E[U] = c_0 \left( \frac{1}{4} + \frac{E[T] + 1}{4} + \frac{E[U] + 1}{2} \right) + (1 - c_0)(E[V] + 1). \quad (2)$$

Finally, for $E[V]$, because parental pairs are chosen uniformly at random with replacement among $N$ possible pairs, two individuals in separate mating pairs are sibs with probability $\frac{1}{N}$. In the previous generation, if the two individuals are sibs, then the two alleles can either coalesce, be in the same parent in the previous generation (state 1), or be in separate parents (state 2). If they are not sibs, then the alleles lie in two individuals in separate mating pairs in the previous generation (state 3). Combining these cases, we have:

$$E[V] = \frac{1}{N} \left( \frac{1}{4} + \frac{E[T] + 1}{4} + \frac{E[U] + 1}{2} \right) + \left( 1 - \frac{1}{N} \right)(E[V] + 1). \quad (3)$$

Eqs. 1-3 form a linear system of equations in $E[T]$, $E[U]$, and $E[V]$, the solution to which is

$$E[T] = 4N(1 - c_0) + 6 \quad (4)$$

$$E[U] = 4N(1 - c_0) + 5 \quad (5)$$

$$E[V] = 4N \left( 1 - \frac{3}{4} c_0 \right) + 4. \quad (6)$$

Note that although CAMPBELL (2015) presented only eq. 4, eqs. 5 and 6 also result from solving the system.

We can immediately observe that $E[V] - E[T] = c_0N - 2$, so that if $c_0$ exceeds $\frac{2}{N}$, or the population has more than 2 consanguineous mating pairs each generation, then $E[V] > E[T]$, and the mean coalescence time for two alleles in different mating pairs exceeds the mean coalescence time within individuals. For $c_0 \leq \frac{2}{N}$, $E[V]$ and $E[T]$ differ by at most 2 generations.

If $c_0 = 0$, then the reduced model of $N$ monogamous diploid pairs with sib mating avoidance produces mean coalescence times close to the mean coalescence time of $4N$ for two lineages chosen uniformly at random in a haploid population of size $4N$ (WAKELEY, 2009, Chapter 6.1). The factors of $1 - c_0$ in eqs. 4 and 5 and $1 - \frac{3}{4} c_0$ in eq. 6 provide linear reductions in mean coalescence time owing to increasing consanguinity $c_0$.

Eqs. 4, 5, and 6, normalized by $4N$, are plotted in Figure 4 as functions of $N$ for various values of $c_0$. As $N$ increases, the constant terms in eqs. 4-6 become unimportant, and the mean coalescence times are dominated by a product of $4N$, the number of allelic copies in the population, and the reduction factor due to consanguinity, $1 - c_0$ or $1 - \frac{3}{4} c_0$. 

![Figure 3](https://example.com/figure3.png)
Next, we extend the model to first cousin mating and again derive $E[T]$, $E[U]$, and $E[V]$ in the same manner as in the sib mating case. In each generation, the fraction of first cousin mating pairs is a constant value $c_1$. Similarly to the sib mating case, both chance first cousin mating and chance sib mating are forbidden among the remaining non-consanguineous pairs. Consanguineous pairs are assumed not to be to double first cousins, and chance double first cousin mating is also forbidden among non-consanguineous pairs.

$E[T]$ is the same as with sib mating: if two alleles are present within one individual (state 1), then they must have been present in two individuals in a mating pair in the previous generation (state 2), and eq. 1 still holds.

For $E[U]$, if two alleles are in two individuals of a mating pair (state 2), then with probability $c_1$, those individuals are first cousins. If they are first cousins, then each has a parent who is the offspring of the shared grandparental mating pair (Figure 5B). For each individual in the first cousin mating pair, the probability that the sampled allele is inherited from the sib parent is $\frac{1}{2}$. Consequently, the probability that the sampled alleles in both individuals are inherited from the sib parents is $\frac{1}{4}$. If both alleles are inherited from the sib parents, then—similar to sib mating—two generations ago, three cases are possible. First, with probability $\frac{1}{4}$, the two alleles coalesce, giving a coalescence time of 2 generations. With probability $\frac{1}{4}$, they are the two alleles of the same individual (state 1), giving a mean coalescence time of $E[T] + 2$. Finally, with probability $\frac{1}{2}$, they are two alleles in the two individuals of a mating pair (state 2), generating mean coalescence time $E[U] + 2$.

With probability $1 - \frac{1}{4}$, two alleles in two individuals in a mating pair are not inherited from a shared grandparental mating pair. Because both chance sib and first cousin mating are forbidden, two generations ago the alleles are in separate mating pairs, giving mean coalescence time $E[V] + 2$. Combining the cases gives

$$E[U] = \frac{c_1}{4} \left( \frac{2}{4} + \frac{E[T]}{4} + \frac{E[U]}{2} + 2 \right) + \left( 1 - \frac{c_1}{4} \right) (E[V] + 2).$$

(7)

Lastly, for $E[V]$ the formula is the same as Eq. 3 because parental pairs of individuals are still chosen uniformly at random with replacement from the $N$ pairs.

Eqs. 1, 7, and 3 form a linear system of equations, with solution

$$E[T] = 4N \left( 1 - \frac{1}{4} c_1 \right) + 10$$

(8)

$$E[U] = 4N \left( 1 - \frac{1}{4} c_1 \right) + 9$$

(9)

$$E[V] = 4N \left( 1 - \frac{3}{16} c_1 \right) + 7.$$  

(10)

We first note that $E[V] - E[T] = \frac{c_1 N}{4} - 3$, so if $c_1 > \frac{12}{N}$ or the number of consanguineous mating pairs exceeds 12, then the mean coalescence time for two alleles in different mating pairs, $E[V]$, exceeds the mean coalescence time for two alleles within an individual, $E[T]$. As $c_1$ approaches 0, the mean coalescence times are near $4N$, the mean coalescence time for two lineages from a randomly mating haploid population of size $4N$. On the other hand, for $c_1$ near 1, $E[T] \approx 3N$ and $E[V] \approx \frac{13}{4} N$. The mean coalescence times are reduced linearly due to consanguinity, by a factor of $1 - \frac{3}{16} c_1$ in Eq. 8 and 9, and by a factor of $1 - \frac{3}{16} c_1$ in Eq. 10.

Eqs. 8 and 10, normalized by $4N$, are plotted in Figure 6. As the number of mating pairs $N$ increases, the mean coalescence times approach the product of $4N$ and a reduction factor due to consanguinity. In contrast to sib mating, for which $E[T]/(4N)$ decreases to 0 and $E[V]/(4N)$ to $\frac{1}{4}$ for large $N$ and $c_0 = 1$, $E[T]/(4N)$ is bounded below by $\frac{3}{4}$ and $E[V]/(4N)$ by $\frac{13}{16}$.
nth cousins

The similarity of the derivation in the cases of sib mating and first cousin mating suggests a generalization to nth cousin mating, where \( n = 1 \) represents first cousin mating and \( n = 0 \) represents sib mating. As before, \( c_n \) is the fraction of mating pairs that represent nth cousins. It will be convenient to assume that chance mating of ith cousins is forbidden for all \( i \) from 0 to \( n \). Beginning with \( \mathbb{E}[T] \), if two alleles are within one individual, then as before, they must have been in two individuals in a mating pair in the previous generation. Eq. 1 continues to hold.

For \( \mathbb{E}[U] \), with probability \( c_n \), the individuals in the mating pair are nth cousins. They then share an ancestral mating pair \( n + 1 \) generations in the past and have ancestors that are sibs \( n \) generations ago (Figure 5C). The probability that a pair of alleles, one in an offspring of one sib and one in an offspring of the other sib, both trace to the sibs is \( \frac{1}{4^n} \). For each of the next \( n - 1 \) generations connecting the sibs to the nth cousins, the conditional probability that the transmitted alleles are both from the sibs given that they are from the sibs in the previous generation is \( \frac{1}{4^n} \). Consequently, with probability \( \frac{1}{4^n} \), the sampled alleles in the current generation are inherited from the sib ancestors.

Conditional on tracing to the sibs, three cases exist for the two alleles in the shared ancestral mating pair: with probability \( \frac{1}{4} \), the alleles coalesce \( n + 1 \) generations ago. With probability \( \frac{1}{4} \), the two alleles are in state 1 and have mean coalescence time \( \mathbb{E}[T] + n + 1 \). Lastly, with probability \( \frac{1}{2} \), the two alleles are in state 2 and have mean coalescence time \( \mathbb{E}[U] + n + 1 \). If the two alleles are not inherited from the ancestral sibs, or if the individuals in the mating pair in the current generation are not nth cousins, then because chance mating of cousins of degree 0, 1, 2, ... , \( n \) is forbidden, the two alleles are in separate mating pairs \( n + 1 \) generations ago and have mean coalescence time \( \mathbb{E}[V] + n + 1 \). Combining the cases gives

\[
\mathbb{E}[U] = \frac{c_n}{4^n} \left( \frac{n + 1}{4} + \frac{\mathbb{E}[T] + n + 1}{4} + \frac{\mathbb{E}[U] + n + 1}{2} \right) + \left( 1 - \frac{c_n}{4^n} \right) (\mathbb{E}[V] + n + 1). \tag{11}
\]

Lastly, for \( \mathbb{E}[V] \), because parental pairs are chosen uniformly at random with replacement from the \( N \) possible pairs, eq. 3 continues to hold.

Eqs. 1, 11, and 3 form a linear system of equations, the solution to which is

\[
\begin{align*}
\mathbb{E}[T] &= 4N \left( 1 - \frac{1}{4^n} c_n \right) + 4n + 6 \tag{12} \\
\mathbb{E}[U] &= 4N \left( 1 - \frac{1}{4^n} c_n \right) + 4n + 5 \tag{13} \\
\mathbb{E}[V] &= 4N \left( 1 - \frac{3}{4^n+1} c_n \right) + 3n + 4. \tag{14}
\end{align*}
\]

Note that eqs. 12-14 give eqs. 4-6 as a special case when \( n = 0 \), and eqs. 8-10 when \( n = 1 \). We can consider the difference \( \mathbb{E}[V] - \mathbb{E}[T] = c_n N/4^n - (n + 2) \). If \( c_n > 4^n(n + 2)/N \), or the number of consanguineous pairs exceeds \( 4^n(n + 2) \), then the mean coalescence time for two alleles in different mating pairs, \( \mathbb{E}[V] \), exceeds the mean coalescence time for two alleles within an individual, \( \mathbb{E}[T] \). As \( n \) increases, the first term \( c_n N/4^n \) approaches zero and the two means differ by approximately \( n + 2 \). For fixed \( n \), the mean coalescence times are reduced linearly due to consanguinity, by a factor of \( 1 - c_n/4^n \) in eq. 12 and 13, and by \( 1 - 3c_n/4^n+1 \) in eq. 14.

Eqs. 12 and 14, normalized by \( 4N \), are plotted in Figure 7 as functions of the degree \( n \) of the cousin relationship. The terms in these equations that reduce coalescence times are \( c_n/4^n \) in eq. 12 and \( 3c_n/4^n+1 \) in eq. 14. As the degree \( n \) of the cousin relationship increases, these terms decrease exponentially to zero, and the mean coalescence times approach \( 4N \). The ratio \( \mathbb{E}[V]/\mathbb{E}[T] \), taking the ratio of eqs. 14 and 12, is plotted in Figure 8 as a function of \( c_n \) for \( n \) from 0 to 5. As the fraction of cousin mating \( c_n \) increases, the ratio increases above 1, so \( \mathbb{E}[V] > \mathbb{E}[T] \); however as the degree of the relationship \( n \) increases for fixed \( c_n \), the ratio decreases toward 1.

Superposition of multiple mating levels

We now combine all forms of consanguinity examined thus far into a superposition of levels of cousin mating in which ith cousin mating is permitted for each \( i \) from 0 to \( n \). For each \( i \) from 0 to \( n \), let \( c_i \) be the fraction of ith cousin mating pairs in each generation, and let \( n \) be the degree of the most distant cousin relationship allowed. For each \( i \leq n \), chance ith cousin mating is prohibited. We assume individuals in a consanguineous mating pair cannot be related by more than one path; for example, they cannot be both first and third cousins. This assumption is designed for use with a large population and a small value of \( \sum_{i=0}^{n} c_i \ll 1 \). For fixed \( n \), as \( N \) becomes large, the probability that two individuals in a
Figure 7: Normalized mean coalescence times as a function of degree \( n \) of the relationship and the fraction \( c_n \) of \( n \)th cousin mating pairs. \( N \to \infty \) is assumed. (A) \( E[T]/(4N) \), eq. 12. (B) \( E[V]/(4N) \), eq. 14.

![Figure 7](image)

Figure 8: The ratio \( E[V]/E[T] \), or eq. 14/eq. 12, as a function of the fraction of cousin mating \( c_n \) and the degree of the cousin cousin relationship \( n \). \( N \to \infty \) is assumed.

![Figure 8](image)

consanguineous mating pair share more than one recent ancestor is regarded as negligible.

\( E[T] \) is the same as in the previous models: two alleles within one individual must have been in two individuals in a mating pair in the previous generation (eq. 1).

For \( E[U] \), for each \( i \leq n \), with probability \( c_i \) the individuals in the mating pair are \( i \)th cousins. As was seen with \( n \)th cousins, with probability \( 1/n \), the two alleles were inherited from sib ancestors \( i \) generations ago. Then \( i + 1 \) generations ago there are three possible cases: with probability \( 1/n \), the alleles coalesce. With probability \( 1/n \), the alleles are in state 1 and have mean coalescence time \( E[T] + i + 1 \). Finally, with probability \( 1/n \), the alleles are in state 2 and have mean coalescence time \( E[U] + i + 1 \).

The probability is \( 1 - \sum_{i=0}^{n} \frac{c_i}{4^i} \) that the two alleles are either not in a consanguineous mating pair for all \( i \) from 0 to \( n \), or not inherited from the shared ancestral mating pair. Then because chance mating of cousins of degree 0, 1, 2, \ldots, \( n \) is forbidden, the two alleles are in separate mating pairs \( n + 1 \) generations ago and have mean coalescence time \( E[V] + n + 1 \). Combining the cases for all \( i \leq n \) gives

\[
E[U] = \sum_{i=0}^{n} \frac{c_i}{4^i} \left( \frac{i + 1}{4} + \frac{E[T] + i + 1}{4} + \frac{E[U] + i + 1}{2} \right) + (E[V] + n + 1) \left( 1 - \sum_{i=0}^{n} \frac{c_i}{4^i} \right).
\]

(15)

Because parental pairs are chosen uniformly at random with replacement from the \( N \) possible pairs, for two alleles in separate mating pairs, eq. 3 holds as before.

We define \( c \) as the sum over \( i \) of the probability that two alleles in a mating pair chosen at random are inherited by descent from the same allele in a shared ancestral mating pair \( i + 1 \) generations in the past:

\[
c = \sum_{i=0}^{n} \frac{c_i}{4^i}.
\]

(16)

In other words, \( c \) is defined in the same way as the kinship coefficient of the two individuals in a randomly chosen mating pair (Jacquard, 1972; Lange, 1997); it is the probability that two alleles selected at random from a randomly chosen mating pair are identical by descent.

Eqs. 1, 15, and 3 form a system of equations, the solution to which is

\[
E[T] = 4N(1 - 4c) + 4n(1 - 4c) + 16d + 6
\]

(17)

\[
E[U] = 4N(1 - 4c) + 4n(1 - 4c) + 16d + 5
\]

(18)

\[
E[V] = 4N(1 - 3c) + 3n(1 - 4c) + 12d + 4,
\]

(19)

where

\[
d = \sum_{i=0}^{n} \frac{ic_i}{4^i+1}.
\]

(20)

First, if \( c = c_n/4^{n+1} \) for any \( n \), then eqs. 17-19 reduce to eqs. 12-14. The difference \( E[V] - E[T] = 4Nc - n(1 - 4c) - 4d - 2 \) for large \( N \) is approximately \( 4Nc \).

For sufficiently large \( N \), the constant terms in eqs. 17-19 contribute little. Next, for each \( i \), \( c_i \leq 1 \), so \( d < \sum_{i=0}^{\infty} i/4^{i+1} = 5/9 \) and the contributions from 16d in eqs. 17 and 18 and from 12d in eq. 19 are relatively small. Finally, noting that for probabilities \( (c_0, \ldots, c_n) \) with \( \sum_{i=0}^{n} c_i \leq 1 \), the sum in eq. 16 is maximized if \( c_0 = 1 \) and all other \( c_i \) equal 0, \( 0 \leq c \leq 1 \), so \( 0 \leq 1 - 4c \leq 1 \). If \( n \ll N \), then the maximal contribution of \( 4n \) in eqs. 17 and 18 and \( 3n \) in eq. 19 is also relatively small. Then except in the sib mating case of \( c_0 = 1 \) and \( c = 1/3 \), the means in eqs 17-19 are dominated by the product of \( 4N \) and the linear reduction factors \( 1 - 4c \) in eqs. 17 and 18 and \( 1 - 3c \) in eq. 19.

Application to data

**Background:** Previously, Kang et al. (2016) demonstrated that ROH sharing increases with consanguinity. Specifically, in their Figure 7, they observed a positive correlation between population means of the total ROH length and population levels of consanguinity available from demographic studies. This relationship
with our prediction that increased consanguinity reduces within-individual mean pairwise coalescence times $E[T]$ (eqs. 4, 8, 12, 17) and hence increases ROH length.

We use the data of KANG et al. (2016) to test predictions about the relationship between consanguinity, ROH, and IBD. Our model of the effect of consanguinity on coalescence times predicts that increased consanguinity decreases mean coalescence times both for pairs of alleles within individuals ($E[T]$) and for pairs of alleles in individuals in different mating pairs ($E[V]$), with a larger reduction for within-individual coalescence times. Because more recent coalescence times for pairs of lineages are expected to give rise to elevated genomic sharing, we expect IBD and ROH sharing to be correlated, owing to the fact that their associated coalescence times both decrease with increasing consanguinity. In addition, we expect a larger increase in ROH sharing relative to the corresponding increase in IBD, due to the larger relative decrease of coalescence times for $E[T]$ compared to $E[V]$.

Data set: We use data from KANG et al. (2016) consisting of 202 individuals from 18 Jewish populations and 2,903 individuals from 123 non-Jewish populations, with genotypes available at 257,091 SNPs. We focus our analysis on Jewish individuals classified by KANG et al. (2016) into six regional groups: Ethiopian, European, Middle Eastern, North African, South Asian, and Yemenite. The remaining non-Jewish individuals are a combination of the HGDP-CEPH and HapMap III data sets and were only included for phasing.

Data availability: See KANG et al. (2016) for the data used in this study.

Data analysis: ROH lengths for each individual were taken from KANG et al. (2016). Following PEMBERTON et al. (2012), KANG et al. (2016) classified ROH segments into three length categories: Class A for short segments, Class B for segments of intermediate length, and Class C for long segments. KANG et al. (2016) further examined the relationship between length class and consanguinity, demonstrating that the total length of the Class C segments drives the correlation between ROH length and consanguinity.

To calculate IBD, we first phased the full data set with Beagle 4.1 (BROWNING and BROWNING, 2007) using the default parameters (maxdr=5000, lowmem=false, window=50000, overlap=3000, niterations=5, impute=false, cluster=0.005, ne=1000000, err=0.001, seed=-99999, modelscale=0.8) and HapMap GRCh36 genetic maps for the map parameter. From the phased data, we called IBD segments with Refined IBD (BROWNING and BROWNING, 2013) using the default parameters (window=40.0, lod=3.0, length=1.5, trim=0.15, scale=3) and the same map files.

Total ROH length sums segments shared between two haplotypes within an individual, whereas total IBD length is a sum of four haplotype comparisons between two diploid individuals. To make IBD directly comparable with ROH, we calculated total IBD length by summing all segments shared between two individuals (reported by Refined IBD) and dividing by 4. This computation gives the mean total IBD length shared between two haplotypes chosen at random from the two individuals. We averaged this length across all pairs of distinct individuals within populations.

Results: In Figure 9A, we compare the relationship between mean total IBD across all pairs of individuals and mean total ROH across all individuals in 18 Jewish populations. As was noted by KANG et al. (2016), the longest ROH lengths occur primarily in the two South Asian Jewish populations and several of the Middle Eastern Jewish populations. Our new computation of IBD length generally accords with ATZMON et al. (2010), CAMPBELL et al. (2012), and WALDMAN et al. (2016a,b), in that the South Asian Jewish populations have the highest IBD sharing, followed by most Middle Eastern and North African Jewish populations, with European, Syrian, and Ethiopian Jews having the least sharing. Note that the particularly high level of IBD sharing in the Mumbai population has been observed previously in an independent sample (WALDMAN et al., 2016a).

IBD and ROH are positively correlated, with $\rho = 0.63$. The regression has positive slope 0.12 with $P = 0.012$, indicating that at a population level, a 1 Mb increase in mean total ROH is expected to increase total IBD by 120 kb on average. The positive relationship between ROH and IBD is consistent with the prediction under the model of a correlated relationship for within-
and between-individual coalescence times. Moreover, the slope is less than one, reflecting the greater reduction in within-individual coalescence times due to consanguinity compared to between-individual coalescence times.

In Figure 9B-D, we consider the relationship between mean total IBD and each of the ROH length classes. We observe the strongest correlation of IBD with total Class C or long ROH, with \( \rho = 0.62 \). Classes A (short) and B (intermediate) have positive but weaker correlations, with \( \rho = 0.45 \) and \( \rho = 0.44 \) respectively. The regression line for Class C is significant with \( P = 0.004 \), whereas for Classes A and B it is not significant, with \( P = 0.113 \) and \( P = 0.200 \) respectively. The relationship between IBD length and Class C ROH length suggests that in general, IBD and ROH are correlated because both are affected by consanguinity, in agreement with our theoretical predictions. The weaker correlations between IBD length and Classes A and B might result from comparatively less accurate calling of short IBD segments.

Discussion

Summary: We have studied the effect of consanguinity on within- and between-individual coalescence times. We extended the sib mating model of Campbell (2015) to permit first cousin mating, \( n \)th cousin mating, and a superposition of multiple levels of cousin mating, deriving mean coalescence times for two alleles within an individual (\( \mathbb{E}[T] \)) and two alleles in separate mating pairs (\( \mathbb{E}[V] \)). We found that consanguinity linearly reduces both means, with a greater reduction for within-individual coalescence times. To test our theoretical predictions, we studied ROH and IBD patterns in 18 Jewish populations, finding that they are correlated, and that the correlation is driven by long Class C ROH. These results support the prediction of the modeling framework that ROH and IBD levels are both amplified by consanguinity.

In each of our various models, for large \( N \), \( \mathbb{E}[T] \) and \( \mathbb{E}[V] \) are approximately equal to the product of \( 4N \), the mean \( T_{MRCA} \) in a haploid population of size \( 4N \), and a linear reduction term that depends on the fraction of consanguineous pairs and their degree of consanguinity. Thus, although the model considers diploids with a rigid monogamous mating structure, its coalescence times produce a close relationship to those of the standard haploid model.

The difference \( \mathbb{E}[V] - \mathbb{E}[T] \) is approximately \( c_n N/4^n \) for \( n \)th cousin mating and \( 4Nc \) for the superposition of different mating levels. The quantity \( 4Nc \) can be viewed as the expected number of coalescence events due to consanguinity, as it is the product of the number of pairs of alleles in two individuals in a mating pair (\( 4N \)) and the probability that two alleles in a mating pair are identical by descent and therefore coalesce quickly rather than on a coalescent timescale (\( c \)). In other words, two alleles in the same individual have probability \( c \) of having a coalescence time near zero, so that on average their coalescence time is expected to be \( 4Nc \) less than that of two alleles that are in different mating pairs and that do not have the probability \( c \) of near-immediate coalescence. Note that this perspective, based on the superposition case, also applies in the \( n \)th cousin mating case as the difference \( \mathbb{E}[V] - \mathbb{E}[T] \) for \( n \)th cousins is \( c_n N/4^n = 4Nc_n/4^{n+1} \), and \( c_n/4^{n+1} \) is the probability that two alleles in two individuals in a mating pair are identical by descent in this case.

Theoretical population genetics of ROH and IBD: If two genomes share a recent common ancestor at a site, then the length of the shared segment surrounding that site is likely to be long, because recombination has had little time to break down the segment. If the genomes share a distant common ancestor, then the surrounding segment is likely to be short because recombination will have had many generations to break it. In this way, recombination produces an inverse relationship between coalescence times at a genomic site and the length of the surrounding shared segment (Palamara et al., 2012; Carmi et al., 2014; Browning and Browning, 2015). The results of our model, that increased consanguinity decreases both within-individual and between-individual coalescence times, suggest that populations with higher rates of consanguinity will have more recent coalescence times and will share longer ROH and IBD segments. Moreover, the result \( \mathbb{E}[T] < \mathbb{E}[V] \) suggests that the reduction is greater for ROH than for IBD, and that consanguinity will have a stronger effect on ROH sharing.

To study ROH and IBD together in the same model, we generalized a diploid consanguineous model of sib mating. The Campbell (2015) model and its generalization represent examples of the increasing integration of coalescent perspectives into models that consider a diploid pedigree structure (Wolﬀenberg and Avise, 1998; Wakeley et al., 2012, 2016; Wilton et al., 2017; King et al., 2018). For example, Wakeley et al. (2012) found that in a pedigree-based consanguineous model, compared to a standard haploid model, the distribution of pairwise coalescence times for random pairs of individuals was altered, most strongly for the most recent coalescence times. Otherwise, the two models have similar coalescence time distributions. In our case, consideration of the pedigree—with no consanguinity—produced mean pairwise coalescence times close to the haploid mean pairwise coalescence time of \( 4N \). The inclusion of consanguinity in the model decreased the mean coalescence time by a linear factor dependent on the kinship coefficient of a randomly chosen mating pair.

Contrasting two hypotheses—one in which both within-individual ROH and between-individual IBD are increased by consanguinity, and the other in which consanguinity increases ROH but not IBD—we found support for the former rather than the latter view. According to our model, consanguinity inflates relatedness not only within families, but in the population in general, so that mean pairwise coalescence times decrease with increasing consanguinity, both for pairs of alleles within individuals and for pairs of alleles in separate mating
pairs. We can understand this phenomenon through the concept of coalescent effective population size. In this perspective (Sjödin et al., 2005), mean pairwise coalescence times have a direct relationship with effective size. Consequently, the direct relationship that we observed between coalescence times associated with ROH and those associated with IBD can be viewed as resulting from a decreased coalescent effective size that in turn results from consanguinity, and which decreases coalescence times both within and between individuals.

Previously, Jacquard (1970) studied the effect of inbreeding avoidance on effective population size. He modeled a two-sex, diploid population of $N$ individuals with equally many males and females, considering cases with and without sib mating avoidance (p. 175, 245). Sib mating avoidance generated a slightly larger effective size compared to the case in which sib mating was permissible, analogous to our observation that coalescence times decrease with increasing sib mating.

The sib mating case of our model is also similar to models of partial selfing in plants (Charlesworth, 2003). Such models can be viewed as having a linear combination of “consanguinity” (selfing) and “random mating” (outcrossing). In our sib mating model, two alleles have probability $\frac{2n}{4N}$ of coalescing in the previous generation, whereas under partial selfing, alleles have probability $\frac{s}{2N}$ of coalescence in the previous generation, where $s$ is the selfing rate. In a partially selfing population of $2N$ diploid individuals, taking $N \rightarrow \infty$, the effective population size is $2N(1 - \frac{s}{2})$ individuals (Pollak, 1987; Nordborg and Donnelly, 1997), a product of the effective size in a randomly mating population and a linear reduction factor proportional to the probability of coalescence, similar to our findings for $E[T]$ and $E[V]$. Our results are analogous to those of Milligan (1996), who studied the effect of partial selfing on within- and between-individual coalescence times, $T_w$ and $T_b$, respectively, finding $E[T_w] = 4N(1 - s)$ and $E[T_b] = 4N(1 - \frac{s}{2})$ for a population of size $2N$ diploid individuals. The greater reduction in coalescence time for within-versus between-individual comparisons echoes our results for $E[T]$ and $E[V]$. Moreover, $E[T_b] - E[T_w] = 4N\frac{s}{2}$, the product of the number of alleles and the probability of rapid coalescence from selfing, analogous to the difference $4Nc$ that we found for $E[V] - E[T]$.

**IBD in Jewish populations:** Our population ordering by ROH and IBD accords with previous studies in Jewish populations (Atzmon et al., 2010; Campbell et al., 2012; Waldman et al., 2016a,b). Kang et al. (2016) observed that their ordering of populations by mean total ROH lengths was similar to the ordering reported by Waldman et al. (2016b). We find that the ordering of mean total IBD length in the data of Kang et al. (2016) is also similar to that of Waldman et al. (2016b). For the populations included in both studies, Waldman et al. (2016b) reported, in decreasing order, Mumbai, Cochin, Iranian, Libyan, Italian, Iraqi, Tunisian, Georgian, Yemenite, Syrian, Ashkenazi, Moroccan, Algerian, and Sephardi. Here we find a similar ordering: Mumbai, Cochin, Iranian, Libyan, Georgian, Moroccan, Ashkenazi, Yemenite, Iraqi, Italian, Tunisian, Algerian, and Sephardi. Although some specific rankings differ, South Asian Jewish populations generally share the most IBD, followed primarily by some of the Middle Eastern and North African Jewish populations, with European Jewish populations tending toward intermediate and lower levels.

From our model, we expect ROH and IBD to be correlated because $E[T]$ and $E[V]$ both depend on consanguinity. Because $E[T] < E[V]$, we expect a stronger effect of consanguinity on ROH than on IBD. Indeed, we find that ROH and IBD are correlated with positive regression slope less than 1, reflecting the weaker effect of consanguinity on IBD. In particular, the correlation is strongest with Class C (long) ROH, though Classes A and B might produce larger correlations if IBD calling for short segments was more accurate. Long ROH in a population reflect consanguinity because long segments are the most likely to share a recent ancestor (Pemberton et al., 2012; Kang et al., 2016); the correlation between Class C ROH and IBD supports the prediction of our model that ROH and IBD are correlated because they are both amplified by consanguinity.

**Limitations and extensions:** Our analysis has a number of limitations. First, we assumed a constant population size and a constant fraction of consanguineous unions each generation. It might be possible to generalize these assumptions to accommodate temporal changes in population size and consanguinity that could affect ROH and IBD distributions. We did not consider population substructure, which is potentially relevant if consanguinity is practiced as a culturally transmitted trait in subgroups of a population. Substructure would affect within- and between-individual coalescence probabilities and, in turn, coalescence times. In the same manner that inbreeding and substructure can be viewed as forms of the same general phenomenon of deviation from random mating, it is possible that a structured population in which random mating occurs within subpopulations but not between them could produce similar phenomena to those we have seen in our consanguinity model.

Second, we focused only on neutral loci. Loci experiencing balancing selection can exhibit evidence of excess genetic differences for pairs of alleles sampled within individuals compared to that seen between individuals, so that a reverse effect $E[T] > E[V]$ might be observed. For example, for the HLA locus, Robertson et al. (1999) studied identity of haplotypes for haplotypes in the same individual and for haplotypes in different individuals, quantities expected to be inversely related to pairwise coalescence times under a neutral model. In a population with no first cousin and closer matings, they found an excess in the number of within-individual versus between-individual haplotype differences compared to a neutral prediction, suggesting an increase in within-individual versus between-individual coalescence times at the HLA locus. This result, which contrasts with our prediction of greater difference for between-individual comparisons,
suggests that caution is warranted in interpreting ROH and IBD with our model for regions experiencing balancing selection.

A third limitation is that for $n$th cousin mating, we assumed $n \ll N$. However, in practical scenarios, if $n$ is large, then randomly mating pairs are related to some degree, and if $N$ is small, then double first cousin mating is non-negligible. It might therefore be unrealistic to consider $n$ large in our model. A fourth limitation is that we did not examine the full distributions of $T$ and $V$; further information about these distributions will be important for clarifying the theoretical relationship between ROH and IBD more precisely.

The same approach we took here can also be applied to other consanguinity regimes. Double first cousins, for example, have twice the number of paths to a recent common ancestor as first cousins. In eq. 7, $\frac{3}{4}$ is the probability that two alleles in a mating pair are inherited from a shared grandparental mating pair. If instead we consider double first cousins, and if $d_1$ is the fraction of double first cousin mating pairs, then $\frac{7}{4}$ is the probability that two alleles in the two individuals of a mating pair are inherited from a shared grandparental mating pair. Substituting $\frac{3}{4}$ for $\frac{7}{4}$ in eq. 7 gives a computation for double first cousin mating.

Our model has implications for empirical studies of ROH and IBD. Studies have used properties of IBD for inferring effective population size and those of consanguinity. One method for distinguishing between the effects of small population size and those of consanguinity is to examine the relationship between the number and length of ROH segments (Ceballos et al., 2018). Our results suggest that examining the number and length of IBD segments could also assist in disentangling these effects, as such features of IBD segments are also affected by consanguinity.

The reduction we observed in coalescence times owing to consanguinity implies that assuming random mating when inferring effective population size may produce underestimates. Under random mating, the mean coalescence time is $4N$, whereas we find that with consanguinity, it is $4N(1 - 3c)$. Thus, in populations with consanguinity, an apparent estimate of $4N$ might actually be an estimate of $4N(1 - 3c)$. Lastly, our finding that $E[T]$ and $E[V]$ depend on the population size $N$ and the kinship coefficient $c$ suggests that given the full distributions of these random variables, it may be possible to infer $N$ and $c$ from joint analysis of ROH and IBD sharing.

We have introduced a model for the simultaneous analysis of ROH and IBD, finding that both are driven by the same phenomena of consanguinity and reduction in effective population size. ROH and IBD have often been analyzed separately, with different motivations and techniques. Our results provide a formal connection between ROH and IBD, demonstrating the utility of considering them together in the same analysis.

Acknowledgments. We thank J. Kang for bioinformatics assistance. Support was provided by NIH grant R01 HG005855, US-Israel Binational Science Foundation grant 2017024, and by a National Science Foundation Graduate Research Fellowship.

References

ATZMON, G., L. HAO, I. PE’ER, C. VELEZ, A. PEARLMAN, et al., 2010 Abraham’s children in the genome era: Major Jewish diaspora populations comprise distinct genetic clusters with shared Middle Eastern ancestry. American Journal of Human Genetics 86: 850–859.

BROWNING, B. L., and S. R. BROWNING, 2013 Improving the accuracy and efficiency of identity by descent detection in population data. Genetics 194: 459–471.

BROWNING, S. R., and B. L. BROWNING, 2007 Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. American Journal of Human Genetics 81: 1084–1097.

BROWNING, S. R., and B. L. BROWNING, 2012 Identity by descent between distant relatives: Detection and applications. Annual Review of Genetics 46: 617–633.

BROWNING, S. R., and B. L. BROWNING, 2015 Accurate non-parametric estimation of recent effective population size from segments of identity by descent. American Journal of Human Genetics 97: 404–418.

CAMPBELL, C. L., P. F. PALAMARA, M. DUBROVSKY, L. R. BOTIGUÉ, M. FELLOUS, et al., 2012 North African Jewish and non-Jewish populations form distinctive, orthogonal clusters. Proceedings of the National Academy of Sciences USA 109: 13865–13870.

CAMPBELL, R. B., 2015 The effect of inbreeding constraints and offspring distribution on time to the most recent common ancestor. Journal of Theoretical Biology 382: 74–80.

CARMI, S., P. R. WILTON, J. WAKELEY, and I. PE’ER, 2014 A renewal theory approach to IBD sharing. Theoretical Population Biology 97: 35–48.

CEBALLOS, F. C., P. K. JOSHI, D. W. CLARK, M. RAMSAY, and J. F. WILSON, 2018 Runs of homozygosity: windows into population history and trait architecture. Nature Reviews Genetics 19: 220–234.

CHARLESWORTH, D., 2003 Effects of inbreeding on the genetic diversity of populations. Philosophical Transactions of the Royal Society of London B 358: 1051–1070.

HARRIS, K., and R. NIELSEN, 2013 Inferring demographic history from a spectrum of shared haplotype lengths. PLoS Genetics 9: e1003521.
Kang, J. T. L., 1970 The Genetic Structure of Populations. Springer-Verlag, New York.

Jacquard, A., 1972 Genetic information given by a relative. Biometrics 28: 1101–1114.

Kang, J. T. L., A. Goldberg, M. D. Edge, D. M. Behar, and N. A. Rosenberg, 2016 Consanguinity rates predict long runs of homozygosity in Jewish populations. Human Heredity 82: 87–102.

Karafet, T. M., K. B. Bulayeva, O. A. Bulayev, F. Gurgenova, J. Omarova, et al., 2015 Extensive genome-wide autozygosity in the population isolates of Daghestan. European Journal of Human Genetics 23: 1405–1412.

King, L., J. Wakeley, and S. Carmi, 2018 A non-zero variance of Tajima’s estimator for two sequences even for infinitely many unlikely loci. Theoretical Population Biology 122: 22–29.

Kirin, M., R. McQuillian, C. S. Franklin, H. Campbell, P. M. McKeigue, et al., 2010 Genomic runs of homozygosity record population history and consanguinity. PLoS One 5: e13996.

Lange, K., 1997 Mathematical and Statistical Methods for Genetic Analysis. Springer, New York.

Milligan, B. G., 1996 Estimating long-term mating systems using dna sequences. Genetics 142: 619–627.

Nordborg, M., and P. Donnelly, 1997 The coalescent process with selection. Genetics 146: 1185–1195.

Palamara, P. F., T. Lencz, A. Darvasi, and I. Pe’er, 2012 Length distributions of identity by descent reveal fine-scale demographic history. American Journal of Human Genetics 91: 809–822.

Pemberton, T. J., D. Absher, M. W. Feldman, R. M. Myers, N. A. Rosenberg, et al., 2012 Genomic patterns of homozygosity in worldwide human populations. American Journal of Human Genetics 91: 275–292.

Pollak, E., 1987 On the theory of partially inbreeding finite populations. I. partial selfing. Genetics 117: 353–360.

Ralph, P., and G. Coop, 2013 The geography of recent genetic ancestry across Europe. PLoS Biology 11: e1001555.

Robertson, A., D. Charlesworth, and C. Ober, 1999 Effect of inbreeding avoidance on hardy-weinberg expectations; Examples of neutral and selected loci. Genetic Epidemiology 17: 165–173.

Scott, E. M., A. Haleys, Y. Itan, E. G. Spencer, Y. He, et al., 2016 Characterization of greater middle eastern genetic variation for enhanced disease gene discovery. Nature Genetics 48: 1071–1076.

Sjödin, P., I. Kaj, S. Krone, M. Lascoux, and M. Nordborg, 2005 On the meaning and existence of an effective population size. Genetics 169: 1061–1070.

Thompson, E. A., 2013 Identity by descent: Variation in meiosis, across genomes, and in populations. Genetics 194: 301–326.

Wakeley, J., 2009 Coalescent Theory: An Introduction. Roberts & Company, Greenwood Village, CO.

Wakeley, J., L. King, B. S. Low, and S. Ramachandran, 2012 Gene genealogies within a fixed pedigree, and the robustness of Kingman’s coalescent. Genetics 190: 1433–1445.

Wakeley, J., L. King, and P. R. Wilton, 2016 Effects of the population pedigree on genetic signatures of historical demographic events. Proceedings of the National Academy of Sciences USA 113: 7994–8001.

Waldman, Y. Y., A. Biddanda, N. R. Davidson, P. Billing-Ross, M. Dubrovsky, et al., 2016a The genetics of Bene Israel from India reveals both substantial Jewish and Indian ancestry. PLoS ONE 11: e0152056.

Waldman, Y. Y., A. Biddanda, M. Dubrovsky, C. L. Campbell, C. Oddoux, et al., 2016b The genetic history of Cochin Jews from India. Human Genetics 135: 1127–1143.

Wilton, P. R., P. Baduel, M. Landon, and J. Wakeley, 2017 Population structure and coalescence in pedigrees: Comparisons to the structured coalescent and a framework for inference. Theoretical Population Biology 115: 1–12.

Wollenberg, K., and J. C. Avise, 1998 Sampling properties of genealogical pathways underlying population pedigrees. Evolution 52: 957–966.

Woods, C. G., J. Cox, K. Springell, D. J. Hampshire, M. D. Mohamed, et al., 2006 Quantification of homozygosity in consanguineous individuals with autosomal recessive disease. American Journal of Human Genetics 78: 889–896.