EFFICIENT SIMULATION OF IBD SPECTRA IN INBRED POPULATIONS USING NETWORK CONVOLUTION

TIAN CHEN ZENG **CS 224W** **Dec 9, 2018**
https://github.com/tczeng/CS-224W

INTRODUCTION

Consanguineous (close-kin) marriage is common in humans, and is or was the preferred or even ideal form of marriage in many societies at various levels of social complexity (Walker and Bailey 2014; A. Korotayev 2000). The discovery of the rates and types of consanguineous marriage in ancient societies using ancient DNA is an exciting prospect. In this project, I will develop an algorithm for the rapid simulation of IBD (identity-by-descent) and ROH (Runs-Of-Homozygosity) spectra given some population parameters (inbreeding of type i with frequency *p* of all marriages) using network methods, which will represent an implementation of some ways, that I developed in a previous project, to distinguish the types and rates of consanguineous marriage using genetic data alone.[1] In particular, my algorithm will rely on the intuition that the accumulation of coefficients of relationship between pairs of persons in a population over many generations through shared common ancestors is equivalent to network convolution, which in this case will take the form of repeated multiplication of sparse matrices.

The majority of consanguineous marriages are cousin marriages. By the demographer's definition (unions between second cousins and closer), such marriages account for ~25% of all contracted in South India (Andhra Pradesh, Telangana, Tamil Nadu, Karnataka) (A. H. Bittles et al. 1987; Richards 1914; A. H. Bittles, Coble, and Appaji Rao 1993), and up to 50% in large parts of the Middle East (Donnan and Holy 1990). A comprehensive taxonomy of consanguineous marriages is provided by Thornhill (1990), but for this investigation, we will be prioritising *first cousin marriages*. As defined from the perspective of a male ego, there are four types of first cousin marriage:

1. Father's sister's daughter (FZD)
2. Mother's brother's daughter (MBD)
3. Father's brother's daughter (FBD)
4. Mother's sister's daughter (MZD)

---

[1] https://docs.google.com/document/d/1R73aBVj8g5v9NbeNGfaiiBzye_BIUIyAADrJiQnJXqw/edit#
Some parts of the text of this document outside the methods and the results section are summaries of work in the link. Appendix 1 and 2 are presentations of work I did in this document.

The widespread occurrence of cousin marriage renders it a significant consideration in public health, and influences the epidemiology of many genetic disorders (A. H. Bittles 2007; Alan H. Bittles, n.d.). This testifies to the significant evolutionary cost of the practice, as inbreeding accumulates over time as a function of the number of consanguineous unions in a given pedigree. The varied distribution of the occurrence, level of preferredness, and sociocultural correlates of cousin marriage have inspired much research into its causes and consequences, covering various aspects of kin selection theory, optimal outbreeding theory (Thornhill 1990), and theories in economics, anthropology and political science regarding the effects of social structure on cooperative behaviours (Edlund 2018). The sociocultural correlates of cousin marriage include kinship systems (Murdock 2015), the security situation and intensity of feuding (Kang 1976; Otterbein 1968), presence of honour and revenge-based culture (Shaw 2001), types of social cooperation and institutional development (Enke 2017; Schulz 2016), level of political complexity and degree of market penetration (A. V. Korotayev 2003), and geographic mobility (Fuster and Colantonio 2004). For all these reasons and more, it may be useful to develop methods to infer the quantity and types of consanguineous marriages in a population's history from genetic data alone, with either ancient or modern DNA as the data source. This will open an unprecedented window into the social structure and dynamics of historic and prehistoric peoples.

I. PREVIOUS WORK

A pedigree constitutes a large directed acyclic graph (DAG), and, with complete outbreeding, this pedigree does not contain any cycles. Each individual would then be the root of a binary tree that expands backwards and doubles in size per generation. An inbreeding coefficient specifies the probability that a person has inherited two copies of the same allele in an ancestor in his pedigree, made possible by the fact that the ancestor was a progenitor of *both the individual's mother and father*. In other words, it specifies the probability that both alleles at any given locus in an individual are identical-by-descent (IBD) or *autozygous* from the fact that copies from a common ancestor were inherited through both the maternal and paternal sides. When the individual has a father and mother sharing an ancestor, a cycle, known as a *pedigree collapse,* develops in the pedigree. Inbreeding coefficients for a given individual I are given by the following formula:

$$F_I = \sum_A (\frac{1}{2})^n (1 + F_A) \quad (1)$$

Where A is a common ancestor of ego's father and mother, n is the number of nodes in the cycle in the pedigree excluding ego, and F_A is the inbreeding coefficient of A. This definition is provided by Sewall Wright (Wright 1922). An equivalent definition is provided by Malécot (1970), using the following system of recursive equations, where I is ego, X and Y are its parents, and H another individual who is not directly related to I:

$$F_I = F_{XY} \quad (2a)$$

$$F_{II} = \frac{1}{2}(1 + F_{XY}) \quad (2b)$$

$$F_{HI} = \frac{1}{2}(F_{HX} + F_{HY}) \quad (2c)$$

Note that Malecot's equations generalise to provide a measure of coancestry, or IBD between two individuals (F_{XY}). More information is provided on Malecot's equations, including their application on example pedigrees, in Appendix 1.

At first glance, there are very few differences in terms of patterns of genetic variation produced by variation in cousin marriage type. However, that does not mean there are none. Over the summer while working in Joseph Henrich's lab, I developed two methods in this work[2], one of which involves comparing X-chromosome vs Autosome inbreeding levels, and another of which involves an observation about the way different types of inbreeding would affect *uniparental subpopulations*. The X-chromosomal method is relatively straightforward and analytically soluble, and in fact can be simulated using the algorithm I will present, with some trivial changes. The method of uniparental subpopulations, on the other hand, is intuitive and interesting, but analytically intractable. This presents an opportunity to use network methods for simulation. I present it below.

ELEVATED COANCESTRY WITHIN UNIPARENTAL SUBPOPULATIONS

Uniparental markers are passed along a single sex--males for the Y-chromosome, and females for mitochondrial DNA. Pairs of members of *uniparental subpopulations* (populations having identical Y-chromosomes or mtDNA sequences) have a nonzero probability of sharing an ancestor of the same sex as themselves along a *single-sex pedigree*. Under some combinations of frequencies of cousin marriage types, the average coancestry between males with the same Y chromosome, or males and females with the same mitochondrial DNA sequence, will exceed the average coancestry of two randomly-drawn members of the population. This occurs because the aforementioned elevated probability of sharing an ancestor along a single-sex pedigree, plus the fact that the relative abundance of cousin marriages on single-sex pedigrees versus an average pedigree is affected by the prevalence of each cousin marriage types. The reason why this happens is explained in Figure 1, below.

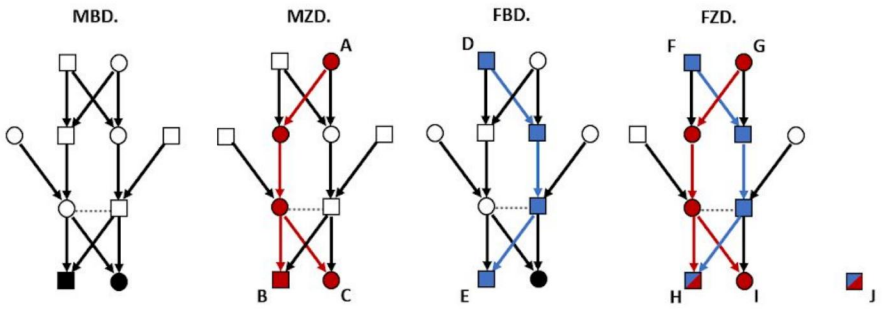


FIGURE 1 | The individual J at bottom right is identical-by-descent uniparentally with D and F in the Y-chromosome (represented by the blue color) and A and G in mtDNA (represented by red).

At left, no single-sex chains result from MBD marriage. However, MZD and FZD provide a chain of females (marked in red, terminating in B, C and H, I respectively) while FBD and FZD provide a chain of males (marked in blue, terminating in E and H respectively). Notice that, as a rule, the individual at the end of the chain invariably shares both elevated autosomal IBD and uniparental IBD with the individual at the beginning, and thus by extension with J as well. Note that, unlike daughters, who do

[2]The Henrich Lab in the Harvard Department of Human Evolutionary Biology

not inherit their Y-chromosome from their fathers and cannot pass it on, sons do inherit their mtDNA from mothers even if they cannot transmit it, and thus a male terminating a chain of females nevertheless shares uniparental ancestry with the female progenitor of the chain.

An expression for the expected value of excess relatedness between pairs of persons drawn from a uniparental subpopulation, relative to a randomly-drawn pair of persons, is given by the equation below, where J_n^i is the increase in average relatedness between the random pair and K_n^i the increase in average relatedness between the uniparental pair, S is the rate of cousin marriage, and f^i is the *i*th Fibonacci number.

$$\frac{K_i^n}{J_i^n} = \prod_{i=1}^{n} (f^{i+2})^{\sum_{x=1}^{n} x \sum_{k=0}^{n} (\sum_{y=x-1}^{n-ik} \frac{1-S^y}{2(1-2S)^y})^2}$$

Table 1 describes the pattern of relatedness resulting from the occurrence of each type of cousin marriage, relative to the relatedness of a random pair in a population without inbreeding. The derivation of this expression is very complex and presented in Appendix 2.

	Random Pair	Pair from Y-chromosome Uniparental subpopulation	Pair from Mitochondrial Uniparental subpopulation
No Cousin Marriage	1	$1 + P(Y_{IBD}) \cdot e$	$1 + P(mtDNA_{IBD}) \cdot e$
FBD	J^n_i	$J^n_i + P(Y_{IBD}) \cdot e \cdot K^n_i$	$J^n_i + P(mtDNA_{IBD}) \cdot e$
FZD	J^n_i	$J^n_i + P(Y_{IBD}) \cdot e \cdot K^n_i$	$J^n_i + P(mtDNA_{IBD}) \cdot e \cdot K^n_i$
MZD	J^n_i	$J^n_i + P(Y_{IBD}) \cdot e$	$J^n_i + P(mtDNA_{IBD}) \cdot e \cdot K^n_i$
MBD	J^n_i	$J^n_i + P(Y_{IBD}) \cdot e$	$J^n_i + P(mtDNA_{IBD}) \cdot e$

TABLE 1 | Amount of inbreeding for each type of pair, relative to a random pair from a population with no cousin marriage.

While the complex equation relating K and J seems to suggest that K is much larger than J, and therefore that the types of cousin marriage should be separable using the rules presented in Table 1, there is no guarantee that this is so. The simulation attempted here is an attempt at solving this issue.

METHODS

PROOF THAT NETWORK CONVOLUTION IS EQUIVALENT TO THE CALCULATION OF INBREEDING COEFFICIENTS ON LARGE PEDIGREES

Define the parent-child matrix for generations 0 and 1:

$$M_{ij}^{1,0} = \begin{cases} 0 & \text{if j is not a parent of i} \\ 1 & \text{if j is a parent of i} \end{cases} \quad (1)$$

Where i is from generation 1 and j is from generation 0. This matrix then is an adjacency matrix of a bipartite network involving parents and children from generations 0 and 1. I will prove that the matrix of relatedness coefficients of generation G^t, at generation t, is given by equation (1):

$$\begin{aligned}
G^t &= \frac{1}{4}M^{(t,t-1)}G^{t-1}(M^{(t,t-1)})^T \\
&= (\frac{1}{4})^2 M^{(t,t-1)}M^{(t-1,t-2)}G^{t-2}(M^{(t-1,t-2)})^T(M^{(t,t-1)})^T \\
&= (\frac{1}{4})^3 M^{(t,t-1)}M^{(t-1,t-2)}M^{(t-2,t-3)}G^{t-3}(M^{(t-2,t-3)})^T(M^{(t-1,t-2)})^T(M^{(t,t-1)})^T \\
&= \dots \\
&= (\frac{1}{4})^t M^{(t,t-1)}M^{(t-1,t-2)}\dots M^{(0,1)}G^0(M^{(0,1)})^T\dots(M^{(t-1,t-2)})^T(M^{(t,t-1)})^T
\end{aligned}$$

Where G^0 is a matrix of relatedness of the members of the first generation within the pedigree, and t is the total number of generations in the pedigree. This equation, (1), is equivalent to stating that the relatedness matrix of the final generation of a large pedigree may be calculated from the relatedness matrix of the first generation multiplied by the parent-child matrices of all intervening generations.

We will prove this through induction. First assume that we have a pedigree of only 2 generations; this means that there is only one parent-child matrix available, and t = 1. Notice that the coefficient of relatedness of any two persons X and Y from generation 1, with parents AB, and CD, respectively, is given by the following, as derived from Malecot's equations:

$$\begin{aligned}
G^1_{XY} &= \frac{1}{2}(F_{XC} + F_{XD}) \\
&= \frac{1}{4}(F_{AC} + F_{AD} + F_{BC} + F_{BD}) \\
&= \frac{1}{4}(g^0_{AC} + g^0_{AD} + g^0_{BC} + g^0_{BD})
\end{aligned}$$

This is a simple expansion using Malecot's second equation, with the terms of the parent-child matrix $M^{(1,0)}$ substituted in. Now note the following:

$$\begin{aligned}
\frac{1}{4}(M^{(1,0)}G^0(M^{(1,0)})^T)_{XY} &= \frac{1}{4}\sum_{j=1}^{l}(\sum_{k=1}^{n} m^{(1,0)}_{Xk}g^0_{kj})m^{(1,0)}_{Yj} \\
&= \frac{1}{4}\sum_{j=1}^{l}(m^{(1,0)}_{XA}.g^0_{Aj} + m^{(1,0)}_{XB}.g^0_{Bj})m^{(1,0)}_{Yj} \\
&= \frac{1}{4}(m^{(1,0)}_{XA}.g^0_{AC}.m^{(1,0)}_{YC} + m^{(1,0)}_{XB}.g^0_{BC}.m^{(1,0)}_{YC} + m^{(1,0)}_{XA}.g^0_{AD}.m^{(1,0)}_{YD} + m^{(1,0)}_{XB}.g^0_{BD}.m^{(1,0)}_{YD}) \\
&= \frac{1}{4}(g^0_{AC} + g^0_{AD} + g^0_{BC} + g^0_{BD}) \\
&= \frac{1}{4}(F_{AC} + F_{AD} + F_{BC} + F_{BD}) \\
&= F_{XY} \\
&= G^1_{XY}
\end{aligned}$$

This proves our base case, that $G^1 = 1/4(M^{(1,0)}G^0(M^{(1,0)})^T)$. Now, for the recursive case, note that, by the same sequence of steps as above,

$$G_{XY}^{p+1} = \frac{1}{4}(M^{(p+1,p)}G^p(M^{(p+1,p)})^T)_{XY}$$
$$= \frac{1}{4}(M^{(p+1,p)}G^p(M^{(p+1,p)})^T)_{XY}$$

This proves that equation (1) is true for all values of t where t is a natural number. Therefore repeated matrix multiplication of adjacency matrices of bipartite parent-child networks is equivalent to the calculation of inbreeding coefficients for a large population given its complete pedigree.

A LARGE PEDIGREE CAN BE REPRESENTED BY A SEQUENCE OF PARENT-CHILD MATRICES

Given a large pedigree, a sequence of parent-child matrices will completely represent the pedigree, as each parent-child matrix is simply an adjacency matrix of a pair of generations, and the combination of adjacency matrices for all generations is sufficient to capture all genealogical relationships. A corollary of this is that the transmission of uniparental markers is very easy to model using the sequence of parent-child matrices. In fact, the following equation represents the uniparental groups to which each member of the population belongs to:

$$Y^t = M^{(t,t-1)}I^s M^{t-1,t-2)}I^s M^{(t-2,t-3)} \dots M^{(2,1)}I^s M^{(1,0)}I^s Y^0$$

Where Y^t represents the uniparental haplogroup membership (represented by an integer) of each person in the last generation, and I^s is a matrix of zeros with a block along the diagonal filled with the identity matrix representing the quadrant where father-son (S = M) or mother-daughter (S = F) relationships are captured in the parent-child matrix. This is equivalent to passing flow through the pedigree along relationships with particular properties, and has many potential applications outside of the one presented here.

DESCRIPTION OF THE ALGORITHM

The algorithm is developed in five steps, which I present in point-form below:

1. Generation of a single matrix, G^0, representing the base level of relatedness between randomly-drawn members of the population in the first generation. Generation of random parent-child matrices to populate each generation of the pedigree. This step involves the pairing of men and women in the parent generation through a polygamy funcion, and the assignment of a fitness parameter to each marriage, once again drawn from a function, to calibrate fecundity. For simplicity's sake, the results presented here are carried out in a population with no polygamy, full marriage, and uniform fecundity (children per person = 2).
2. To capture the products of cousin marriage in the parent-child pedigree, the assignment of marriages is changed. Given a particular target frequency of cousin marriages of a certain type, all cousin marriages of that type that occur in the random marriage graph are accounted for. Then rewiring of marriage partners, i.e. the introduction of cousin marriages of a particular type, takes place, with the newly partnerless pair left over from the new cousin marriage connected to each other. As the marriages here take place between cousins, parent-child matrices from previous generations are used to discover eligible pairs of cousins.

As target frequencies of FBD, MZD, FZD and MBD marriages are reached sequentially, certain numbers of off-target cousin marriages (that form due to the pair left over during the rewiring) may be introduced. However, because the probability of off-target cousin marriages is equivalent to the probability that a random pair of persons in the population are cousins, the probability of an off-target cousin marriage taking place falls dramatically as the total population size in the simulation increases, and is not a significant consideration for our purposes (N = 500).

3. The relationship coefficients are calculated, and a sequence of parent-child matrices representing the entire pedigree is produced. Uniparental markers are passed through the pedigree through the flow process described previously.
4. Statistics for uniparental relatedness and random-pair relatedness are calculated using the final uniparental marker configuration.

The operation of the algorithm has an intuitive visual representation when successive generations of the matrix G are plotted, as seen below:

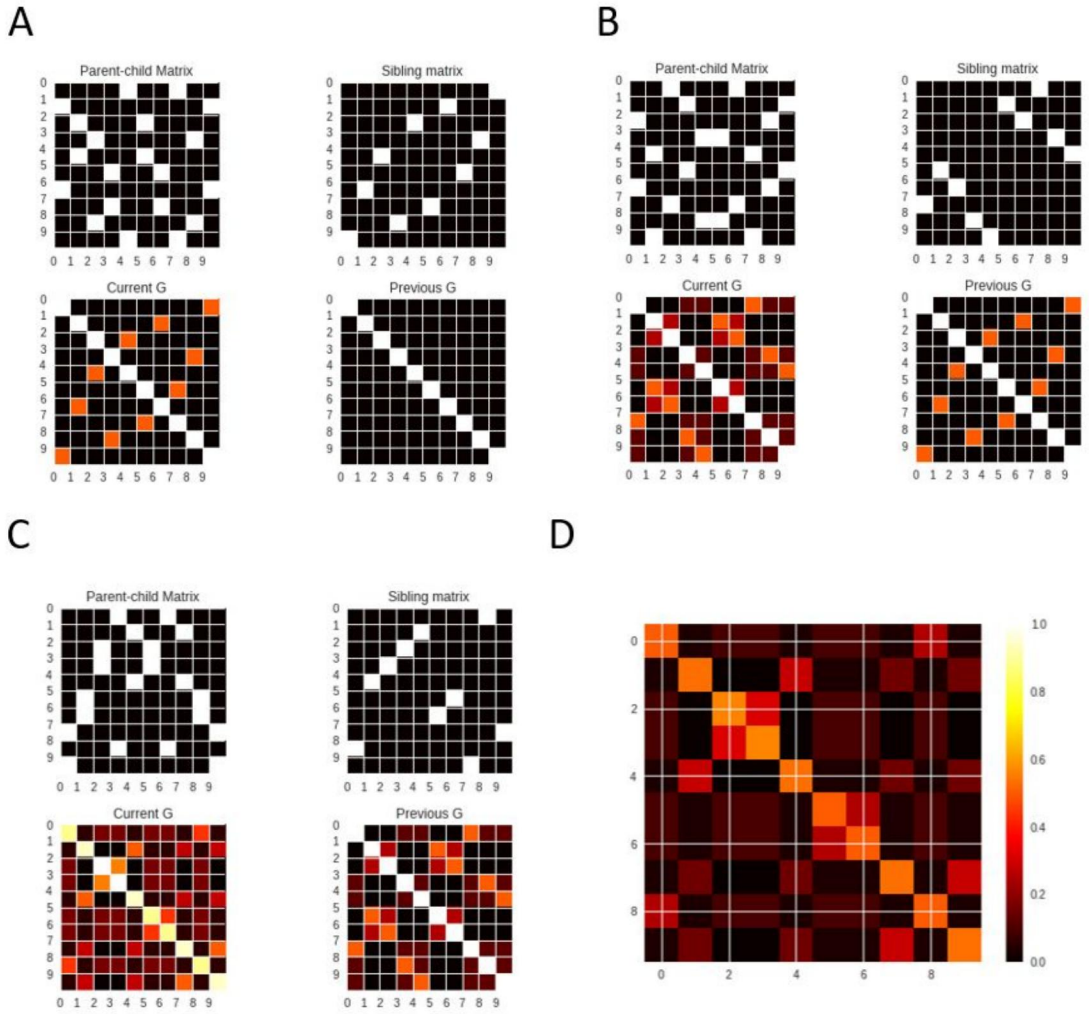


FIGURE 2 | A, B and C contain the parent-child matrices and relatedness matrices for generations 0, 1 and 2 respectively. The sibling matrix of generation 0 is used to discover eligible cousin pairs in generation 2 for the creation of FBD marriages in this scenario. Notice the reflection of sibling-relatedness in the relatedness matrix G of each current generation, as well as the effects of previous relatedness matrices on the later generations of G.

RESULTS

FINITE POPULATION EFFECTS PREVENT THE SIMULATION OF EXCESSIVELY LARGE NUMBERS OF GENERATIONS

Careful readers and users of the algorithm will realise that the inbreeding coefficients produced by the algorithm is not equal to the equilibrium inbreeding coefficients derived in Appendix A. This is due to *finite population effects* (Crow 1970), which cause random pairs from successive generations of a small population to become more and more closely related to each other, due to increasing numbers of pedigree collapses as the number of ancestors increases backwards in time. This causes the signal introduced by our specified level of cousin marriage to be drowned out if the number of generations is too high. To simulate a historically realistic timeline of cousin marriage practice, as well as to prevent finite population effects from overtaking the algorithm, the number of generations I use in later simulations does not exceed 50, which translates, using a generation time of 25 years, to 1250 years. For some sense of scale, cousin marriage set in as a practice in the Middle East only with the advent of Islam, after the 7th century AD (A. V. Korotayev 2003).

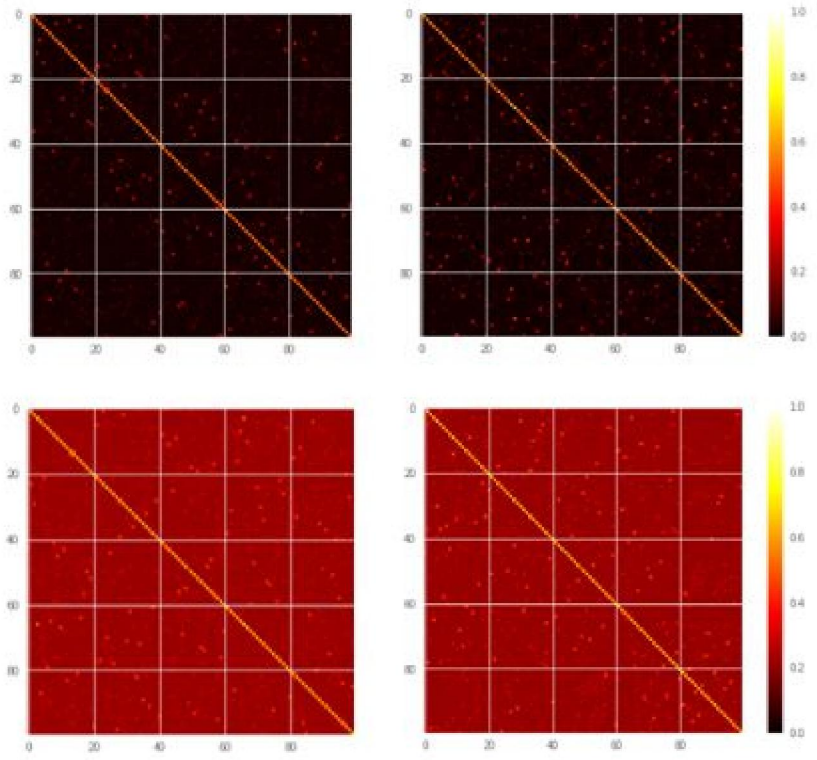


FIGURE 3 | At top right, we have a pedigree of 10 generations with no inbreeding. The average inbreeding coefficient is 0.03449, and the inbreeding coefficient of the Y-chromosomal subpopulations is 0.0594. With the imposition of 10% FBD marriage, seen at top right, the average inbreeding is 0.03426 and the inbreeding coefficient of the Y-chromosomal subpopulations is 0.7421--a noticeable increase. At bottom right, we have a pedigree of 100 generations with no cousin marriage. As can be seen, extensive pedigree collapse causes elevated background

relatedness among random pairs. Here the average inbreeding coefficient is 0.23064. With the imposition of 10% FBD cousin marriage, it is 0.2265--variation caused by differences in the background stochastic pedigree collapse swamp differences imposed by inbreeding. Likewise, the inbreeding coefficient of the Y-chromosomal subpopulations does not increase.

ASYMMETRIC PARALLEL-COUSIN MARRIAGES STRONGLY INCREASE UNIPARENTAL RELATEDNESS RELATIVE TO RELATEDNESS OF RANDOM PAIRS

In general terms, the method I proposed works very well, at least in indicating the presence of cousin marriage of the parallel-cousin type (FBD and MZD). Surprisingly, the metric that works the most well is the ratio of uniparental relatedness to the relatedness of random pairs. This indicates that parallel cousin marriage introduces increased relatedness in the uniparental subpopulations of the marker that is not implicated in uniparental pedigrees for that type of cousin marriage (i.e. FBD and mtDNA, and MZD and Y-chromsomes), in some way. This motivates me to look more carefully at this problem.

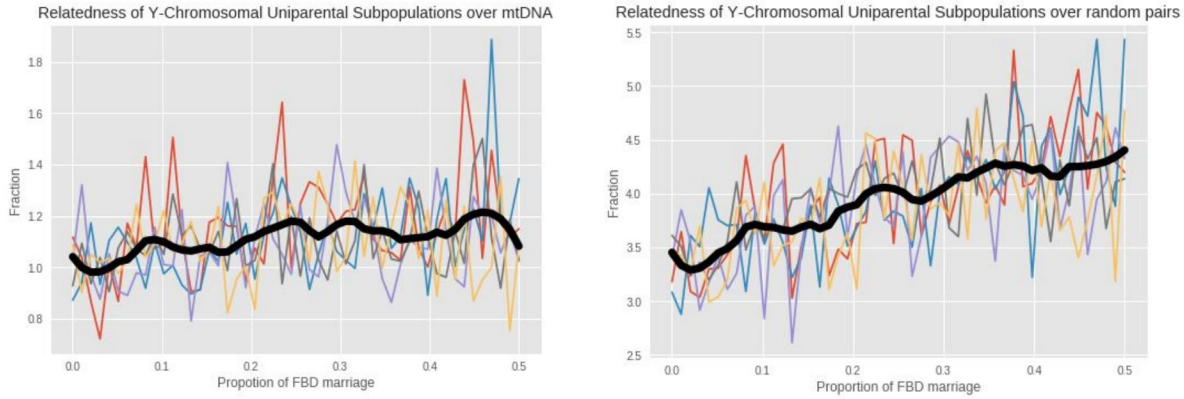


FIGURE 4 | The ratio of relatedness of Y-chromosomal subpopulations to the relatedness of mtDNA subpopulations increases with increasing fraction of FBD marriage, seen at left. The ratio of relatedness of Y-chromosomal subpopulations to random pairs increases even more rapidly, as seen at right. The individual colored trajectories represent single simulations.

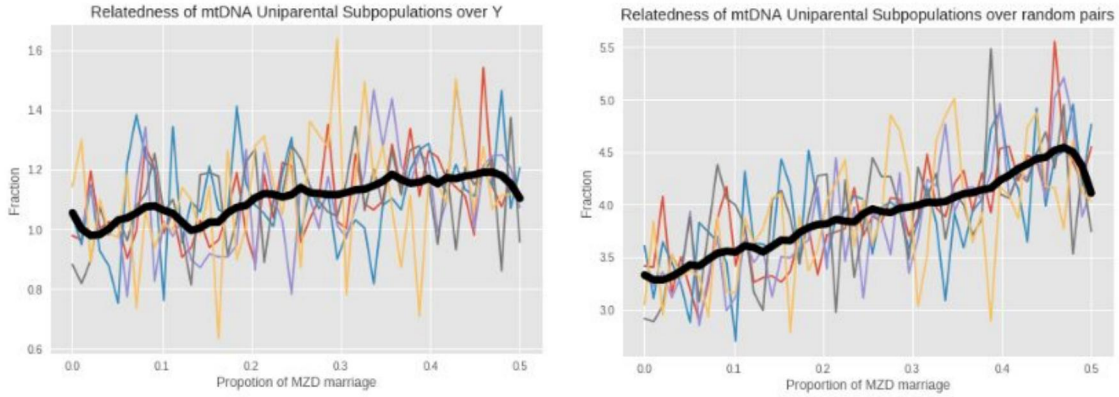


FIGURE 5 | The ratio of relatedness of mitochondrial subpopulations to the relatedness of Y-chromosomal subpopulations increases with increasing fraction of FZD marriage, seen at left. The ratio of relatedness of mitochondrial subpopulations to random pairs increases even more rapidly, as seen at right. The individual colored trajectories represent single simulations.

THE PROBABILITY OF RECENT UNIPARENTAL IBD IS ALWAYS LARGE

Notice also that the baseline excess relatedness of the uniparental subpopulations over the randomly selected pairs is already large. This tells us that the $P(Y_{IBD})$ and $P(mtDNA_{IBD})$ terms in TABLE 1 are larger in naturalistic populations. For more explanation, see Appendix 2.

GREATER RESOLUTION FOR UNIPARENTAL MARKERS OR GREATER DIVERSITY IN UNIPARENTAL MARKERS STRONGLY INCREASE THE STRENGTH OF THE SIGNAL BY INCREASING PROBABILITY OF UNIPARENTAL IBD

The high variance of the method, manifested in the large spread of the plotlines of each single simulation in figure 5, is reduced when the number of uniparental clades in the population is larger; i.e., that $P(Y_{IBD})$ and $P(mtDNA_{IBD})$ begin from a higher baseline. When the number of clades is 5, the inbreeding coefficient of Y-chromosomal subpopulations exceeded that of mitochondrial subpopulations 73% of the time in 100 simulations on a population of 500 for 10 generations, while it exceeded that of mitochondrial subpopulations 98% of the time when the simulations were repeated with the number of clades increased to 50.

This validates well a prediction I made when developing this method; a finer segmentation of the population into uniparental subpopulations increases the likelihood that any uniparentally identical pair shares a single-sex pedigree in a relatively shallow time-frame. Therefore, when the method is used in real datasets, high-resolution deep sequencing will lend the method the most power.

APPENDIX 1

Calculating Equilibrium Inbreeding Values (F) for the Autosome and X Chromosome given Cousin Marriage without Regard to Cousin Marriage Rules

In many works focusing on demography (e.g. Bittles 1987; Freire-Maia 1967), one calculates the average inbreeding coefficient of a population using the following formula:

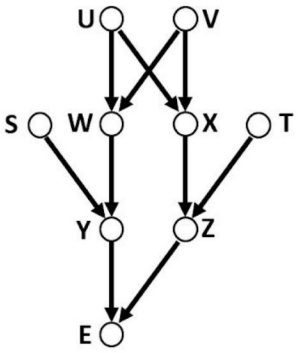
$$\alpha = \sum_{i=1}^{m} p_i f_i$$

Where p_i is the proportion of the *i*th type of inbreeding, and f_i is the inbreeding coefficient associated. While the average inbreeding approximates the equilibrium inbreeding when rates of consanguineous marriage are low, the two diverge for high values, as 1) accumulated inbreeding and 2) consanguineous marriages that simultaneously fall under two different inbreeding types are ignored. The general method for deriving equilibrium inbreeding coefficients for each type of inbreeding given some frequency of such inbreeding is given in Hedrick ((P. W. Hedrick 1986; Philip W. Hedrick and Cockerham 1986), through the analytical solution of recursive equations, using Malecot's method of coancestry.

Intuitively, the equilibrium inbreeding F_e may be conceived as the result of a Markov process involving a population of single or branched chains, possibly overlapping, that have some probability of 1) extension vs. termination and 2) spontaneous generation, where only currently unterminated chains contribute to the inbreeding coefficient. Therefore, recursions can be used to find the equilibrium inbreeding value. Hedrick gives us the following for the inbreeding of a population with partial first cousin marriage, occurring with probability *S*:

$$F_e = \frac{S}{16 - 15S}$$

I demonstrate the derivation given the pedigree below:



$$\begin{aligned}
F_E &= F_{YZ} \\
F_E &= \frac{1}{2}(F_{YX} + F_{YT}) \\
F_E &= \frac{1}{4}(F_{SX} + F_{ST} + F_{WX} + F_{WT}) \\
&= \frac{1}{4}(F_{SX} + F_{ST} + F_{WT}) + \frac{1}{8}(F_{WU} + F_{WV}) \\
F_E &= \frac{1}{4}(F_{SX} + F_{ST} + F_{WT}) + \frac{1}{16}(F_{UU} + 2F_{UV} + F_{VV}) \\
&= \frac{1}{16}(4F_{SX} + 4F_{ST} + 4F_{WT} + 2F_{UV} + F_{UU} + F_{VV})
\end{aligned}$$

$$= \frac{1}{16}(4F_{SX} + 4F_{ST} + 4F_{WT} + 2F_{UV} + \frac{1}{2}(1 + F_U) + \frac{1}{2}(1 + F_V))$$

Then, by using equation 2a in Malecot's set:

$$F_{t+3} = \frac{S}{16}(12F_{t+2} + 2F_{t+1} + 1 + F_t)$$

At equilibrium,

$$F_e = \frac{S}{16}(15F_e + 1)$$

$$F_e = \frac{S}{16 - 15S}$$

A general equation that approximates the equilibrium inbreeding value for any combination of inbreeding types is is:

$$f_e = \frac{\sum_j \frac{S_j}{2^j}}{1 - \sum S_j[1 - (\frac{1}{2})^j]}$$

Where *j* is the *degree* of relationship of the inbreeding partners, and S_j the proportion of persons involved in that inbreeding type.

APPENDIX 2

The prevalence of each cousin marriage type would affect relatedness among subpopulations carrying the same sets of uniparental markers. Uniparental subpopulations here refer to subdivisions of the population into groups with the same set of Y chromosomes for males, or mtDNA for females and males. Y chromosomes are passed from father to son, and record unilineal relationships among males only, while mitochondrial DNA is passed from mother to both daughter and son, and record unilineal relationships in mother-daughter chains or mother-son terminal relationships.

Under some combinations of frequencies of cousin marriage types, the average coancestry between males with the same Y chromosome, or males and females with the same mitochondrial DNA sequence, will exceed the average coancestry of two randomly-drawn members of the population. This occurs because the members of unilineal subpopulations have an elevated probability of sharing an ancestor along a single-sex pedigree, and the relative abundance of cousin marriages on single-sex pedigrees versus an average pedigree is affected by the prevalence of each cousin marriage types.

Individuals from Uniparental Subpopulations have a nonzero Probability of Sharing a Common Ancestor in a Single-Sex Pedigree

Two males A and B with identical Y-chromosomal haplotypes may share a male ancestor, C, with some probability, whose Y chromosome has been transmitted via two unbroken lines of males to A and B, as seen in FIGURE 2. Likewise, two persons of either sex D and E with identical mitochondrial haplotypes may share a common female ancestor F who transmitted her mitochondrial haplotype to them via two unbroken

sequences of daughters. On the other hand, a pair of persons, X and Y, without the same uniparental chromosomes cannot possibly share a common ancestor with a single-sex pedigree.

In general, randomly-drawn pairs of persons within a population of finite size will have nonzero relatedness, sharing some numbers of ancestors with pedigrees of the type exemplified by ZXY; the same is true for individuals drawn from uniparental subpopulations. However, only individuals from uniparental subpopulations have a nonzero probability of having a pedigree of the types exemplified by CAB and FDE.

.

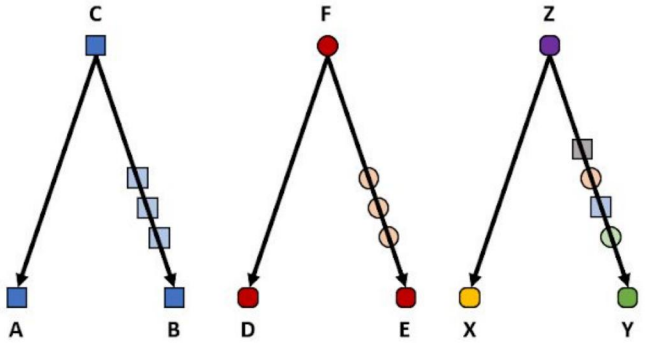


FIGURE 2 | At left, two males A and B with the same Y-chromosomal haplogroup, represented by the identical blue coloration, have some probability of sharing *one* male ancestor C some indeterminate number of generations ago, who is connected to both descendants by an uninterrupted chain of males of some length, represented here by the transparent males along the arrows. The same applies for two individuals of either sex (D and E) who share the same mtDNA sequence, at center, who are connected to *one* common female ancestor F through an uninterrupted chain of females. At right, two individuals of either sex (X and Y) who do not share any uniparental chromosomes, whether Y-chromosomes or mtDNA, may share any number of ancestors Z through pedigrees with any combination of males and females, with no preservation of uniparental chromosomes throughout the pedigree, indicated by the different colors of the individuals along the arrows.

Notice that, among pairs of persons with significant coancestries, almost all the coancestry will be expected to be due to pedigrees of the ZXY type, i.e. non-single-sex pedigrees. Indeed, among individuals with different uniparental markers, all the coancestry must be due to such pedigrees. Only individuals who have the same uniparental markers can possible share ancestry due to single-sex pedigrees from a common ancestor. We will call the sharing of uniparental chromosomes through descent from a common ancestor *uniparental IBD*, or *uniparental coancestry*.

Cousin Marriages Affect the Accumulation of Inbreeding Along Single-Sex Chains

Now notice that FBD, FZD and MZD marriages produce inbreeding coefficients in persons at the ends of single-sex chains--chains of females for MZD, chains of males for FBD, and chains of males and females for FZD--but MBD precludes this possibility (FIGURE 3).

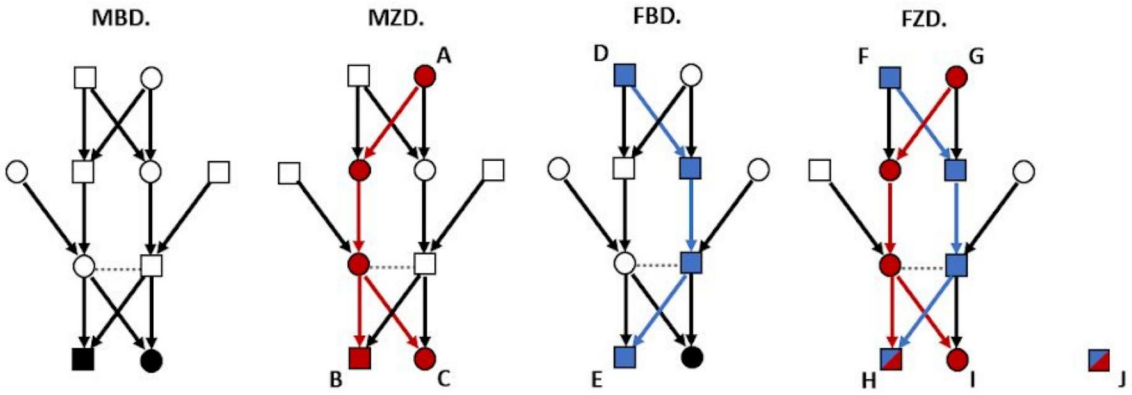
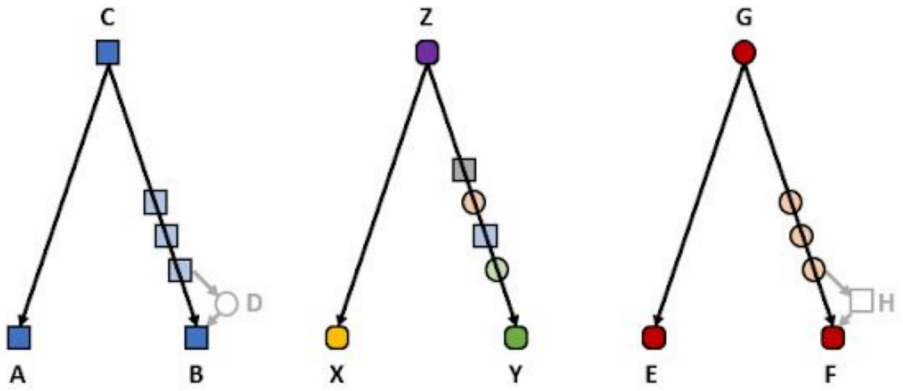


FIGURE 3 | The individual J at bottom right is identical-by-descent uniparentally with D and F in the Y-chromosome (represented by the blue color) and A and G in mtDNA (represented by red).

At left, no single-sex chains result from MBD marriage. However, MZD and FZD provide a chain of females (marked in red, terminating in B, C and H, I respectively) while FBD and FZD provide a chain of males (marked in blue, terminating in E and H respectively). Notice that, as a rule, the individual at the end of the chain invariably shares both elevated autosomal IBD and uniparental IBD with the individual at the beginning, and thus by extension with J as well. Note that, unlike daughters, who do not inherit their Y-chromosome from their fathers and cannot pass it on, sons do inherit their mtDNA from mothers even if they cannot transmit it, and thus a male terminating a chain of females nevertheless shares uniparental ancestry with the female progenitor of the chain.

Figure 3 demonstrates that FBD, FZD, and MZD marriages cause inbreeding accumulation in persons at the ends of single-sex chains. Because such persons may preserve uniparental IBD with distant relatives, their relatedness with these relatives is likewise amplified. This suggests that these three types of cousin marriage may increase the average relatedness of all persons who have uniparental coancestry. Given that all pairs of persons with uniparental coancestry are also members of the same uniparental subpopulation, FBD may be expected to increase average relatedness among members of Y-chromosomal subpopulations, MZD among members of mitochondrial subpopulations, and FZD for both. This intuition is illustrated more clearly in BOX 4.



BOX 4 I Given a large population with a naturalistic pedigree, randomly drawn pairs X and Y are likely to share some numbers of ancestors Z at some distribution of generational depths in mixed-sex pedigrees, illustrated in the lambda-shaped pedigree ZXY above. However, subdivisions of the main population with identical Y chromosomes (A and B) or identical mtDNA (E and F) may by uniparentally IBD, in which case they also share a common ancestor (C and G respectively) in a lambda-shaped pedigree comprised of only sons or daughters, respectively. Cousin marriages along the pedigrees above introduce a loop (see the position of D and H respectively), increasing the total number of paths along which AB, XY, or EF share ancestry (for example, if a cousin marriage occurs at the position represented by D, A and B share ancestry along two paths: ACB and ACDB).
Now notice that FBD will cause a large excess of cousin marriages to occur pedigree CAB relative to ZXY. MZD will likewise cause a large excess in GEF relative to ZXY, while FZD will cause excesses in both CAB and GEF relative to ZXY. MZD, FZD and FBD marriages increase the the number of paths and thus the total relatedness of either or both of the uniparental IBD pairs AB and EF relative to the random pair XY, as some function of their respective frequencies. Conversely, MBD preferentially loads on ZXY and cannot occur at all on the pedigrees CAB and GEF, and thus increases the relatedness of random pairs compared to uniparentally identical pairs.
Also notice that two persons with the same Y chromosome or mtDNA profile are already more likely to be related than random pairs; therefore the increase in relatedness in uniparental subpopulations occurs relative to some baseline expectation, which can be calculated using algorithms/simulations such as those in (Mikkel M. Andersen 2017)). This will appear in the math later.

Calculation of the Expected Increase in Coancestry Among Uniparental Subpopulations

I. Inbreeding Accumulation in Chainlike Pedigrees due to Cousin Marriage

Before we can formalise the intuition in the previous section, we calculate how cousin marriages increase levels of relatedness among individuals in a lambda-shaped pedigrees of the form seen in FIGURE 2. Cousin marriages may take place along one arm of a lambda-shaped pedigree, which would be a single chain; each cousin marriage increases relatedness among the person at the beginning and end of the chain by introducing loops, i.e, multiple possible paths, as seen in FIGURE 4.

We term cousin marriages that interact with lineal pedigrees to introduce loops *intersections* between a linear pedigree and a cousin marriage. Only lineal pedigrees that pass from A or A's wife in 4 to H or H's brother are intersections. Note that a lineal pedigree that passes through the mother of E or the mother

of F in FIGURE 4 to Y does not introduce loops into the pedigree, and cannot increase the relatedness between X and Y relative to a pedigree without the cousin marriage; interactions of this type are not intersections.

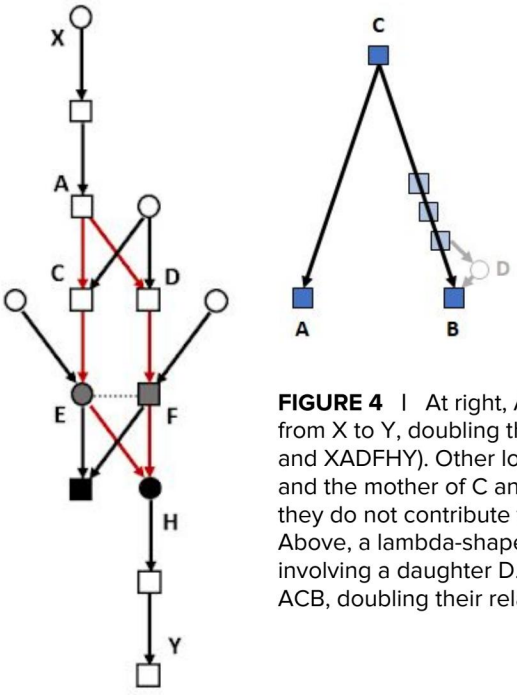


FIGURE 4 | At right, A cousin marriage introduces a loop into a chain from X to Y, doubling the number of paths leading from X to Y (XACEHY and XADFHY). Other loops in the pedigree (e.g. that formed by CDEF and the mother of C and D and the son of E and F) are not relevant, as they do not contribute to increased coancestry between X and Y. Above, a lambda-shaped pedigree CAB gains a loop via a FBD event, involving a daughter D. This leads to two paths from A to B, ACDB and ACB, doubling their relatedness levels, or their coefficient of coancestry.

In FIGURE 5, we demonstrate variation in number and degree of overlap in cousin marriages. Note that consecutive cousin marriages always overlap.

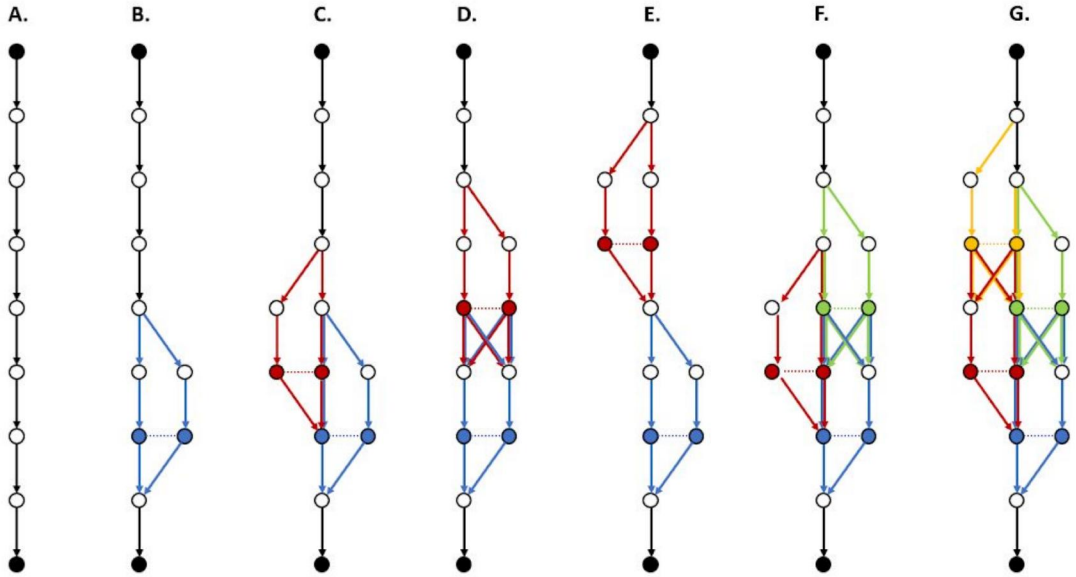


FIGURE 5 | At right, A demonstrates a chain of persons without cousin marriage. Here, all other parents of all persons are assumed not to share significant ancestry, and are thus omitted. The rest of the pedigrees demonstrate varying numbers of cousin marriages in various configurations.

The coancestry between the first and last individual of pedigree A, without inbreeding, is:

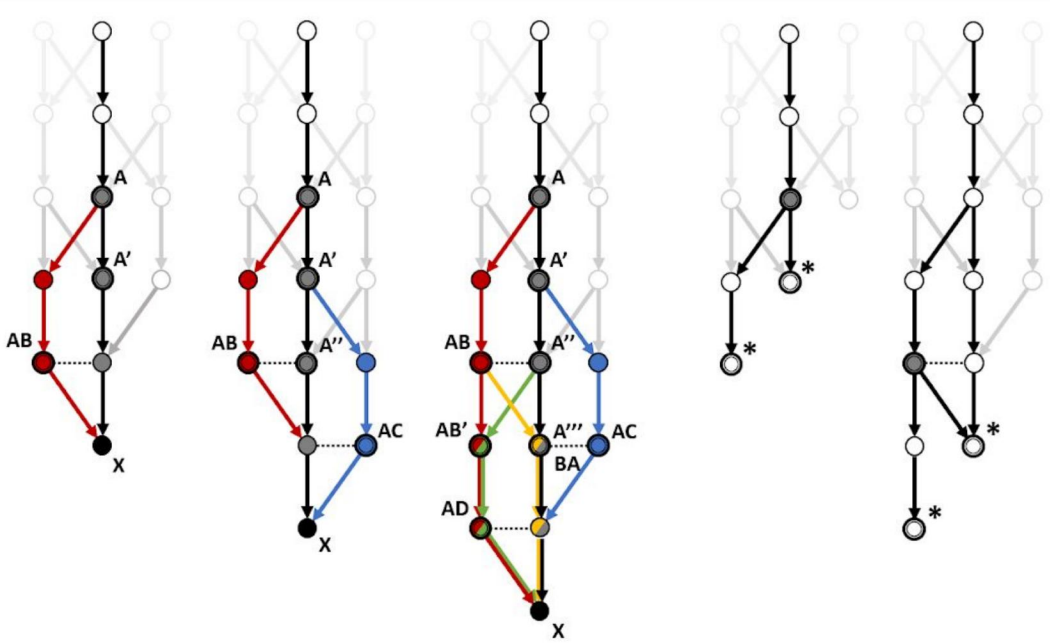
$$F_{XY} = (\frac{1}{2})^n \cdot \frac{1+F}{2} = C \tag{8}$$

Where n is the number of generations intervening, and *F* is the coefficient of inbreeding for the first node in the pedigree. Inbred pedigrees then increase the coancestry between the first and last persons in the chain (and by extension terminal nodes in a lamba-shaped pedigree) by increasing the number of paths connecting them.

For example, the coancestry for the pedigree FIGURE 5B, with a single cousin marriage, is then 2*C*, due to the existence of two chains connecting the ancestor with the descendant which only differ in the paths traced by the blue arrows. 5D and E represent *non-overlapping* cousin marriages, where each cousin marriage independently doubles the coefficient of coancestry, by doubling the number of paths; the coancestry then is (2^2)C = 4*C*.

The pedigree in FIGURE 5C represents a pair of *overlapping* cousin marriages, or *consecutive* cousin marriages, where one generation of cousin marriage (highlighted in red) is followed by another also involved (highlighted in blue). Consecutive cousin marriages do *not* independently double relatedness between the first and last persons in the chain; rather, the total number of paths given *m* consecutive cousin marriages is given by the $m+2^{nd}$ Fibonacci number. The coancestry coefficient is then 3*C* for 5C (2 cousin consecutive marriages), 5*C* for 5F (with three) and 8*C* for 5G (with four). Why this is the case is explained in BOX 5.

BOX 5 | A model of a natural process that produces the Fibonacci Sequence is that presented by Fibonacci himself: of reproducing rabbits. The important insight is the role of *generative pairs* in the process. Each generative pair rests for one turn, and produces another generative pair in all subsequent turns. The Fibonacci sequence begins with one generative pair.



Here we describe how the number of possible paths between an ancestor, Y, to a descendant X, increases as one increases the number of consecutive cousin marriages along the chain. Y exists an indeterminate number of generations ago, and is not depicted in the figure above. X is the descendant that terminates the chain. This demonstration works regardless of the number of previous cousin marriages, and the fact that prior cousin marriages may or may not exist is indicated by the faded nodes and arrows.

We introduce the idea of a *generative node*, denoted by A in the leftmost panel in the figure above. A cousin marriage induces a generative node to "reproduce", creating two more generative nodes, AB and A'. One of them is always one generation away from ego, and another is always two generations away from ego, as indicated in the two panels at right. Also note that each generative node descended from the parent node represents one possible path connecting the parent node to X. Here, the production of two generative nodes linking A to X implies the existence of two paths from Y to X.

We then add another cousin marriage, indicated in the second panel from left. A' immediately produces two more generative nodes, A'' and AC. There is no way to pass from the path through AB to the new generative node, AC, so one of the generative nodes produced in the previous cousin marriage (AB) rests, without producing any more nodes. In general, of a pair of generative nodes produced by a cousin marriage, one always rests and the other always reproduces with the addition of the next cousin marriage.

As we continue to add cousin marriages, the pattern continues. With the third cousin marriage:

1. Node A'' reproduces, producing A''' and AD.
2. Node AB reproduces, producing BA and AB'.
3. Node AC, produced in the previous generation, does not reproduce.

Note that, with the addition of more cousin marriages, single persons in the pedigree may carry multiple generative node labels, as each node may be reached from A via multiple routes--analogous to how each position in the pedigree may be produced via descent from multiple generative node sequences starting from from A.

The total number of paths from A to X given *m* cousin marriages is equal to the total number of current terminal nodes; i.e. the total number of nodes that have yet to reproduce, and thus terminate the generative node tree, after *m* generations of generative node reproduction. The pattern of production of generative nodes is identical with the pattern of reproduction of Fibonacci's rabbits, with the exception that the first cousin marriage immediately increases the number to 2, i.e. a cousin marriage number of one ($m = 1$) immediately causes the number of paths from A to X to increase to the third Fibonnaci number.

Therefore the number of possible paths (*p*) contributed by *m* cousin marriages is

$$p^m = f^{m+2}$$

Where f^m is the m^{th} Fibonacci number.

Each separate instance of sequences of cousin marriages independently generates new paths, and each instance independently multiplies the total number of paths. The expected total relatedness of persons Y and X at the beginning and end of a chainlike pedigree of large length *n*, where cousin marriages occur with frequency S and can possibly overlap in sequences of length *i*, is then

$$F_{YX} = C \cdot \prod_{i=1}^{n}((f^{i+2})^{\mathbb{E}(N_n^i)}) \tag{9}$$

Where *C* is the same expression as in equation 8, f^{i+2} is the $(i+2)^{nd}$ Fibonacci number, and $\mathbf{E}(N^i_n)$ is the expected number of sequences of consecutive cousin marriages of length *i* (terminated on either side by non-cousin marriages) occurring in a chain-like pedigree of of *n* generations. This is equal to the expected number of separate sequences of *i* consecutive successes occurring in *n* Bernoulli trials (where *p* and *q* are *S* and *1-S* in this case).

Let M^i_n refer to the number of separate sequences of consecutive successes of length *i* **or more** occurring in n Bernoulli trials. Then the following relation exists between M^i_n and N^i_n:

$$\begin{aligned} P(N_n^i = x) &= P(M_n^i - M_n^{i+1} = x) \\ &= \sum_{k=0}^{n} P(M_n^i = k + x) \cdot P(M_n^{i+1} = k) \end{aligned} \tag{10}$$

An expression for M^i_n term is given by a rather complicated expression given in (Museli 1996). Note that S is the probability that any given marriage is a cousin marriage.

$$P(M_n^i = x) = \sum_{y=x-1}^{n-ix} S^{n-y}(1-S)^y \sum_{m=x}^{min(y+1,\frac{n-y}{i})} (-1)^{m-x}\binom{m}{x}\binom{y+1}{m}\binom{n-mi}{y} \tag{11}$$

Note that equations 10 and 11 can be substituted into equation 9 (with the usual formula to derive expectation from probability, $\mathbb{E}(x) = \sum xP(x)$ for a numeric solution given some *S*. However numeric calculation is less the point here: the properties of 10 and 11 will be used to calculate a ratio between expected relatedness of uniparentally identical and nonidentical persons under different values of *S* later on.

The value for X and Y when they are on either side of a lambda-shaped pedigree is very much the same as that given in equation 9, where F is replaced by F_Z, the inbreeding coefficient of the common ancestor of Y and X, and the exponent of ½ is doubled to account for double the number of genealogical degrees:

$$F_{XY} = (\frac{1}{2})^{2n} \cdot \frac{1 + F_Z}{2} = C \tag{8b}$$

The increase in relatedness caused by cousin marriage is the same and can likewise be calculated through the substitution of equations 10 and 11 into 8b.

II. Excess Inbreeding Accumulation in Chainlike Pedigrees of a Single Sex due to Cousin Marriage

Consider a population with a massive, complete, naturalistic pedigree, from which we produce a sample, Q. Within this sample, no persons have signals of recent close ancestry.

Two persons are drawn randomly from Q, X and Y, who do not have identical Y chromosomes or mtDNA sequences. They may share some set of possible common ancestors Z^i, $0 < i < k$, in a numbered set $\{Z\}$ at some generational depth n_Z^i, with probability $P\{Z^i\}$ along a pedigree of any shape, depicted in Figure 4. Because of their discordant Y chromosomes, no pedigree between X and Y can cause them to have uniparental IBD; i.e, no pedigrees to any common ancestors Z can be single-sex male pedigrees leading back to a male ancestor, or single-sex female pedigrees leading back to a female ancestor.

Also consider another random pair A and B, randomly drawn from Q, who do happen to have the same Y chromosome, and thus belong to a single Y-chromosomal subpopulation (Figure 4). Other than the ancestors Z^i that A and B share by chance in pedigrees of the same type as XYZ in figure 4, A and B have a nonzero probability $P(Y_{IBD})$ of having a common male ancestor, C, at some generational depth, n_y, who passed on his Y-chromosome to A and B along a single-sex male pedigree.

Likewise consider a third random pair E and F drawn from Q, who have the same mtDNA sequence, and thus belong to the same mitochondrial subpopulation. Other than the ancestors Z^i that E and F share by chance in pedigrees of the same type as XYZ in figure 4, E and F have a nonzero probability $P(mtDNA_{IBD})$ of having a common male ancestor, G, at some generational depth, n_{mtDNA}, who passed on her mitochondria to E and F along a single-sex female pedigree.

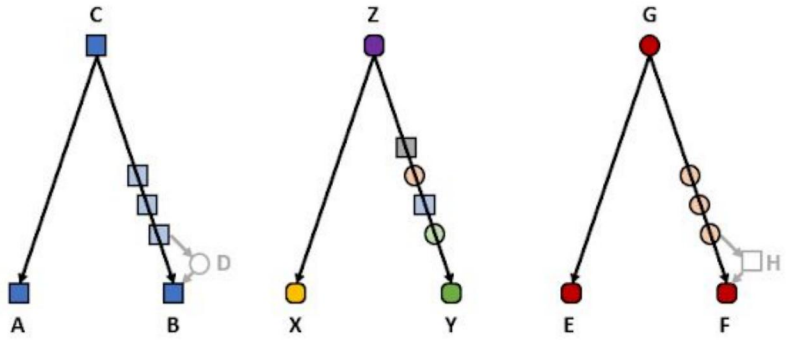


FIGURE 5 | Two randomly drawn members of P, X and Y, share some ancestors Z^k at generational depth n_Z^k with probability P_Z^k. Because there is no requirement that X and Y are uniparentally identical-by-descent in either the mtDNA or the Y chromosome (represented by the different colors), they may have a pedigree leading back to Z, who may themselves be of either sex, involving any proportion of persons of both sexes in any order. On the other hand, A and B, drawn from a uniparental subpopulation, have a nonzero probability P_Y of sharing a male ancestor C at generational depth n_Y through a single-sex pedigree of only males.

Without inbreeding, the coancestry between X and Y is simply the summation of coancestry along all shared ancestors in set $\{Z\}$, using Malecot's definition of coancestry:

$$F_{XY} = \sum_{i=1}^{k} P(Z^i) \cdot (\frac{1}{2})^{2n_Z^i} \cdot \frac{1 + F_{Z^i}}{2} \tag{12.a}$$

In other words, it is a summation of relatedness among all lambda-shaped pedigrees of the form ZXY. Note that the term:

$$(\frac{1}{2})^{2n_Z^i} \cdot \frac{1 + F_{Z^i}}{2}$$

Is identical to the expression for *C* in equation 8.

Let us indicate the value of coancestry from an ancestor Z^i at n generations ago with no cousin marriages by the following:

$$C_i^n = (\frac{1}{2})^{2n_Z^i} \cdot \frac{1 + F_{Z^i}}{2} \tag{13a}$$

Then equation 12 can be restated as:

$$F_{XY} = \sum_{i=1}^{k} P(Z^i) \cdot C_i^n \tag{12.a.i}$$

The coancestry between A and B, then includes all ancestry shared due to ancestors in Z^i, plus the relatedness resulting from the single-sex pedigree leading back to ancestor C:

$$F_{AB} = \sum_{i=1}^{k} P(Z^i) \cdot C_i^n + P(Y_{IBD}) \cdot (\frac{1}{2})^{2n_Y} \cdot \frac{1 + F_C}{2} \tag{12.b}$$

Likewise, the coancestry between E and F also includes the relatedness resulting from the single-sex pedigree leading back to G:

$$F_{EF} = \sum_{i=1}^{k} P(Z^i) \cdot C_i^n + P(mtDNA_{IBD}) \cdot (\frac{1}{2})^{2n_{mtDNA}} \cdot \frac{1 + F_G}{2} \tag{12.c}$$

Now, analogously to equation 13 and its expression for C, let us define the following terms:

$$C_Y^{n_Y} = (\frac{1}{2})^{2n_Y} \cdot \frac{1 + F_C}{2} \tag{13.b}$$

$$C_{mtDNA}^{n_{mtDNA}} = (\frac{1}{2})^{2n_{mtDNA}} \cdot \frac{1 + F_G}{2} \tag{13.c}$$

It follows that equations 12b and 12c can be reexpressed:

$$F_{AB} = \sum_{i=1}^{k} P(Z^i) \cdot C_i^n + P(Y_{IBD}) \cdot C_Y^{n_Y} \tag{12.b.i}$$

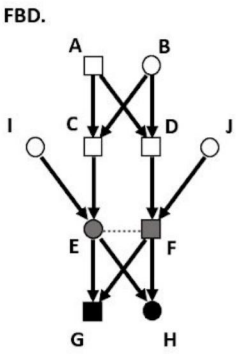
$$F_{EF} = \sum_{i=1}^{k} P(Z^i) \cdot C_i^n + P(mtDNA_{IBD}) \cdot C_{mtDNA}^{n_{mtDNA}} \tag{12.c.i}$$

The probability that two persons with a Y-chromosomal match or mtDNA match are related through shared ancestry more than the average is significant in most real-world situations. Therefore, the extra terms associated with the EF and AB pairs, compared to the XY pair (compare equations 12, 12.b.i and 12.b.ii) indicate that, under conditions of no cousin marriage, persons with uniparental IBD are already more likely to be related than random pairs. One may estimate the excess relatedness of persons with uniparental IBD using methods such as those detailed in Andersen (2017). The problem is analogous to one in forensics: how likely are two persons with a Y-chromosome or uniparental match actually related, and if so, how deeply? This depends on a variety of demographic parameters and the quality of the match, which may be estimated separately from aspects of the autosomal data.

More importantly, with no cousin marriage plus random mating, one may expect F_{EF} and F_{AB} to be relatively equal. The same stochastic population processes generate uniparental matches in mtDNA and the Y-chromosome, and there is therefore no reason to expect persons with matrilineal uniparental matches to be more related along the autosome than those with patrilineal ones.

Now, consider how each type of cousin marriage affects the calculations of each pair in the lambda-shaped pedigree. Notice that, because the chains CA and CB are all male, and GE and GF are all female, there would be a large excess of FBD and FZD marriages intersecting with the lineal arms of the pedigree ABC relative to XYZ, and a large excess of MZD and FZD marriages intersecting with the pedigree GEF relative to XYZ. In fact, if the frequency of either FBD, FZD or MZD marriages is *S* for XYZ, it is *2S* in the single-sex pedigrees. At the same time, if the frequency of MBD marriages in random pedigrees is *S*, it is 0 in unisex pedigrees. To see why, see BOX 5.

BOX 5 | Suppose we have a naturalistic pedigree of infinite size of some population *P*. Within this pedigree, cousin marriages of the FBD type occur at proportion *S* of all marriages. Observe a pedigree of one FBD cousin marriage, on the left.



Considering sequences of four parent-child relationships, there are 8 different sequences of length four from the common ancestors (A, B) to the descendants (G, H) distinguished by the order of sexes, which we list here in the set *L*:

ADFG (MMMM), ACEG (MMFM), ADFH (MMMF), ACEH (MMFF),
BDFH (FMMF), BCEG (FMFM), BDFG (FMMM), BCEH (FMFF)

Note that in all such sequences, the third element (E or F) is the one involved in the cousin marriage. Also note that every FBD marriage event, in a sense, "produces" a single instance of all 8 different sequences at once, at a single location within the pedigree of *P*.

Now we reiterate the definition of an *intersection*: When a chain of parent-child relationships in the pedigree contains an FBD cousin marriage event by

intersecting the FBD configuration depicted at right through one of the sequences listed above (in other words, the chain contains one of the members of the pairs (A, B), and (G, H), linked through one of the sequences listed above), we will call this an *intersection* of a chain with a FBD marriage event. An intersection in the chain XY is depicted below, where the chain passes through ACEH (MMFF) and ADFH (MMMF).

There are altogether 16 (2^4) different ways to arrange sequences of four with respect to sex, of which only 8 listed above in *L* can possibly be placed where a chain intersects a single FBD marriage. It is not possible for the other sequences exist where a FBD marriage intersects with a chain in the manner defined here.

Now let us define the set of {all sequences of four parent-child links in the pedigree of *P*}, or *N*. Notice that all individuals who have reproduced at least once in the pedigree of *P* can serve as third elements within sequences of four parent-child links, as they by definition have reproduced, and needs must have parents and grandparents. This implies that the set {all sequences of four parent-child links}, or *N*, effectively represents all persons who reproduce. Thus *N* is also, effectively, a set of all marriage events.

Now let us select a chain of four consecutive parent-child links from the pedigree, i.e. an element of *N*, at random, and call this element *M*. If the configuration of sexes in this sequence is such that it falls in set *L*, then it may intersect a FBD marriage; let us call the probability that this sequence is involved in a FBD marriage S^L. If the configuration of sexes in this sequence is such that it does not fall in set *L*, let us call the probability that this sequence is involved in a FBD marriage $S^{!L}$. Because all marriages are represented in *N*, it follows that the probability that any element of *N* intersects with FBD marriage obeys the following identity:

$$P(M \text{ intersects } FBD) = S^L \cdot P(M \in L) + S^{!L} \cdot P(M \notin L)$$

We have shown above that the set N is effectively the set of all marriages, so P(M is FBD) is equal to *S*.

$$P(M \text{ intersects } FBD) = S = S^L \cdot P(M \in L) + S^{!L} \cdot P(M \notin L)$$

Now we know that L contains 8 sequences of four sexes, while the total possible is 16, implying that the probability of any random sequence of four from the pedigree having a sequence of sexes in *L* is ½. So $P(M \in L) = 1/2$. In addition, we know that $S^{!L}$ is 0, because a four-member sequence whose sequence of sexes is not in L cannot intersect with a FBD marriage.

It follows that:

$$S = \frac{1}{2}S^L + 0$$

So the probability that a given four-member sequence with a sequence of sexes in *L* intersects with a FBD marriage is 2*S* when the overall probability of FBD marriage is *S*. Without loss of generality, notice once can repeat the same demonstration for FZD, MZD and MBD marriages.

Now notice that unisex pedigrees only provide elements *M* from the set *N* of the form MMMM or FFFF. By the reasoning above, the probability that any given MMMM and FFFF sequence intersects with a cousin marriage, if the overall probability is *S*, would be the folllowing:

	FBD	FZD	MZD	MBD
FFFF	0	2S	2S	0

MMMM	2S	2S	0	0

It follows that if the frequency of either FBD, FZD or GEF marriages is S for random lambda-shaped pedigrees, it is 2S in the single-sex lambda-shaped pedigrees, and if the probability of MBD is S for the random lambda-shaped pedigrees, it is 0 for single-sex pedigrees.

If MBD is taking place at frequency S out of all marriages, the system of equations become:

$$F_{XY} = \sum_{i=1}^{k} P(Z^i) \cdot C_i^n \cdot \prod_{i=1}^{n} (f^{i+2})^{\mathbb{E}(N_n^i|p=S)} \tag{14.a.i}$$

$$F_{AB} = \sum_{i=1}^{k} P(Z^i) \cdot C_i^n \cdot \prod_{i=1}^{n} (f^{i+2})^{\mathbb{E}(N_n^i|p=S)} + P(Y_{IBD}) \cdot C_Y^{n_Y} \cdot \prod_{i=1}^{n} (f^{i+2})^{\mathbb{E}(N_n^i|p=0)}$$

$$= \sum_{i=1}^{k} P(Z^i) \cdot C_i^n \cdot \prod_{i=1}^{n} (f^{i+2})^{\mathbb{E}(N_n^i|p=S)} + P(Y_{IBD}) \cdot C_Y^{n_Y} \tag{14.b.i}$$

Because the probability of MBD marriages occurring in an all-male pedigree is 0, MBD is expected to increase the average relatedness of persons without and with uniparental IBD along the Y-chromosome equally. The same applies for mtDNA:

$$F_{EF} = \sum_{i=1}^{k} P(Z^i) \cdot C_i^n \cdot \prod_{i=1}^{n} (f^{i+2})^{\mathbb{E}(N_n^i|p=S)} + P(mtDNA_{IBD}) \cdot C_{mtDNA}^{n_{mtDNA}} \tag{14.c.i}$$

On the other hand, for all other types of cousin marriage, the relatedness between uniparentally identical individuals behaves differently. For FBD, the equations are much the same as those in set 14, except for the second:

$$F_{AB} = \sum_{i=1}^{k} P(Z^i) \cdot C_i^n \cdot \prod_{i=1}^{n} (f^{i+2})^{\mathbb{E}(N_n^i|p=S)} + P(Y_{IBD}) \cdot C_Y^{n_Y} \cdot \prod_{i=1}^{n} (f^{i+2})^{\mathbb{E}(N_n^i|p=2S)} \tag{15}$$

Here, FBD increases the relatedness between individuals with uniparental IBD along the Y-chromosome disproportionately (notice that p=2S). The absolute increase in relatedness among the Y-chromosome subpopulations is, after substituting equation 11 and some algebra:

For MZD, the equations are likewise, except for the third:

$$F_{EF} = \sum_{i=1}^{k} P(Z^i) \cdot C_i^n \cdot \prod_{i=1}^{n} (f^{i+2})^{\mathbb{E}(N_n^i|p=S)} + P(mtDNA_{IBD}) \cdot C_{mtDNA}^{n_{mtDNA}} \cdot \prod_{i=1}^{n} (f^{i+2})^{\mathbb{E}(N_n^i|p=2S)} \tag{16}$$

For FZD, both the second and third equations are modified.

Given the following:

$$\prod_{i=1}^{n}(f^{i+2})^{\mathbb{E}(N_n^i|p=S)} = J_i^n$$
$$\prod_{i=1}^{n}(f^{i+2})^{\mathbb{E}(N_n^i|p=2S)} = K_i^n$$

Then, with some algebraic manipulation after substitution of equations 10 and 11 into 9 for the value of $\mathbf{E}(N^i_n)$:

$$\frac{K_i^n}{J_i^n} = \prod_{i=1}^{n}(f^{i+2})^{\sum_{x=1}^{n} x \sum_{k=0}^{n}(\sum_{y=x-1}^{n-ik} \frac{1-S^y}{2(1-2S)^y})^2}$$

$$K_i^n >> J_i^n$$

Table 1 describes the pattern of relatedness resulting from the occurrence of each type of cousin marriage, relative to the relatedness of a random pair in a population without inbreeding:

	Random Pair	Pair from Y-chromosome Uniparental subpopulation	Pair from Mitochondrial Uniparental subpopulation
No Cousin Marriage	1	$1 + P(Y_{IBD}) \cdot e$	$1 + P(mtDNA_{IBD}) \cdot e$
FBD	J^n_i	$J^n_i + P(Y_{IBD}) \cdot e \cdot K^n_i$	$J^n_i + P(mtDNA_{IBD}) \cdot e$
FZD	J^n_i	$J^n_i + P(Y_{IBD}) \cdot e \cdot K^n_i$	$J^n_i + P(mtDNA_{IBD}) \cdot e \cdot K^n_i$
MZD	J^n_i	$J^n_i + P(Y_{IBD}) \cdot e$	$J^n_i + P(mtDNA_{IBD}) \cdot e \cdot K^n_i$
MBD	J^n_i	$J^n_i + P(Y_{IBD}) \cdot e$	$J^n_i + P(mtDNA_{IBD}) \cdot e$

TABLE 1 | Amount of inbreeding for each type of pair, relative to a random pair from a population with no cousin marriage.

Notice that the signal of excess inbreeding produced within single-sex lineages is greater when $P(Y_{IBD})$ and $P(mtDNA_{IBD})$ are themselves significant. This is a limitation of this method. Note also that there is no clear way to derive S values from the relatedness of random pairs and uniparentally identical pairs. This method may be more suited to simulations, and may work well with Approximate Bayesian Computation (ABC).

BIBLIOGRAPHY

Ackerman, Matthew S., Parul Johri, Ken Spitze, Sen Xu, Thomas G. Doak, Kimberly Young, and Michael Lynch. 2017. "Estimating Seven Coefficients of Pairwise Relatedness Using Population-Genomic Data." *Genetics* 206 (1): 105–18.

Bittles, A. H. 2007. "Consanguineous Marriage and Childhood Health." *Developmental Medicine & Child Neurology* 45 (8): 571–76.

Bittles, A. H., J. M. Coble, and N. Appaji Rao. 1993. "Trends in Consanguineous Marriage in Karnataka, South India, 1980–89." *Journal of Biosocial Science* 25 (01). https://doi.org/10.1017/s0021932000020356.

Bittles, A. H., A. Radha Rama Devi, H. S. Savithri, Rajeswari Sridhar, and N. Appaji Rao. 1987. "Consanguineous Marriage and Postnatal Mortality in Karnataka, South India." *Man* 22 (4): 736.

Bittles, Alan H. n.d. "Demographic and Socioeconomic Aspects of Consanguineous Marriage." In *Consanguinity in Context*, 58–73.

Broushaki, Farnaz, Mark G. Thomas, Vivian Link, Saioa López, Lucy van Dorp, Karola Kirsanow, Zuzana Hofmanová, et al. 2016. "Early Neolithic Genomes from the Eastern Fertile Crescent." *Science* 353 (6298): 499–503.

Donnan, Hastings, and Ladislav Holy. 1990. "Kinship, Honour and Solidarity: Cousin Marriage in the Middle East." *Man* 25 (4): 732.

Edlund, Lena. 2018. "Cousin Marriage Is Not Choice: Muslim Marriage and Underdevelopment." *AEA Papers and Proceedings* 108: 353–57.

Enke, Benjamin. 2017. "Kinship Systems, Cooperation and the Evolution of Culture." https://doi.org/10.3386/w23499.

Freire-Maia, N., and A. Freire-Maia. 1961. "The Structure of Consanguineous Marriages and Its Genetic Implications." *Annals of Human Genetics* 25 (1): 29–39.

Fuster, V., and S. E. Colantonio. 2004. "Socioeconomic, Demographic, and Geographic Variables Affecting the Diverse Degrees of Consanguineous Marriages in Spain." *Human Biology* 76 (1): 1–14.

Gallego-Llorente, M., S. Connell, E. R. Jones, D. C. Merrett, Y. Jeon, A. Eriksson, V. Siska, et al. 2016. "The Genetics of an Early Neolithic Pastoralist from the Zagros, Iran." *Scientific Reports* 6 (August): 31326.

Grossman, M., and E. J. Eisen. 1989. "Inbreeding, Coancestry, and Covariance between Relatives for X-Chromosomal Loci." *The Journal of Heredity* 80 (2): 137–42.

Hedrick, Philip W., and C. Clark Cockerham. 1986. "PARTIAL INBREEDING: EQUILIBRIUM HETEROZYGOSITY AND THE HETEROZYGOSITY PARADOX." *Evolution; International Journal of Organic Evolution* 40 (4): 856–61.

Hedrick, P. W. 1986. "Average Inbreeding or Equilibrium Inbreeding?" *American Journal of Human Genetics* 38 (6): 965–70.

Kang, Gay Elizabeth. 1976. "Conflicting Loyalties Theory: A Cross-Cultural Test." *Ethnology* 15 (2): 201.

Korotayev, Andrey. 2000. "Parallel-Cousin (FBD) Marriage, Islamization, and Arabization." *Ethnology* 39 (4): 395.

Korotayev, Andrey V. 2003. "Unilineal Descent Organization and Deep Christianization: A Cross-Cultural Comparison." *Cross-Cultural Research: Official Journal of the Society for Cross-Cultural Research / Sponsored by the Human Relations Area Files, Inc* 37 (1): 133–57.

Kousathanas, Athanasios, Christoph Leuenberger, Vivian Link, Christian Sell, Joachim Burger, and Daniel Wegmann. 2017. "Inferring Heterozygosity from Ancient and Low Coverage Genomes." *Genetics* 205 (1): 317–32.

Kudo, A., and K. Sakaguchi. 1963. "A METHOD FOR CALCULATING THE INBREEDING COEFFICIENT. II. SEX-LINKED GENES." *American Journal of Human Genetics* 15 (December): 476–80.

MacLeod, Iona M., Denis M. Larkin, Harris A. Lewin, Ben J. Hayes, and Mike E. Goddard. 2013. "Inferring Demography from Runs of Homozygosity in Whole-Genome Sequence, with Correction for Sequence Errors." *Molecular Biology and Evolution* 30 (9): 2209–23.

Malécot, Gustave. 1970. *The Mathematics of Heredity*. W.H. Freeman.

Manichaikul, Ani, Josyf C. Mychaleckyj, Stephen S. Rich, Kathy Daly, Michèle Sale, and Wei-Min Chen. 2010.

"Robust Relationship Inference in Genome-Wide Association Studies." *Bioinformatics* 26 (22): 2867–73.
Mikkel M. Andersen, David J. Balding. 2017. "How Convincing Is a Matching Y-Chromosome Profile?" *PLoS Genetics* 13 (11). https://doi.org/10.1371/journal.pgen.1007028.
Murdock, George Peter. 2015. *Social Structure*. Sagwan Press.
Museli, Marco. 1996. "Simple Expressions for Success Run Distributions in Bernoulli Trials." *Statistics & Probability Letters* 31 (2): 121–28.
Otterbein, Keith F. 1968. "Internal War: A Cross-Cultural Study." *American Anthropologist* 70 (2): 277–89.
Richards, F. J. 1914. "97. Cross Cousin Marriage in South India." *Man* 14: 194.
Schulz, Jonathan F. 2016. "The Churches' Bans on Consanguineous Marriages, Kin-Networks and Democracy." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2877828.
Shaw, Alison. 2001. "Kinship, Cultural Preference and Immigration: Consanguineous Marriage Among British Pakistanis." *The Journal of the Royal Anthropological Institute* 7 (2): 315–34.
Thornhill, Nancy Wilmsen. 1990. "The Evolutionary Significance of Incest Rules." *Ethology and Sociobiology* 11 (2): 113–29.
Walker, Robert S., and Drew H. Bailey. 2014. "Marrying Kin in Small-Scale Societies." *American Journal of Human Biology: The Official Journal of the Human Biology Council* 26 (3): 384–88.
Wang, J. 2017. "Estimating Pairwise Relatedness in a Small Sample of Individuals." *Heredity* 119 (5): 302–13.
Wright, Sewall. 1922. "Coefficients of Inbreeding and Relationship." *The American Naturalist* 56 (645): 330–38.