

1 The effect of inbreeding constraints and offspring
2 distribution on time to the most recent common
3 ancestor

4 R. B. Campbell

5 *Department of Mathematics, University of Northern Iowa, Cedar Falls, IA 50614-0506*

6 *<http://www.math.uni.edu/~campbell>*

7 *campbell@math.uni.edu*

8 *(319)273-2447*

9 *fax:(319)273-2546*

10 **Abstract**

11 The expected time to the most recent common ancestor (MRCA) of two al-
12 leles in a diploid individual is $4N + 2$ under random mating with a Poisson
13 progeny distribution, but $8N - 2$ under maximum avoidance of inbreeding,
14 which entails two progeny per mating pair. (N is the number of mating pairs,
15 hence $2N$ is the number of individuals, hence $4N$ is the number of alleles.)
16 The interrelationship of inbreeding constraints and offspring distribution is
17 investigated by varying the level of sib mating: Prohibiting sib mating in-
18 creases the time to MRCA by four generations and decreases the variance
19 of the offspring distribution by $2/N$. With two progeny per mating pair,
20 the expected time to the MRCA is $8N - 2$ under both random mating and
21 sib mating prohibited, as well as under maximum avoidance of inbreeding,
22 but this result does not hold for all mating structures with two progeny per
23 mating pair.

24 *Keywords:* coalescent, genetic identity, pedigrees

25 *2008 MSC:* 92D15, 92D25

26 **1. Introduction**

27 All alleles are ultimately identical by descent, the time to the most recent
28 common ancestor (MRCA) governs the likelihood that they are identical by

29 state (a mutation has not occurred since their common ancestor). The time
30 to a common ancestor is impacted by the progeny distribution and the mating
31 structure of the population, with increased variance of the progeny distribu-
32 tion and consanguineous mating reducing the time to a common ancestor.
33 Consanguineous marriages were common in the millennia preceding the com-
34 mon era including extensive sib mating among the pharaohs. More recently,
35 there has been cousin mating among the royal houses of Europe. There has
36 also been prohibition of consanguineous marriages, even third cousin mating
37 required dispensation in the Catholic church until 1917 (Cavalli-Sforza and
38 Bodmer, 1971). This dichotomy remains today with some cultures banning
39 consanguineous marriages while other cultures favor them (Cavalli-Sforza
40 and Bodmer, 1971; Krawczak and Barnes, 2010).

41 Pedigree information from parish records has allowed calculation of the
42 frequency of consanguineous matings (Bittles and Egerbladh, 2005; Pettay
43 et al., 2007; Wakeley et al., 2012). Some of these include information up to
44 sixth degree cousins. A question which will be addressed is whether this is
45 sufficient information to determine genetic identity.

46 The mating structure of a population is only important within the con-
47 text of the population size. If the population size is infinite, the expected
48 time since a common ancestor will be infinite as long as the pedigree grows
49 (Lachance, 2009), because an ancestor which escapes the pedigree structure
50 will never return. If the mating structure provides a de facto finite ancestral
51 population size (e.g., full sib mating or double first cousin mating), then the
52 actual population size is not relevant and the expected time to a common
53 ancestor is finite. We want more generality than regular systems of inbreed-
54 ing (Arzberger, 1988; Kimura and Crow, 1963; Wright, 1921) because they
55 generally require too much structure, often entailing a finite ancestral pop-
56 ulation size, and often requiring exactly two progeny per individual, which
57 confounds the effect of the progeny distribution with the effect of the pedi-
58 gree structure (Campbell, 1993). However, we shall employ regular systems
59 of inbreeding to bound genetic behavior.

60 **2. The Model**

61 This analysis is based on the observation by Kingman (1982a) that the
62 Poisson progeny distribution is equivalent to each individual “chooses its
63 parents at random” which introduced the coalescent into population genetics
64 (Hein et al., 2005; Kingman, 1982a, 1982b; Wakeley, 2008). In particular, we

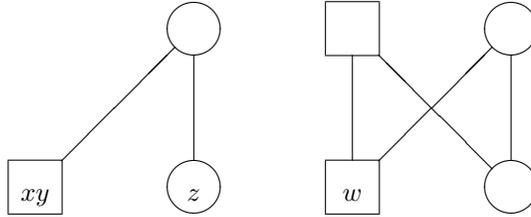


Figure 1: Location of alleles. Alleles x and y are in the same individual, alleles x and z are in a mating pair, but not in the same individual, and alleles x and w are in two different mating pairs. Squares represent males, circles represent females, and lines extend upward to progeny.

65 study the trajectory of two alleles in an individual backward in time to find
 66 the coalescent time for those two alleles, which is the time to their MRCA. We
 67 employ a dioecious diploid model with N monogamous mating pairs, hence
 68 $2N$ individuals, hence $4N$ alleles (at a single locus). Because this model is
 69 diploid with monogamous mating pairs, an individual chooses a mating pair
 70 for its parents rather than choosing a single individual as occurs in haploid
 71 models or two individuals as occurs under unconstrained random mating.

72 Specifically, the model is that every generation has N monogamous mat-
 73 ing pairs. The $2N$ individuals in those pairs are randomly assigned to monog-
 74 amous mating pairs (their parents) the previous generation. If an assignment
 75 violates an inbreeding constraint, that individual will be randomly reassigned
 76 until no constraint is violated; inbreeding constraints may specify either that
 77 inbreeding occurs or is avoided. The inbreeding constraints will alter the
 78 progeny distribution so that it is not Poisson. The pedigree goes backward
 79 in time, but we shall employ generation 1 as the generation before generation
 80 0, so that increasing the generation number goes back further in time. All
 81 calculations are in real time without rescaling.

82 3. Full Sib Mating Specified

83 The first model specifies the amount of full sib mating with no other
 84 constraints. Random mating is a special case of this for which the time to
 85 the MRCA is $4N + 2$.

86 We employ s for the fraction of matings which are between (full) sibs
 87 (since we are assuming monogamous mating pairs, there are no half-sibs).
 88 The model entails N mating pairs each generation, sN of them are randomly

89 chosen to be randomly assigned with replacement a single mating pair as par-
 90 ents, while the rest are assigned with replacement two distinct mating pairs
 91 as parents (i.e., one mating pair the previous generation for each individual
 92 in the mating pair). Going forward in time, this entails sN mating pairs
 93 chosen (with replacement) to parent mating pairs (i.e., full sib matings the
 94 next generation), and then mating pairs are chosen (with replacement) to
 95 parent the remaining individuals subject to the constraint that no mating
 96 pair is the parent of both individuals in a mating pair. An alternative model
 97 is presented in the appendix.

98 The mating pairs assigned two sets of parents must be assigned two dif-
 99 ferent sets of parents, but they may be assigned mating pairs chosen by indi-
 100 viduals in other mating pairs, including individuals in full sib mating pairs.
 101 This process is iterated through the ancestral generations. This process can
 102 be analyzed using three quantities: T , the expected time to a common an-
 103 cestor of two alleles in an individual; U , the expected time to a common
 104 ancestor of two alleles in two individuals within a mating pair; and V , the
 105 expected time to a common ancestor of two alleles in two different mating
 106 pairs (see Figure 1).

107 This provides a system of three equations:

$$T = U + 1 \tag{1}$$

108 because if two alleles are in an individual, one was in each parent the previous
 109 generation - two alleles in an individual in generation 0 were two alleles in
 110 two individuals within a mating pair in generation 1.

$$U = s(.25 + .25(1 + T) + .5(1 + U)) + (1 - s)(1 + V) \tag{2}$$

111 If that mating pair in generation 0 were sibs (shared parents in generation
 112 1), the alleles would be copies of the same allele .25 of the time, copies
 113 of two alleles in an individual .25 of the time, and copies of alleles in two
 114 individuals in a mating pair .5 of the time. If that mating pair were not
 115 sibs, their ancestral alleles would have been in two different mating pairs in
 116 generation 1.

$$V = (1/N)(.25 + .25(1 + T) + .5(1 + U)) + (1 - 1/N)(1 + V) \tag{3}$$

117 If the alleles were in two different mating pairs in generation 0, $1/N$ of the
 118 time they would have been in the same mating pair in generation 1 by random

119 choice of parents, and $(1-1/N)$ of the time they would have been in different
 120 mating pairs in generation 1. (Even if one or both individuals in generation
 121 0 were the result of sib mating, they could have chosen the same parents.)

122 Equations (1-3) have the solution $T = 6 + (1-s)4N$. If $s = 1$ (mandatory
 123 sib mating), T is equal to 6; this can be easily shown directly since mandatory
 124 full sib mating provides each individual an ancestral population of size 2. If
 125 $s = 0$ (sib mating precluded), $T = 4N + 6$. The expected time to a common
 126 ancestor decreases linearly as s increases. Random mating is provided by
 127 $s = 1/N$, which yields $T = 4N + 2$; this can be derived directly by noting
 128 that equations (2) and (3) for U and V above are the same if $s = 1/N$, so
 129 combining them with (1) provides the single equation $T = 1 + 1/(4N) + (1 +$
 130 $T)/(4N) + T(1 - 1/(2N))$.

131 The assumption of constant population size requires that the expected
 132 number of progeny per mating pair is two, but inbreeding constraints will
 133 cause the variance of the progeny distribution to change. With random
 134 mating (individuals randomly choose parents), each of the N mating pairs
 135 will have probability $1/N$ of parenting each of the $2N$ progeny, and the
 136 binomial distribution provides that the variance of the number of progeny
 137 per mating pair is $2N(1/N)(1 - 1/N)$. If sib mating is precluded, each
 138 mating pair will have probability $2/N$ of having exactly one progeny in each
 139 of the N mating pairs the next generation, and by the binomial distribution
 140 the variance of the number of progeny will be $N(2/N)(1 - 2/N)$. With
 141 mandatory sib mating, each of the N mating pairs will have probability $1/N$
 142 of parenting each of the mating pairs (i.e., both progeny in the mating pair)
 143 the following generation, and by the binomial distribution the the variance
 144 of the number of progeny will be $4 \times N(1/N)(1 - 1/N)$. The initial 4 is
 145 present because each parental mating pair has 2 or 0 progeny in a mating
 146 pair the next generation rather than 1 or 0. Random mating entails that
 147 the fraction of sib-mating is $s = 1/N$, and $2N(1/N)(1 - 1/N) = (1/N) \times$
 148 $4N(1/N)(1 - 1/N) + (1 - 1/N) \times N(2/N)(1 - 2/N)$. The variance of the
 149 progeny distribution is a linear function of the amount of sib mating.

This variance may be approximated as

$$s \times 4 + (1 - s) \times 2 = 2(1 + s) \tag{4}$$

150 which is obtained if one uses the Poisson approximation to the binomial
 151 distribution. Thus as the proportion of sib mating pairs (s) increases from 0
 152 to 1, the variance of the progeny distribution increases from (approximately)
 153 2 to 4, and the time since the MRCA decreases from $4N + 6$ to 6.

154 4. Multiple Generations of Inbreeding Precluded

155 We have shown that changing from random mating to prohibiting full
156 sib mating increases the time since a common genetic ancestor from $4N + 2$
157 to $4N + 6$ generations, i.e., by 4 generations. An alternative derivation for
158 the juxtaposition of random mating with sib mating precluded provides the
159 basis for an heuristic argument that the time since a common ancestor is
160 extended by approximately 4 generations for each generation that the number
161 of generations without inbreeding is extended.

162 Whether or not sib mating is precluded, two alleles in an individual are
163 two alleles in different individuals in a mating pair the previous generation.
164 Under random mating, two alleles in different individuals in a mating pair
165 have probability $1/N$ of being in the same mating pair (perhaps the same
166 allele, perhaps different alleles in an individual, perhaps in different members
167 of the mating pair) the previous generation. If they are not in the same
168 mating pair, they will be in different mating pairs. Each generation they are
169 in different mating pairs, they will have probability $1/N$ of being in the same
170 mating pair the previous generation.

171 If sib mating is precluded, two alleles in different individuals in a mating
172 pair cannot be in the same mating pair the previous generation, and will
173 need to be in different mating pairs the previous generation. From this state
174 they will have probability $1/N$ of being two alleles in the same individual the
175 previous generation each generation until they are two alleles in a mating pair.
176 Thus prohibiting sib mating extends by one generation the time until two
177 alleles in an individual are first two alleles in a mating pair, that additional
178 generation being generation 2 when two alleles in different individuals in a
179 mating pair must have come from different mating pairs.

180 When ancestors of two alleles in an individual are first in the same mating
181 pair, one fourth of the time they will be the same allele, and common ancestry
182 will be established. One fourth of the time they will be two alleles in an
183 individual, from which state prohibiting sib mating increases the expected
184 time until they are in the same mating pair by one generation as was just
185 shown. One half the time they will be in different individuals in a mating
186 pair, from which state it was also shown prohibiting sib mating extends
187 the expected time until they are in the same mating pair by one generation.
188 Hence after the initial occurrence as two alleles in an individual, whenever the
189 ancestors of those alleles are in the same mating pair, one fourth of the time
190 common ancestry occurs and three quarters of the time the expected time to

191 common ancestry is increased by one generation under sib mating precluded
192 versus random mating. This provides an overall increase of $\sum_{i=0}^{\infty} (3/4)^i = 4$
193 generations in the expected time until common ancestry under sib mating
194 precluded.

195 This argument suggests that prohibiting inbreeding for n generations
196 (e.g., three generations by prohibiting sib mating and first and second cousin
197 mating) will extend the expected time to a common ancestor by $n \times \sum_{i=0}^{\infty} (3/4)^i =$
198 $4n$ generations because there will be an initial delay of n generations before
199 a common ancestor can occur, and that same delay will recur if presence in a
200 single mating pair does not result in common ancestry. However, this argu-
201 ment is only rigorous for $n = 1$ because it assumes that the only constraints
202 on matings are from the two alleles being monitored. If $n = 1$, matings
203 involving those alleles will not be impacted by any other alleles because the
204 inbreeding avoidance, lasting only one generation, will terminate as soon as
205 two alleles are in different mating pairs. But if $n > 1$, common ancestry
206 of other alleles being disallowed may make common ancestry for the allele
207 pair being studied more likely. This interference should not be significant
208 for small values of n (generations of inbreeding avoidance) and large values
209 of N (population size) since there will not be many constraints on matings,
210 and $4n$ will be a good approximation, but overestimate, of the time to a
211 common genetic ancestor. However, $4n$ will not be a good approximation as
212 the inbreeding proscription approaches maximum avoidance of inbreeding as
213 discussed below.

214 5. Two progeny per mating pair

215 The constraint of two progeny per mating pair is that every mating pair
216 has exactly two progeny, which provides that the population size is constant.
217 Going backward in time, this requires that exactly two individuals choose
218 each mating pair as parents. This puts significant constraints on the pedigree
219 structure so that prohibiting sib mating does not increase the expected time
220 to the MRCA by 4 generations as is the case with the Poisson (binomial)
221 progeny distribution. Maximum avoidance of inbreeding entails two progeny
222 per mating pair.

223 Studying two progeny per mating pair provides both the extreme case
224 when the variance of the progeny distribution is zero, and the opportunity to
225 study the impact of inbreeding constraints on the time to the MRCA when
226 the variance of the progeny distribution is not changing. (The variance of

227 the progeny distribution of mating pairs is zero, the variance of the progeny
228 distribution at the haploid (gamete) level is .5.) Analogous to above, we
229 vary the amount of full sib mating under the constraint of two progeny per
230 mating pair. As in equations 1-3, T denotes the expected time to a common
231 ancestor of two alleles in an individual, U denotes the expected time to a
232 common ancestor of two alleles in different individuals in a mating pair, and
233 V denotes the expected time to a common ancestor of two alleles in different
234 mating pairs. The fraction of matings which are between full sibs is denoted
235 by s , sN of the mating pairs in a generation are randomly assigned a mating
236 pair the previous generation as parents. This assignment is done without
237 replacement since each mating pair has only two progeny (the individuals in
238 the mating pair assigned to them), and those parental mating pairs are not
239 available to individuals which are not in full sib mating pairs. The remaining
240 parental mating pairs are assigned to exactly two individuals in different
241 mating pairs of the next generation which do not include the full sib mating
242 pairs.

With the constraint of two progeny per mating pair, the above equations (1-3) become

$$T = U + 1 \tag{5}$$

$$U = s(.25 + .25(1 + T) + .5(1 + U)) + (1 - s)(1 + V) \tag{6}$$

$$V = [(1 - s)N((1 - s)N - 1)/(N(N - 1)(2(1 - s)N - 2))] \times$$

$$(.25 + .25(1 + T) + .5(1 + U)) + \tag{7}$$

$$[1 - ((1 - s)N((1 - s)N - 1)/(N(N - 1)(2(1 - s)N - 2))](1 + V).$$

243 The first two equations are the same as (1,2), but the third is altered if there
244 are exactly two progeny per mating pair because alleles in different mating
245 pairs cannot have both chosen the same parents of a sib mating as parents
246 because both of its progeny are in the same sib mating pair. $1/N$ is replaced
247 by $(1 - s)N((1 - s)N - 1)/(N(N - 1))$ which is the fraction of pairs of mating
248 pairs where neither is a result of sib mating times $1/(2(1 - s)N - 2)$ which is
249 the probability two individuals (in distinct mating pairs) from such pairs are
250 siblings (the last -2 entails that both the first individual and its mate reduce
251 the number of parental assignments available for the second individual). The
252 solutions are $T = 6$ for mandatory sib mating ($s = 1$), $T = 8N - 2$ for random
253 mating ($s = 1/(2N - 1)$), and $T = 8N - 2$ for sib mating precluded ($s = 0$).

254 **6. Maximum Avoidance of Inbreeding**

255 A bound on the time to the MRCA is provided by maximum avoidance
256 of inbreeding. Pedigree common ancestry cannot be avoided for more than
257 $\lfloor \log_2 N \rfloor$ generations, where $\lfloor \cdot \rfloor$ denotes the greatest integer function, be-
258 cause after $\lfloor \log_2 N \rfloor$ generations each allele will have $2^{\lfloor \log_2 N \rfloor} / 2$ mating pairs
259 as ancestors, hence the next generation their ancestry will overlap. If in-
260 breeding is avoided for $\lfloor \log_2 N \rfloor$ generations, there will be a pedigree of size
261 $2^{\lfloor \log_2 N \rfloor}$ which manifests a maximum avoidance of inbreeding regular system
262 of inbreeding as studied by Wright (1921) and Kimura and Crow (1963) em-
263 bedded in the pedigree. This embedded pedigree governs genetic identity,
264 and we shall assume that N is an integral power of 2. The maximum avoid-
265 ance of inbreeding pedigrees provide a bound on how much common ancestry
266 can be postponed by avoiding inbreeding.

267 Maximum avoidance of inbreeding includes full sib mating in a population
268 with one mating pair, double first cousin mating in a population with two
269 mating pairs, quadruple second cousin mating in a population with four
270 mating pairs, A schematic for double first cousin mating is presented in
271 Figure 2, which will help explain higher order avoidance of inbreeding. We
272 shall calculate the expected time since a common ancestor of two alleles in
273 an individual under maximum avoidance of inbreeding.

274 Our question is: what is the expected time since two alleles in an indi-
275 vidual in generation 0 had a common genetic ancestor. We shall consider the
276 top left individual in Figure 2: one of its alleles came from one parent and
277 the other from its other parent in generation 1. The possible locations of the
278 ancestral alleles are indicated with a and b . For clarity of exposition (and as
279 allowed by symmetry) we shall assume that the allele (a) which came from
280 the parent on the left side always came from the parent on the left side in
281 previous generations, although all possible locations are indicated in Figure
282 2. We shall first answer the preliminary question: ‘what is the expected time
283 since the other allele had a parent on the left side?’ and then multiply the
284 answer by 2 to address genetic identity.

285 At generation 1, the ancestral allele (b) of the top left individual which
286 is not on the left side is not in a progeny of a parent on the left side. It has
287 two choices for its parent, one of which is the progeny of a parent on the left
288 side, and the other of which is not. Inspection of the pedigree shows that in
289 each generation one individual is on the left side, one individual is the child
290 of an individual on the left side, and the other two are grandchildren, but not

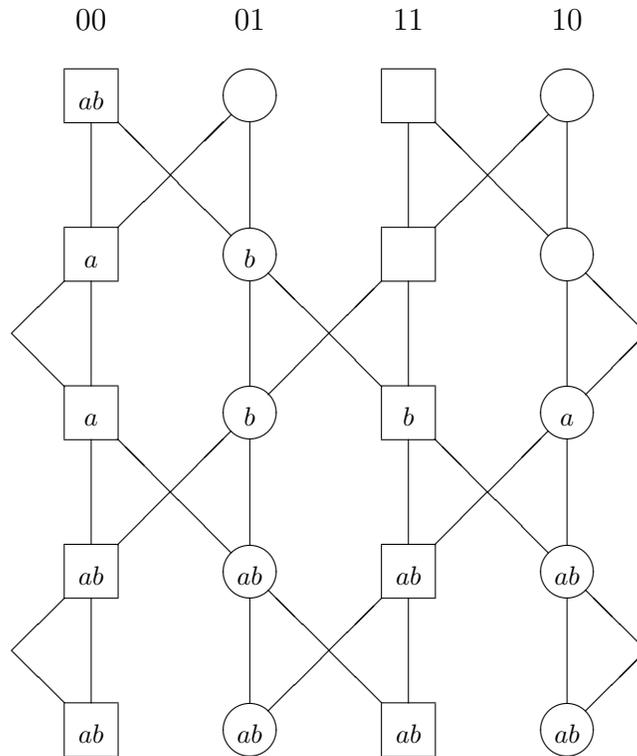


Figure 2: Double First Cousin Mating. Squares represent males, circles represent females, and lines extend upward to progeny, hence downward to parents (diagonal lines on the border continue on the other side as if this figure were wrapped around a cylinder). The figure is symmetric with respect to direction (toward the top or bottom). We shall designate the top row 0 and the rows below 1, 2, ... even though they are going backward in time. The a 's, b 's and 2-digit binary numbers are explained in the text. *cf.* Kimura and Crow (1963).

291 children, of an individual on the left side. Furthermore, if an individual is a
 292 child of an individual on the left side, the other parent is a grandchild, but
 293 not a child of a an individual on the left side; if an individual is a grandchild,
 294 but not a child of an individual on the left side, one parent is a child of an
 295 individual on the left side, but the other parent is a grandchild, but not a
 296 child, of an individual on the left side.

297 The expected time until the other allele (b) is on the left side (i.e., in
 298 the same individual as the allele which remains on the left side) is calculated
 299 by monitoring its probable distance (child or grandchild) from the left side.
 300 Every possible position of the allele in a generation becomes two equally
 301 likely positions the previous generation, so there are 2^{t-1} possible positions
 302 in generation t . But many of these will coincide because there are only four
 303 possible positions in a generation, and symmetry reduces them to three: the
 304 left side, child of the left side, and grandchild of the left side. The left side
 305 represents a common parent, and is treated like an absorbing state. Parents
 306 of children of the left side are the left side half the time and grandchildren
 307 of the left side half the time. Parents of grandchildren of the left side are
 308 children of the left side half the time and grandchildren of the left side half
 309 the time. This provides the recurrence relation for the number of possible
 310 children (c) and grandchildren (g) of the left side in generation t

$$c_t = g_{t-1} \tag{8}$$

$$g_t = c_{t-1} + g_{t-1}. \tag{9}$$

311 Substituting (8) into (9) (which entails shifting the index by 1) provides
 312 the recurrence relation for the Fibonacci numbers $g_t = g_{t-2} + g_{t-1}$ (Chandra
 313 and Weisstein, 2014). In generation 0, the allele (b) is in neither a child nor
 314 grandchild of the left side (by ‘child of the left side’ we mean child of the left
 315 side which is not on the left side), in generation 1, there is a single occurrence
 316 as a grandchild, and this starts the Fibonacci numbers for g , with c being g
 317 with the index shifted by 1.

318 Half the time that the allele is in a child of the left side, it is in the
 319 left side (hence the same individual as the allele (a) which remained on the
 320 left side) the previous generation. Thus the expected time since two alleles
 321 in an individual were previously in an individual is (2^{t-1} converts possible
 322 occurrences to probabilities)

$$2 + \sum_{i=1}^{\infty} i \times \frac{F_i}{2^{i+1}}, \quad (10)$$

323 where F_i is the i^{th} Fibonacci number and the initial 2 is the number of
 324 generations until the allele can be in a child of the left side.

325 When these alleles first choose a common ancestor, only half the time
 326 will they have chosen the same allele within that ancestor, the other half of
 327 the time they will be two alleles in the same individual, which restarts the
 328 clock for time to a common ancestor, and this happens every time a common
 329 ancestor is achieved, hence we must double the time until a common ancestor
 330 to get the time until a common allele within the ancestor:

$$2 \times \left(2 + \sum_{i=1}^{\infty} i \times \frac{F_i}{2^{i+1}} \right) \quad (11)$$

331 is the expected time since a common genetic ancestor of two alleles in the
 332 same individual. This can be summed using Binet's Fibonacci number for-
 333 mula (Weisstein, 2014)

$$F_n = \frac{(1 + \sqrt{5})^n - (1 - \sqrt{5})^n}{2^n \sqrt{5}} \quad (12)$$

334 which produces

$$2 \times \left(2 + \sum_{i=1}^{\infty} i \times \frac{(1+\sqrt{5})^i - (1-\sqrt{5})^i}{2^i \sqrt{5}} \right) =$$

$$4 + \frac{1}{\sqrt{5}} \left(\sum_{i=1}^{\infty} i \times \left(\frac{(1 + \sqrt{5})}{4} \right)^i - \sum_{i=1}^{\infty} i \times \left(\frac{(1 - \sqrt{5})}{4} \right)^i \right) = 14. \quad (13)$$

335 We confirmed this by numerical summation of the first 200 terms, and an al-
 336 ternative derivation which extends to higher order cousin mating is presented
 337 below.

338 It is hard to draw (or comprehend) a schematic for higher order maxi-
 339 mum avoidance of inbreeding, but it can be represented mathematically. A
 340 population of 2^{k+1} individuals can be identified with binary strings of $k + 1$
 341 0's and 1's (i.e, a number between 0 and $2^{k+1} - 1$, inclusive, there are 2^{k+1}

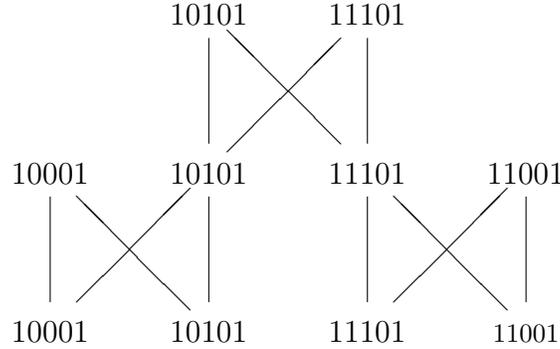


Figure 3: Matings of digitally labelled individuals. The second digit specifies the top mating/sibship and the third digit specifies the bottom matings/sibships.

342 individuals in a generation with 2^k -fold k^{th} cousin mating). Each individual
 343 has a different string (or number). In a generation, matings occur between
 344 individuals which have the same digits (0 or 1) in all positions except the
 345 j -th (the same position for all individuals) at which they have different digits.
 346 This model can be made dioecious (two sex) by having strings with an even
 347 number of 1's be male and strings with an odd number of 1's be female;
 348 because the digits identifying members of mating pairs differ in only one po-
 349 sition, they will be one male and one female). Two progeny are produced by
 350 each mating, and they are given the binary strings of their parents (each is
 351 assigned one of the two strings, so all 2^{k+1} strings occur the next generation,
 352 it does not matter which progeny is assigned which string). Then in the new
 353 generation matings occur between individuals which share the same digits
 354 in all positions except the $(j + 1)^{\text{st}}$ (if j was $k + 1$, then the 1^{st} position is
 355 used; i.e, mod $k + 1$). Figure 2 is consistent with this model if the columns
 356 are labelled 00, 01, 11, and 10, respectively. Figure 3 shows some matings
 357 within a pedigree of 16-fold fourth cousin mating. This model is reversible
 358 for an ancestral analysis which we employ, where individuals form sibships
 359 based upon identity of all but one of the entries in their binary strings and
 360 their parents are assigned the same strings as they are.

361 In order to generalize the analysis of double first cousin mating, we shall
 362 identify the individual $000\dots 0$, which we shall designate as $\mathbf{0}$, with the
 363 left side of figure 2. The notation c and g which suggests children and
 364 grandchildren will be replaced with d_i for individuals which are i generations

365 away from $\mathbf{0}$, d may be remembered as distance, descendant, or degree (the
 366 subscript does not denote time). Under 16-fold fourth cousin mating when
 367 the sibships are defined by the third digit, individual 00100 is in state d_1
 368 because its sib is 00000, hence one of its parents is 00000. Individual 00101
 369 is in state d_3 because its sib, hence one of its parents, is 00001; the sib of
 370 00001 is 00011 because the position defining sibships has shifted by one, and
 371 one of the parents of that sibship is 00001; the sib of 00001 is 00000 because
 372 the position defining sibships has shifted one more, hence 00000 is one of
 373 its parents and 00101 was three generations from $\mathbf{0}$. Individual 10110 is in
 374 state d_4 because it and its ancestors must have the sibs 10010, 10000, 10001
 375 (from which sibship the parent 10000 continues the lineage), and 00000; this
 376 illustrates the mod $k + 1$ counting mentioned above. Generalizations of the
 377 above results follow.

378 For a given locus in a given individual in a mating, there are two indi-
 379 viduals the previous generation which may have supplied the allele there,
 380 four individuals the generation before that, continuing until 2^{k+1} is achieved
 381 $k + 1$ generations prior. The same is true for the other individual in the
 382 mating. Further, the ancestral pools of the individuals in the mating will
 383 not overlap until $k + 1$ generations prior when the ancestral pool for both
 384 individuals is the entire population. The growth of the number of ancestors
 385 occurs because each generation, parents may be either 0 or 1 in the position
 386 defining the sibship/mating (which doubles the number of ancestors). The
 387 ancestral copies of the genetic material in parents of a sibship do not come
 388 together in a sibship until $k + 1$ generations before that sibship because if
 389 the mating is defined by the j^{th} position, the parents and their ancestors will
 390 maintain a difference at that position until the j^{th} position is again the po-
 391 sition where members of mating pairs differ, which will be $k + 1$ generations
 392 earlier. At that time the ancestral copies of their genetic material will be
 393 dispersed throughout 2^k individuals for each parent and come together in all
 394 2^k mating pairs. Postponement of inbreeding may be a better name than
 395 avoidance of inbreeding. This argument confirms that there are the full 2^i
 396 distinct ancestors of each individual i generations prior, because any smaller
 397 ancestral pool would entail genetic material in a mating coming together
 398 before $k + 1$ generations.

399 A related result is that one of the members of every mating pair will be in
 400 state d_{k+1} , because it will have a 1 in the position which defined the mating
 401 pair (following that member's lineage of parents which are 0 in the subsequent
 402 sibship/mating positions will in $k + 1$ generations provide an ancestor with

403 0 in all positions when the 1 in the j^{th} position is changed to 0). The other
 404 member will be in state d_{r-j} (where the subscript is mod $k + 1$) where r is
 405 the position of the last 1 in the sequence common to the members of the
 406 mating pair (each subsequent generation of parents will have a member that
 407 is 0 in the mating position, and when the 1 in the r^{th} position is changed to
 408 0, $\mathbf{0}$ will be attained).

409 For an arbitrary two individuals, they will have a set (perhaps two sets)
 410 of parents (one parent of each individual) which will not have a common
 411 ancestor for $k + 1$ generations, because they will differ in the position which
 412 designates the sibship/mating pair, hence cannot have a common ancestor
 413 until that position is again the position which designates sibships/mating
 414 pairs.

415 The shortest time until a common ancestor of two arbitrary individuals
 416 is the number of positions (generations) after the sibship position until they
 417 have the same digit (both have 0 or both have 1, different positions may
 418 have different digits) in every position following that position (in the sense
 419 of mod $k + 1$). To show this, let r be the last position in which the digits
 420 differ. Then there will be no common ancestor of the two individuals before
 421 the position which defines sibships is r , because the ancestors will differ in
 422 position r . When the mating position is r , the ancestors of both individuals
 423 will have every sequence of 0's and 1's before position r , as generated by the
 424 process of having one parent with 0 and one parent with 1 in the sibship
 425 position. Hence there will be many pairs of one ancestor of each of the
 426 original individuals which share all the digits before r , and all the digits
 427 after r are the same because they will not have changed. If all the digits are
 428 the same, two individuals are the same individual. Therefore the parents of
 429 those mating pairs (i.e., pairs sharing all digits except the r^{th} position when
 430 the mating position is r) will be ancestors of both the original individuals.

431 To calculate the expected time to a common ancestor we start with two
 432 alleles in individual $\mathbf{0}$, which was the left side of Figure 2, and assume one
 433 of the parental alleles is always in individual $\mathbf{0}$ the previous generations.
 434 The parental allele which is not in individual $\mathbf{0}$ will be in state d_{k+1} because
 435 it will be 1 in the position designating the sibship/mating. In general, an
 436 individual in state d_m will have one parent in state d_{k+1} because it will be 1
 437 in the sibship/mating position, and the other parent in state d_{m-1} because
 438 one generation has elapsed so one parent must be one generation closer to $\mathbf{0}$.

439 This allows construction of a recurrence relation for the potential locations
 440 of the allele not in $\mathbf{0}$ analogous to the recurrence relation for double first

441 cousin mating, which will entail generalizations of the Fibonacci numbers
 442 (Noe et al., 2014). But there is an alternative derivation which avoids the
 443 Fibonacci numbers entirely.

444 The expected time since a common ancestor can be obtained recursively
 445 (on k) using the representation of binary 2^{k+1} -tuples to represent the individ-
 446 uals. The trajectory of the allele which did not remain in $\mathbf{0}$ from state d_{k+1}
 447 to $\mathbf{0}$ ($= d_0$) under k^{th} cousin mating is embedded in $k + 1^{\text{st}}$ cousin mating as
 448 the trajectory from state d_{k+2} to d_1 since the transition probabilities are the
 449 same with the index shifted by 1. Under $(k + 1)^{\text{st}}$ cousin mating, state d_1
 450 is followed (preceded in real time) by $\mathbf{0}$ half the time and d_{k+2} half the time
 451 (with one more generation elapsed). Hence it will either terminate in one
 452 more generation, or restart its journey (the additional generation from d_1 to
 453 d_{k+2} under $k + 1^{\text{st}}$ cousin mating corresponds to the initial generation from
 454 $\mathbf{0}$ to d_{k+1} under k^{th} cousin mating). This provides that the expected time to
 455 $\mathbf{0}$ for $k + 1^{\text{st}}$ cousin mating will be 1 more than twice the expected time for
 456 k^{th} cousin mating with the factor of two (“twice”) reflecting the restarting,
 457 and the one being the additional generation when it terminates. Therefore
 458 the time $S(k)$ since a common ancestor for 2^k -fold k^{th} cousin mating satisfies
 459 the recurrence relation

$$S_{k+1} = 1 + 2 \times S_k, \tag{14}$$

460 but since the time until a common genetic ancestor (descendants of the same
 461 allele) is twice the time since a common ancestor was attained, the recurrence
 462 relation for the time since a common genetic ancestor is

$$T_{k+1} = 2 + 2 \times T_k. \tag{15}$$

463 This relation specifies sequence A000918 at oeis.org (OEIS, 2014).

464 We can use the value 14 for double first cousin mating to initiate this
 465 recursion relation, or avoid the Fibonacci numbers by using the initial case
 466 $k = 0$ which is full sib mating (the two progeny mate with each other each
 467 generation). After the two alleles in an individual came from separate par-
 468 ents, they have probability .5 of coming from the same parent each previous
 469 generation. This argument provides the formula $1 + \sum i/2^i$, which is read-
 470 ily summed to 3, hence doubled to 6 for a common genetic ancestor. The
 471 expected time to the MRCA for full sib mating is 6, for double first cousin
 472 mating 14, for quadruple second cousin mating 30, for octuple third cousin

473 mating 62. The general formula for the expected time since a common ge-
474 netic ancestor under (2^k) -fold k^{th} cousin mating is $2^{k+3} - 2 = 8N - 2$ because
475 (2^k) -fold k^{th} cousin mating entails a population of 2^k mating pairs.

476 7. Discussion

477 This manuscript presents explicit formulae for the expected time since a
478 common ancestor for various levels of full sib mating and regular systems of
479 inbreeding. The latter had been studied (Kimura and Crow, 1963; Wright,
480 1921) in the context of homozygosity rather than time since a common an-
481 cestor (if τ is the generation when a common ancestor occurred, they studied
482 $P(\tau < t)$ rather than $E[\tau]$). These results complement those studies. But
483 these results also provide new insights.

484 The most striking result is that the expected time since a common ances-
485 tor is $8N - 2$ for random mating with two progeny per mating pair, full sib
486 mating precluded with two progeny per mating pair, and maximum avoid-
487 ance of inbreeding (which entails two progeny per mating pair). This is not
488 always the expected time since a common ancestor with two progeny per
489 mating pair; equations (5-7) do not have a constant solution, and multiple
490 instances of double first cousin mating or other regular systems of inbreeding
491 in a larger population will provide other exceptions. But contrasting $8N - 2$
492 to $4N + 2$ which is the expected time since a common ancestor under ran-
493 dom mating with a Poisson progeny distribution suggests that the resultant
494 progeny distribution may be a significant determinant of the expected time
495 since a common ancestor when the pedigree structure is changed.

496 Kingman (1982a) demonstrated that the coalescent time is inversely re-
497 lated to the variance of the progeny distribution if the conditions of exchange-
498 ability and independence are met. The variances of the progeny distribution
499 for mating pairs are 2 for the Poisson progeny distribution and 0 for two
500 progeny per mating pair, but these variances are 1 and .5, respectively,
501 at the level of the allele (gamete), which is the basis of the coalescent. This
502 is consistent with the time to the MRCA with two progeny per mating pair
503 being twice the time for the Poisson progeny distribution. This suggests
504 that the variance of the progeny distribution is a primary determinant of
505 the time to the MRCA, which is modified dependent on the extent to which
506 the mating structure deviates from or is equivalent to exchangeability and
507 independence.

508 Imposing sib mating on a random mating population with a Poisson
509 progeny distribution decreased the expected time to the MRCA linearly from
510 $4N + 6$ with no sib mating to 6 with mandatory sib mating, but some of this
511 change may be an indirect effect of the variance of the progeny distribu-
512 tion increasing from 2 to 4. With two progeny per mating pair imposing
513 sib mating does not change the variance of the progeny distribution and the
514 non-linear decrease in the time to the MRCA from $8N - 2$ to 6 is due di-
515 rectly to the change in mating structure. The time $8N - 2$ for maximum
516 avoidance of inbreeding suggests that that must in some sense be equivalent
517 to exchangeability and independence.

518 We can address the question of whether a few generations of pedigree
519 information provides sufficient inbreeding information to approximate the
520 expected time to the MRCA. The observation that each generation that
521 inbreeding is avoided extends the expected time to a common ancestor by
522 about four generations in a large population suggests that the answer is no.
523 For example, six generations of inbreeding avoidance extends the time to
524 the MRCA by only approximately $4 \times 6 = 24$ generations, which is a small
525 number compared to the difference between $4N + 2$ for random mating and
526 $8N - 2$ for maximum avoidance of inbreeding if N is large, hence additional
527 mating information prior to those generations is needed to determine where
528 between $4N + 26$ and $8N - 2$ the expected time since the MRCA lies.

529 For a concrete example, consider a population of $N = 2^{12} = 4096$ mating
530 pairs. The expected time since a common genetic ancestor is 16386 gener-
531 ations under random mating with monogamous mating pairs but 32766
532 generations with random mating with monogamous mating pairs and two
533 progeny per mating pair. If there is not inbreeding for 6 generations, the
534 expected time since a common ancestor could be as low as 510 generations
535 if there are 64 subpopulations practicing 64-fold sixth cousin mating or as
536 high as 32766 generations with 4096-fold twelfth cousin mating. Thus the
537 mating structure including earlier generations is important (and the progeny
538 distribution reflecting which lineages do not go extinct).

539 Avoidance (postponement) of inbreeding and reduced variance of the
540 progeny distribution will increase the time to the MRCA. Maximum avoid-
541 ance of inbreeding which has zero variance of the progeny distribution doubles
542 the expected time since a common ancestor compared to random mating with
543 a Poisson progeny distribution. But avoidance of inbreeding and progeny
544 distribution are interrelated but not totally dependent. The entire pedigree
545 determines the expected time since the MRCA, one cannot calculate the

546 expected time based solely on the variance of the progeny distribution.

547 **References**

548 **8. References**

549 Arzberger, P. 1988. *Results for Generalized Regular Inbreeding Systems*. J.
550 Math. Biol. **26**, 535–550.

551 Bittles, A. H., Egerbladh, I. 2005. *The Influence of Past Endogamy and Con-*
552 *sanguinity on Genetic Disorders in Northern Sweden*. Annals of Human
553 Genetics **69**, 549–558.

554 Campbell, R. B. 1993. *The Importance of Mating Structure versus Progeny*
555 *Distribution for Genetic Identity under Mutation*. Theoretical Population
556 Biology **43**, 129–140.

557 Campbell, R. B. 1999. *The Coalescent Time in the Presence of Background*
558 *Fertility Selection*. Theoretical Population Biology **55**, 260–269.

559 Cavalli-Sforza, L. L., Bodmer, W. F. 1971. *The Genetics of Human Popula-*
560 *tions*. W. H. Freeman, San Francisco.

561 Chandra, P., Weisstein, E. W. 2014. *Fibonacci*
562 *Number*, MATHWORLD – A Wolfram Resource.
563 <http://mathworld.wolfram.com/FibonacciNumber.html>

564 Crow, J. F., Kimura, M. 1970 *An Introduction to Population Genetics The-*
565 *ory*. Harper and Row. New York.

566 Hein, J., Schierup, M. H., Wiuf, C., 2005. *Gene Genealogies, Variation and*
567 *Evolution – A Primer in Coalescent Theory*. Oxford University Press, Ox-
568 ford.

569 Kimura, M., Crow, J. F. 1963. *On the Maximum Avoidance of Inbreeding*.
570 Genetical Research, Camb. **4**, 399–415.

571 Kingman, J. F. C. 1982a. *On the Genealogy of Large Populations*. Journal of
572 Applied Probability, **19A**, 27–43.

573 Kingman, J. F. C. 1982b. *The Coalescent*. Stochastic Processes and their
574 Applications, **13**, 235–248.

- 575 Krawczak, M., Barnes, R. H. 2010. *How Obedience of Marriage Rules May*
576 *Counteract Genetic Drift*. J. Community Genetics, **1**, 23–28.
- 577 Lachance, J. 2009 *Inbreeding, Pedigree Size, and the most recent Common*
578 *Ancestor of Humanity*. J.Theor. Biol. **261**(2), 238–247.
- 579 Noe, T., Piezas, T., Weisstein, E. W. 2014. *Fibonacci n-*
580 *Step Number*. MATHWORLD–A Wolfram Web Resource.
581 <http://mathworld.wolfram.com/Fibonacci-StepNumber.html>
- 582 OEIS Foundation Inc. 2014. *The On-Line Encyclopedia of Integer Sequences*.
583 <http://oeis.org/A000918>.
- 584 Pettay, J. E., Helle, S., Jokela, J., Lummaa, V. 2007. *Natural Se-*
585 *lection of Female Life-History Traits in Relation to Socio-Economic*
586 *Class in Pre-Industrial Human Populations*. PLoS ONE **2**(7): e603,
587 doi:10.1371/journal.pone.0000606.
- 588 Wakeley, J. 2008. *Coalescent Theory: An Introduction*. Roberts & Company
589 Publishers, Greenwood Village, Colorado.
- 590 Wakeley, J., King, L., Low, B. S., Ramachandran, S. 2012. *Gene Genealo-*
591 *gies within a Fixed Pedigree, and the Robustness of Kingman’s Coalescent*.
592 Genetics **190**, 1433–1445.
- 593 Weisstein, E. W. 2014. *Binet’s Fibonacci Number*
594 *Formula*. MATHWORLD – A Wolfram Resource,
595 <http://mathworld.wolfram.com/BinetsFibonacciNumberFormula.html>
- 596 Wright, S. 1921. *Systems of Mating. II. The Effects of Inbreeding on the*
597 *Genetic Composition of a Population*. Genetics **6**, 124–143.

598 **Appendix A. An Alternative Model**

599 This appendix considers an alternative model to generate the proportion
600 of sib mating. Let ζ be the probability that two individuals choose the same
601 parents. It will also be the probability that a mating pair are sibs (because
602 they are two individuals), which was denoted by s in the previous model.
603 Attaining a particular value of ζ will constrain the progeny distribution (and
604 vice-versa). If $\zeta = 1$, then there is only one mating pair in the parental gener-
605 ation that has progeny, since all individuals share parents; choosing parents

606 will not be random, but will be constrained by the progeny distribution. It
 607 is not possible to have $\varsigma = 0$, since there are $2N$ individuals and only N sets
 608 of parents, so some individuals will need to share parents. The smallest ς
 609 can be is $N/(2N(2N - 1)/2) = 1/(2N - 1)$ since two progeny per mating
 610 pair minimizes the number of pairs sharing parents and $2N(2N - 1)/2$ is the
 611 number of possible pairs of $2N$ individuals.

With this model

$$T = 1 + \varsigma(.25 + .25(T + 1) + .5T) + (1 - \varsigma)T \quad (\text{A.1})$$

612 [$T = U + 1 = V + 1$; this is just the U equation above] because it does not
 613 matter whether or not two individuals are in a mating pair, hence $U = V$ as
 614 defined above. This has the solution $T = 4/\varsigma + 2$. If $\varsigma = 1$, $T = 6$ (but in this
 615 model the pedigree shrinks to a single mating pair in one generation which
 616 is not the case with the above model, so the time until a common ancestor
 617 of the entire population will be different). If $\varsigma = 1/N$, $T = 4N + 2$ (and
 618 this is indeed the random mating model with Poisson progeny distribution
 619 as above). If $\varsigma = 1/(2N - 1)$, $T = 8N - 2$, which is the expected time to a
 620 common ancestor with two progeny per mating pair and random mating.