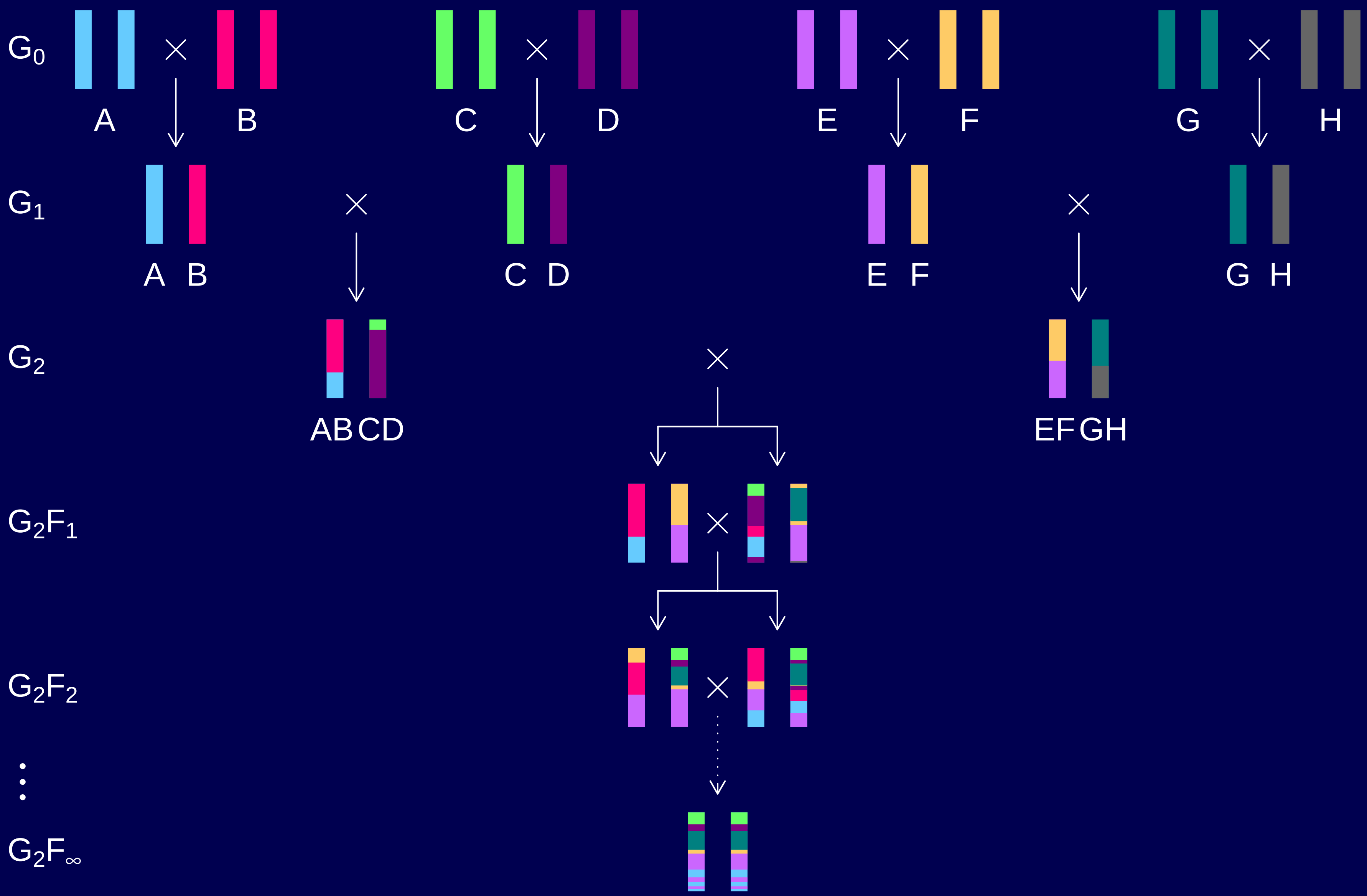
Genotype probabilities in the pre-CC

Karl W Broman

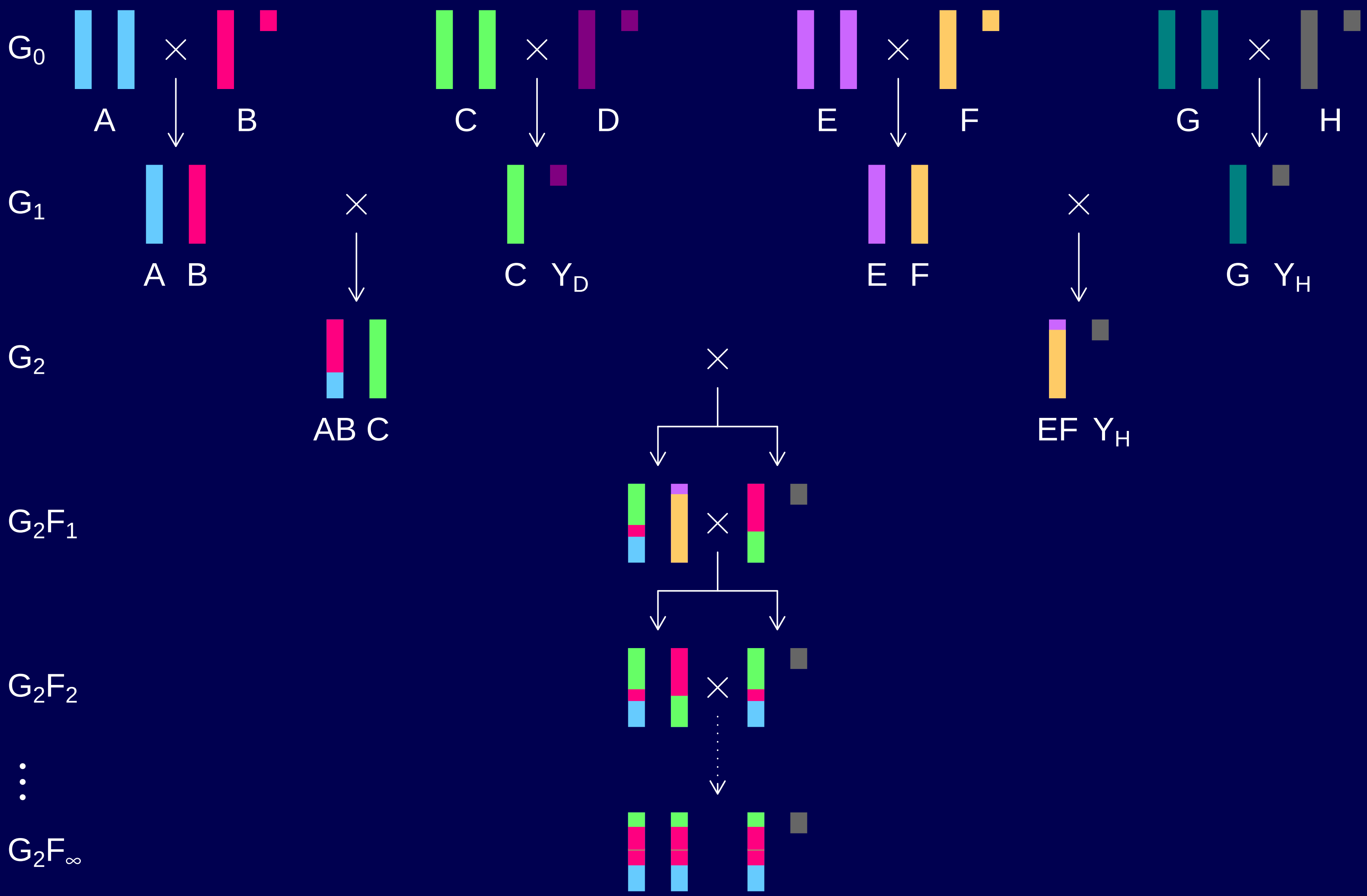
Department of Biostatistics & Medical Informatics
University of Wisconsin – Madison

www.biostat.wisc.edu/~kbroman

The Collaborative Cross



X in the CC



In the CC

Autosomes

$\Pr(G_1 = i) = 1/8$
$\Pr(G_2 = j \mid G_1 = i) = r/(1 + 6r)$ for $i \neq j$

$\Pr(G_2 \neq G_1) = 7r/(1 + 6r)$

X chromosome

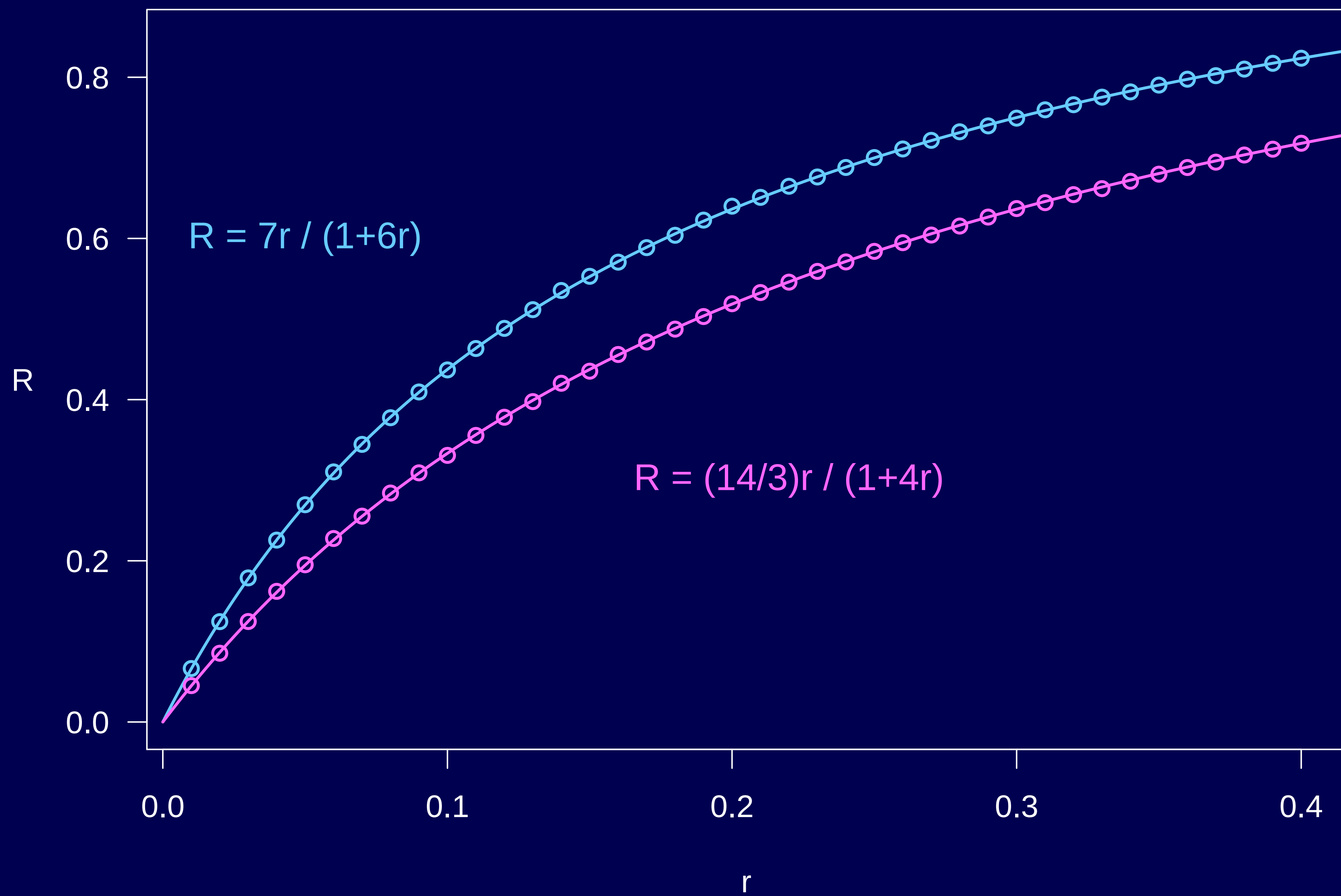
$\Pr(G_1 = A) = \Pr(G_1 = B) = \Pr(G_1 = E) = \Pr(G_1 = F) = 1/6$
$\Pr(G_1 = C) = 1/3$

$\Pr(G_2 = B \mid G_1 = A) = r/(1 + 4r)$
$\Pr(G_2 = C \mid G_1 = A) = 2r/(1 + 4r)$
$\Pr(G_2 = A \mid G_1 = C) = r/(1 + 4r)$

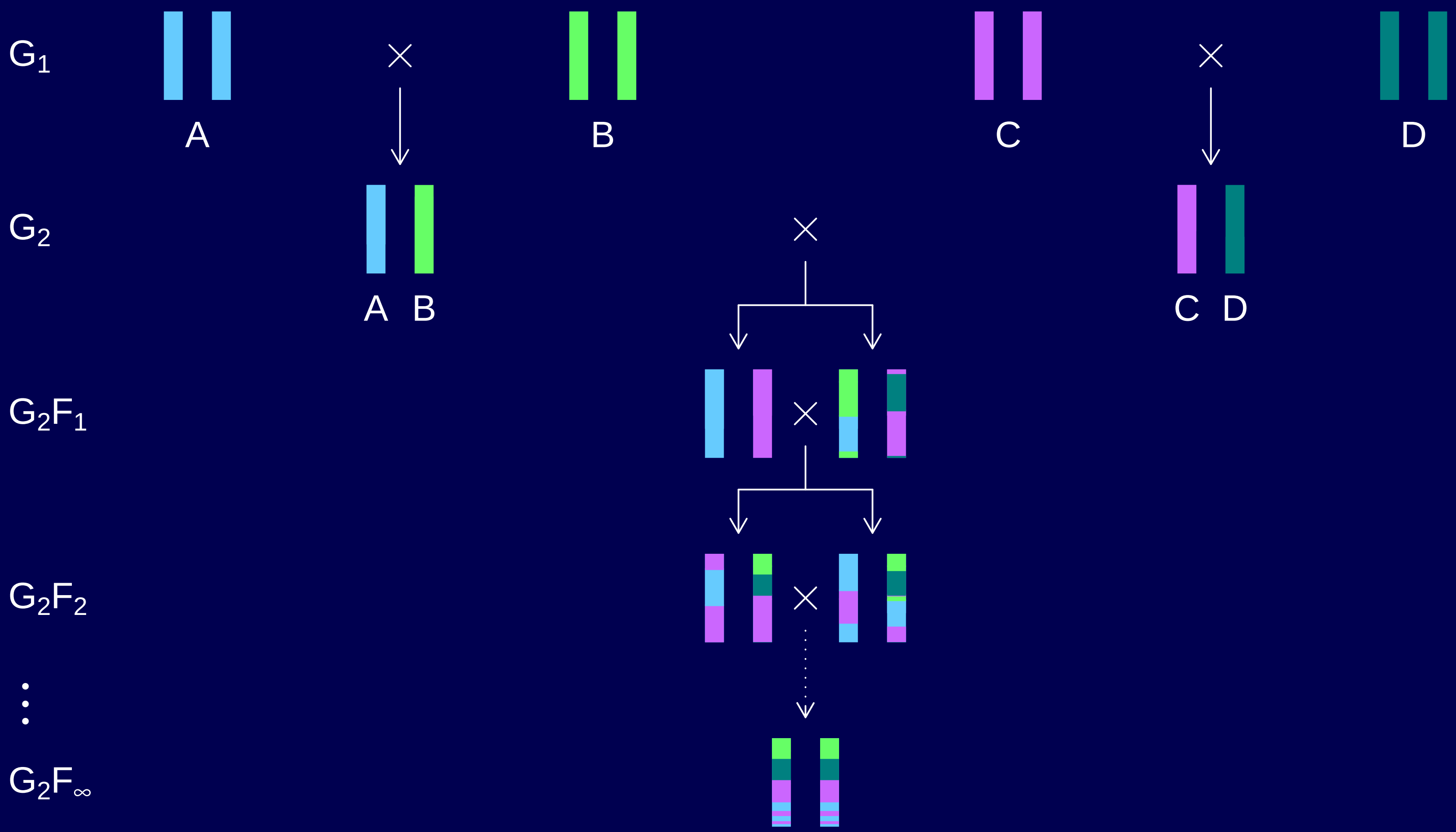
$\Pr(G_2 \neq G_1) = (14/3)r/(1 + 4r)$

Broman, Genetics 169:1133–1146, 2005

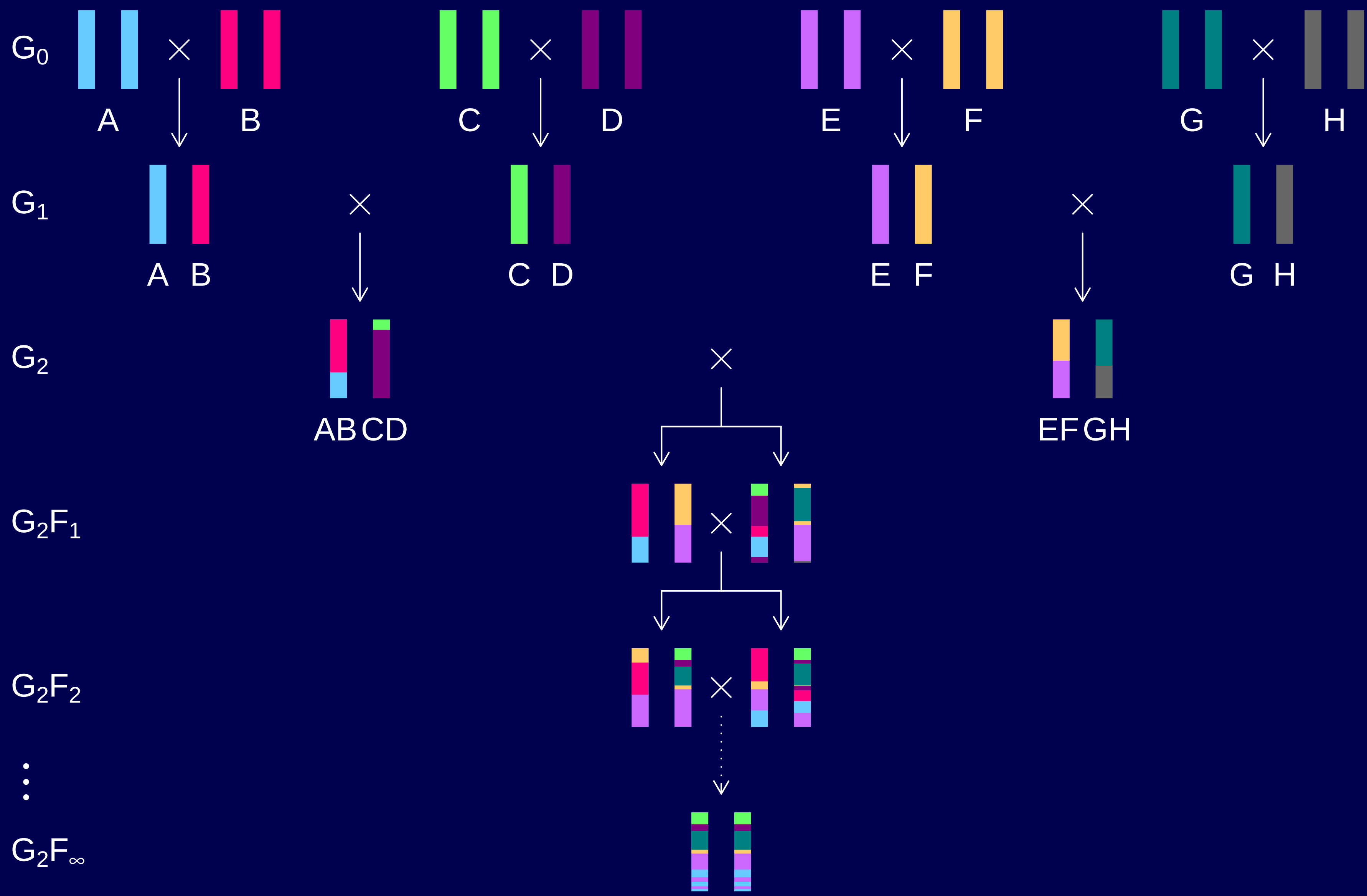
Computer simulations



4-way RIL



8-way RIL



The PreCC

What happens at G_2F_k?

$\Pr(g_1 = i)$ as a function of k

$\Pr(g_1 = i, g_2 = j)$ as a function of k and the recombination fraction

...ideally, the joint probabilities for the pair of individuals.

...but, at least, for a random individual.

1 locus, X chr

G_2 AB|C

1 locus, X chr

G_2 AB|C

G_2F_1 AC|A AC|B BC|A BC|B

1 locus, X chr

G_2	AB|C			
G_2F_1	AC|A	AC|B	BC|A	BC|B
	↓			
G_2F_2	AA|A	AA|C	AC|A	AC|C

1 locus, X chr

G_2	AB|C				
G_2F_1	AC|A	AC|B	BC|A	BC|B	
		↓			
G_2F_2	AA|A	AA|C	AC|A	AC|C	
		AB|A	AB|C	BC|A	BC|C

1 locus, X chr

G_2	AB|C					
G_2F_1	AC|A	AC|B	BC|A	BC|B		
			↓			
G_2F_2	AA|A	AA|C	AC|A	AC|C		
		AB|A	AB|C	BC|A	BC|C	
			AB|B	AB|C	AC|B	AC|C

1 locus, X chr

G_2	AB|C						
G_2F_1	AC|A	AC|B	BC|A	BC|B			
				↓			
G_2F_2	AA|A	AA|C	AC|A	AC|C			
		AB|A	AB|C	BC|A	BC|C		
			AB|B	AB|C	AC|B	AC|C	
				BB|B	BB|C	BC|B	BC|C

1 locus, X chr

G_2	AB|C						
G_2F_1	AC|A	AC|B	BC|A	BC|B			
G_2F_2	AA|A	AA|C	AC|A	AC|C			
		AB|A	AB|C	BC|A	BC|C		
			AB|B	AB|C	AC|B	AC|C	
				BB|B	BB|C	BC|B	BC|C

1 locus, X chr

G_2 AB|C

G_2F_1 AC|A AC|B BC|A BC|B

G_2F_2 AA|A AA|C AC|A AC|C

AB|A AB|C BC|A BC|C

AB|B AB|C AC|B AC|C

BB|B BB|C BC|B BC|C

Females	AA, BB	1/8 each
	AC, BC	1/4 each
	AB	1/4 each
Males	A, B	1/4 each
	C	1/2

Markov chains

- Sequence of random variables $\{X_0, X_1, X_2, \ldots\}$ satisfying

$$\Pr(X_{n+1} \mid X_0, X_1, \ldots, X_n) = \Pr(X_{n+1} \mid X_n)$$

- Here, X_n = "parental type" at generation n.

- Transition probabilities $P_{ij} = \Pr(X_{n+1} = j \mid X_n = i)$

- If π_0 is the starting distribution (at G_2),

  $\pi_k = \pi_0 P^k$ is the distribution at G_2F_k

1 locus, X chr, male

	A	B	C
G_2	0	0	1
G_2F_1	1/2	1/2	0
G_2F_2	1/4	1/4	1/2
G_2F_3	·	·	1/4
G_2F_4	·	·	3/8
G_2F_5	·	·	5/16
G_2F_6	·	·	11/32
G_2F_7	·	·	21/64
G_2F_8	·	·	43/128
G_2F_9	·	·	85/256

1 locus, X chr, male

	A	B	C
G_2	0	0	1
G_2F_1	1/2	1/2	0
G_2F_2	1/4	1/4	1/2
G_2F_3	·	·	1/4
G_2F_4	·	·	3/8
G_2F_5	·	·	5/16
G_2F_6	·	·	11/32
G_2F_7	·	·	21/64
G_2F_8	·	·	43/128
G_2F_9	·	·	85/256

G_2F_k (for $k \geq 1$): $\Pr(C) = 1 - \sum_{i=0}^{k-1} \left(-\frac{1}{2}\right)^i$

1 locus, X chr, female

	AA, BB	AC, BC	AB	CC
G_2	0	0	1	0
G_2F_1	0	1/2	0	0
G_2F_2	1/8	2/8	2/8	0
G_2F_3	1/8	2/8	1/8	1/8
G_2F_4	3/16	3/16	2/16	2/16
G_2F_5	13/64	10/64	6/64	12/64
G_2F_6	15/64	8/64	5/64	13/64
G_2F_7	32/128	13/128	8/128	30/128
G_2F_8	137/512	42/512	26/512	128/512
G_2F_9	143/512	34/512	21/512	137/512

1 locus, autosome, 1 ind'l

	AA,BB,CC,DD	AB,CD	AC,AD,BC,BD
G_2	0	1/2	0
G_2F_1	0	0	1/4
G_2F_2	1/16	2/16	2/16
G_2F_3	3/32	2/32	4/32
G_2F_4	4/32	2/32	3/32
G_2F_5	19/128	6/128	10/128
G_2F_6	43/256	10/256	16/256
G_2F_7	47/256	8/256	13/256
G_2F_8	201/1024	26/1024	42/1024
G_2F_9	423/2048	42/2048	68/2048

Summary

		1 locus	2 loc, hap	2 loc, gen
X chr	Male	✓	✓	—
	Female	✓	✓	×
	Pair	✓	×	×
Autosome	Ind'l	✓	✓	×
	Pair	✓	×	×

Kinship coefficients

$\Phi_{11}^{(k)} =$ kinship coefficient of ind'l w/ itself at G_2F_k

$\Phi_{12}^{(k)} =$ kinship coefficient of sibs at G_2F_k

$$\Phi_{11}^{(0)} = \frac{1}{2},\ \Phi_{12}^{(0)} = 0$$

$$\Phi_{11}^{(k+1)} = \frac{1}{2} + \frac{1}{2}\Phi_{12}^{(k)}$$

$$\Phi_{12}^{(k+1)} = \frac{1}{2}\Phi_{11}^{(k)} + \frac{1}{2}\Phi_{12}^{(k)}$$

$$\Phi_{12}^{(k+2)} = \frac{1}{2}\Phi_{12}^{(k+1)} + \frac{1}{4}\Phi_{12}^{(k)} + \frac{1}{4}$$

$$\Phi_{12}^{(k+3)} = \frac{3}{2}\Phi_{12}^{(k+2)} - \frac{1}{4}\Phi_{12}^{(k+1)} - \frac{1}{4}\Phi_{12}^{(k)}$$

Fibonacci numbers

0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, 377, 610, 987, . . .

$x_k = x_{k-1} + x_{k-2}$, where $x_0 = 0, x_1 = 1$

Characteristic polynomial: $x^2 - x - 1$

Roots: $\varphi = \frac{1+\sqrt{5}}{2}, 1 - \varphi = \frac{1-\sqrt{5}}{2}$

φ is the “golden ratio”

$x_k = A\varphi^k + B(1-\varphi)^k = \ldots = \frac{\varphi^k - (1-\varphi)^k}{\sqrt{5}}$

1 locus, X chr, female

	AA, BB	AC, BC	AB	CC
G_2	0	0	1	0
G_2F_1	0	1/2	0	0
G_2F_2	1/8	2/8	2/8	0
G_2F_3	1/8	2/8	1/8	1/8
G_2F_4	3/16	3/16	2/16	2/16
G_2F_5	13/64	10/64	6/64	12/64
G_2F_6	15/64	8/64	5/64	13/64
G_2F_7	32/128	13/128	8/128	30/128
G_2F_8	137/512	42/512	26/512	128/512
G_2F_9	143/512	34/512	21/512	137/512

1 locus, X chr, female

Pr(AC) satisfies $a_{k+2} = \frac{1}{2}a_{k+1} + \frac{1}{4}a_k$ with $a_0 = 0, a_1 = \frac{1}{2}$

$$a_k = \frac{\sqrt{5}}{5}\left[\left(\frac{1+\sqrt{5}}{4}\right)^k - \left(\frac{1-\sqrt{5}}{4}\right)^k\right]$$

Pr(AB) satisfies $a_{k+2} = \frac{1}{2}a_{k+1} + \frac{1}{4}a_k$ with $a_0 = 1, a_1 = 0$

$$a_k = \left(\frac{5-\sqrt{5}}{10}\right)\left(\frac{1+\sqrt{5}}{4}\right)^k + \left(\frac{5+\sqrt{5}}{10}\right)\left(\frac{1-\sqrt{5}}{4}\right)^k$$

Pr(AA)

$$a_k = \frac{1}{3} + \left(\frac{1}{6}\right)\left(-\frac{1}{2}\right)^k - \left(\frac{5+\sqrt{5}}{20}\right)\left(\frac{1+\sqrt{5}}{4}\right)^k - \left(\frac{5-\sqrt{5}}{20}\right)\left(\frac{1-\sqrt{5}}{4}\right)^k$$

Pr(CC)

$$a_k = \frac{1}{3} + \left(\frac{1}{6}\right)\left(-\frac{1}{2}\right)^{k-1} - \left(\frac{5+\sqrt{5}}{20}\right)\left(\frac{1+\sqrt{5}}{4}\right)^{k-1} - \left(\frac{5-\sqrt{5}}{20}\right)\left(\frac{1-\sqrt{5}}{4}\right)^{k-1}$$

1 locus, autosome, 1 ind'l

	AA,BB,CC,DD	AB,CD	AC,AD,BC,BD
G_2	0	1/2	0
G_2F_1	0	0	1/4
G_2F_2	1/16	2/16	2/16
G_2F_3	3/32	2/32	4/32
G_2F_4	4/32	2/32	3/32
G_2F_5	19/128	6/128	10/128
G_2F_6	43/256	10/256	16/256
G_2F_7	47/256	8/256	13/256
G_2F_8	201/1024	26/1024	42/1024
G_2F_9	423/2048	42/2048	68/2048

1 locus, autosome, 1 ind'l

$\Pr(AC)$ satisfies $a_{k+2} = \frac{1}{2}a_{k+1} + \frac{1}{4}a_k$ with $a_0 = 0, a_1 = \frac{1}{4}$

$$a_k = \frac{\sqrt{5}}{10}\left[\left(\frac{1+\sqrt{5}}{4}\right)^k - \left(\frac{1-\sqrt{5}}{4}\right)^k\right]$$

$\Pr(AB)$ satisfies $a_{k+2} = \frac{1}{2}a_{k+1} + \frac{1}{4}a_k$ with $a_0 = \frac{1}{2}, a_1 = 0$

$$a_k = \left(\frac{5-\sqrt{5}}{20}\right)\left(\frac{1+\sqrt{5}}{4}\right)^k + \left(\frac{5+\sqrt{5}}{20}\right)\left(\frac{1-\sqrt{5}}{4}\right)^k$$

$\Pr(AA)$

$$a_k = \frac{1}{4} - \left(\frac{5+2\sqrt{5}}{40}\right)\left(\frac{1+\sqrt{5}}{4}\right)^{k-1} - \left(\frac{5-2\sqrt{5}}{40}\right)\left(\frac{1-\sqrt{5}}{4}\right)^{k-1}$$

2-locus hap, X chr, male

$r = 0.01 \qquad a_{k+4} = \frac{199}{200} a_{k+3} + \frac{199}{400} a_{k+2} - \frac{74}{200} a_{k+1} - \frac{49}{400} a_k$

$r = 0.02 \qquad a_{k+4} = \frac{198}{200} a_{k+3} + \frac{198}{400} a_{k+2} - \frac{73}{200} a_{k+1} - \frac{48}{400} a_k$

$r = 0.03 \qquad a_{k+4} = \frac{197}{200} a_{k+3} + \frac{197}{400} a_{k+2} - \frac{72}{200} a_{k+1} - \frac{47}{400} a_k$

$$a_{k+4} = \left(\frac{2-r}{r}\right) a_{k+3} + \left(\frac{2-r}{4}\right) a_{k+2} - \left(\frac{3-4r}{8}\right) a_{k+1} - \left(\frac{1-2r}{8}\right) a_k$$

$$\Pr(CC) = \frac{1}{12r+3} + \frac{2}{3r+3}\left(-\frac{1}{2}\right)^k + C\left(\frac{1-r+\sqrt{r^2-10r+5}}{4}\right)^k + D\left(\frac{1-r-\sqrt{r^2-10r+5}}{4}\right)^k$$

where $C = \frac{2r^4+(2r^3-r^2+r)\sqrt{r^2-10r+5}-19r^3+5r}{4r^4-35r^3-29r^2+15r+5}$

$D = \frac{2r^4-(2r^3-r^2+r)\sqrt{r^2-10r+5}-19r^3+5r}{4r^4-35r^3-29r^2+15r+5}$

$\Pr(AC), \Pr(AB), \Pr(AA)$ similarly obnoxious

Fixation probability

Consider a single autosomal locus.

Consider a particular allele (say A).

Pool others into a single allele (B).

Want $\Pr(X_k = AA \times AA)$ [or really $4\times$ that]

Transition matrix, P:

	AA×BB	AA×AB	AB×AB	AB×BB	AA×AA	BB×BB
AA×BB	0	0	1	0	0	0
AA×AB	0	$\frac{1}{2}$	$\frac{1}{4}$	0	$\frac{1}{4}$	0
AB×AB	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{16}$	$\frac{1}{16}$
AB×BB	0	0	$\frac{1}{4}$	$\frac{1}{2}$	0	$\frac{1}{4}$
AA×AA	0	0	0	0	1	0
BB×BB	0	0	0	0	0	1

Fixation probability

We seek $\pi_k = \pi_0 P^k$ where $\pi_0 = (0\ 0\ 0\ 1\ 0\ 0)$ [start at AB×BB].

Write $P = VDV^{-1}$

where D = diagonal matrix of eigenvalues, V = eigenvectors

So $P^k = VD^kV^{-1}$, and so $\pi_k = \pi_0 VD^kV^{-1}$.

Thus, the fixation probability for 4- and 8-way RIL is

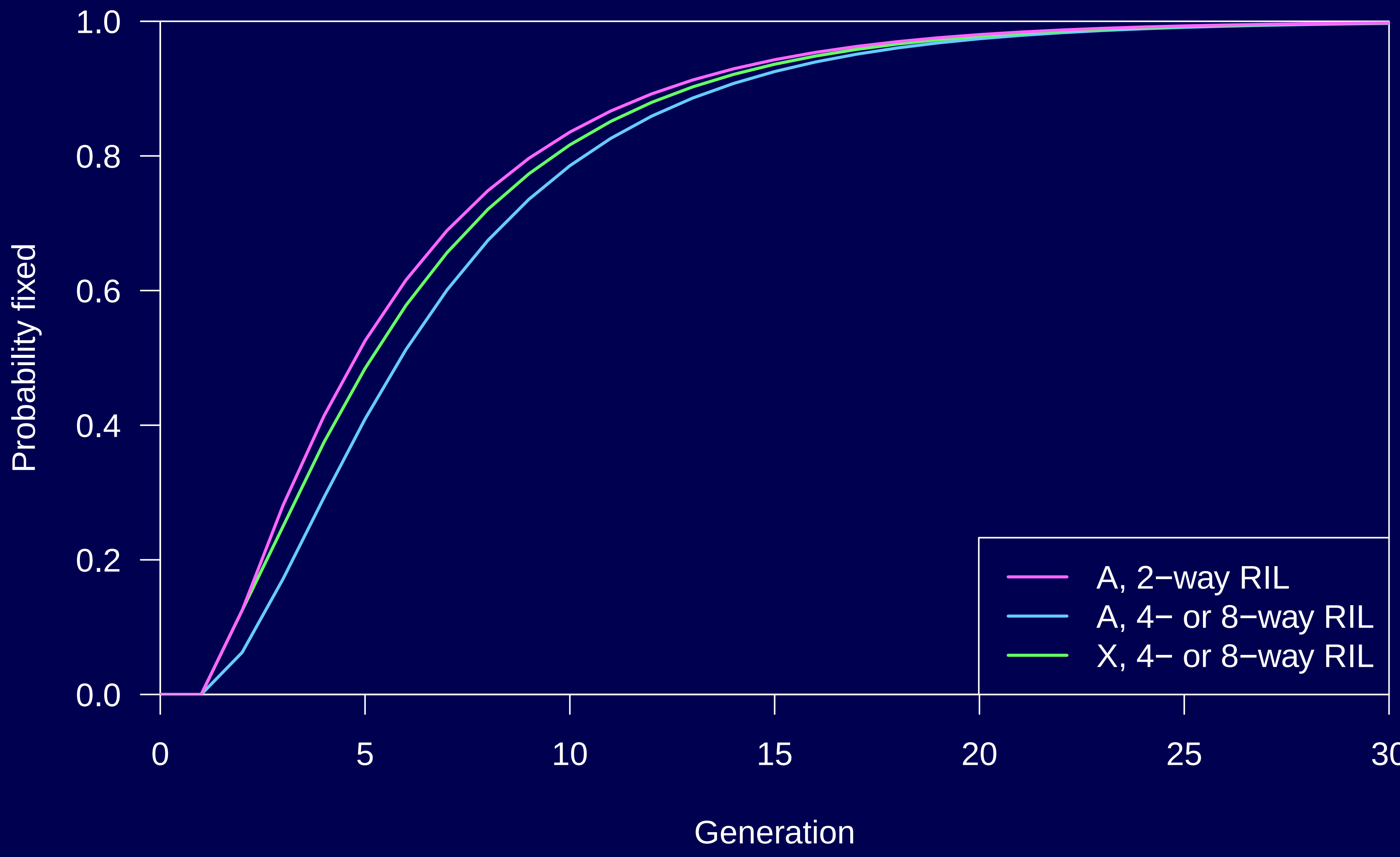
$$1 + \left(\frac{1}{2}\right)^k - \frac{1}{5}\left(\frac{1}{4}\right)^k - \left(\frac{9-4\sqrt{5}}{10}\right)\left(\frac{1-\sqrt{5}}{4}\right)^k - \left(\frac{9+4\sqrt{5}}{10}\right)\left(\frac{1+\sqrt{5}}{4}\right)^k$$

For 2-way RIL, we use $\pi_0 = (1\ 0\ 0\ 0\ 0\ 0)$ [start at AA×BB], and obtain

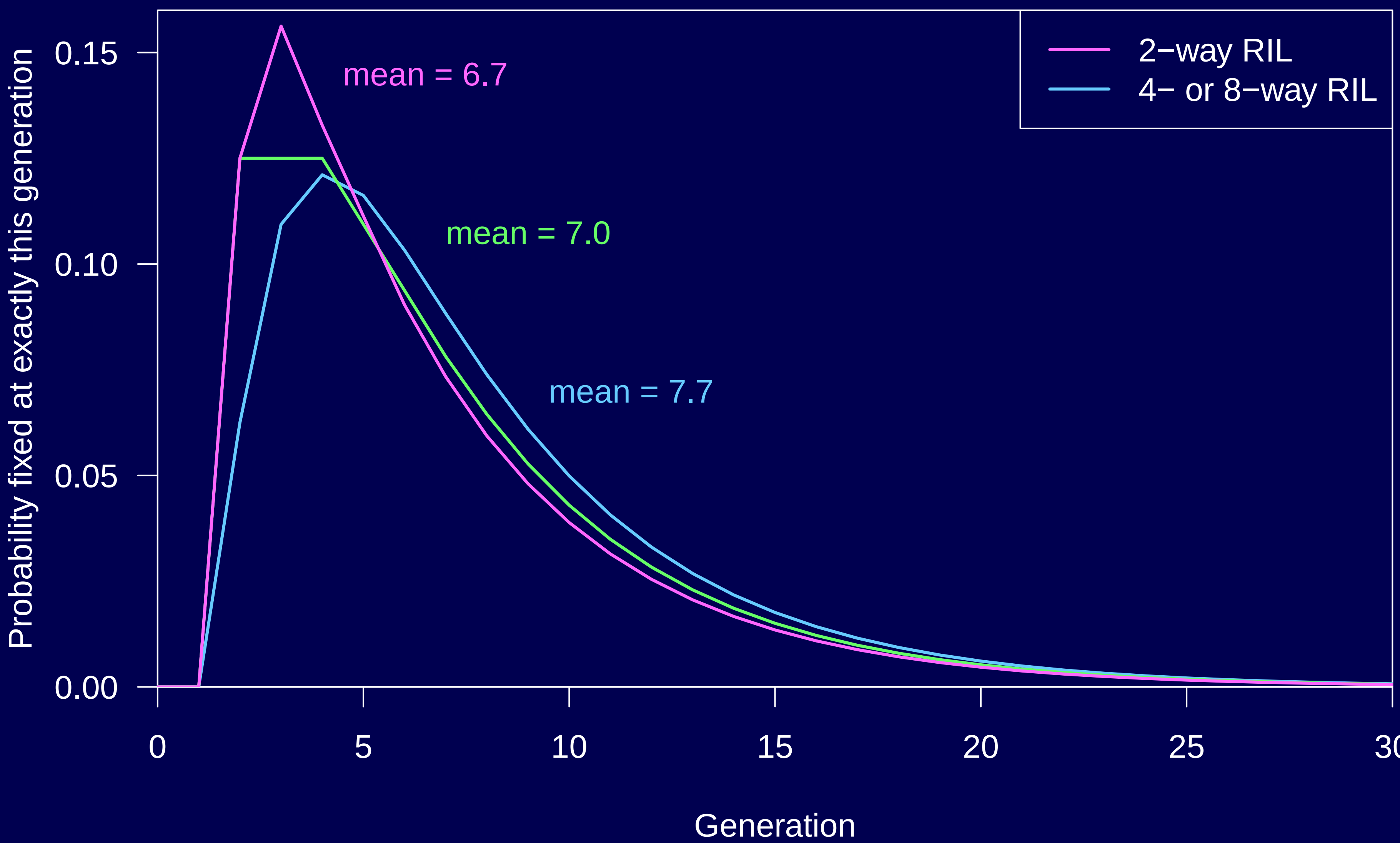
$$1 + \left(\frac{2}{5}\right)\left(\frac{1}{4}\right)^k - \left(\frac{7-3\sqrt{5}}{10}\right)\left(\frac{1-\sqrt{5}}{4}\right)^k - \left(\frac{7+3\sqrt{5}}{10}\right)\left(\frac{1+\sqrt{5}}{4}\right)^k$$

[Caveat: meaning of generation k somewhat different in 2-, 4-, and 8-way RIL]

Fixation probability



Time of fixation



Kimura (1963)

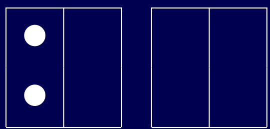
$C_k(xy)$ = Pr(random chr from gen k has x at locus 1 and y at locus 2)

$S_k(xy)$ = Pr(random chr from gen k has x at locus 1 and opposite chr in that ind'l has y at locus 2)

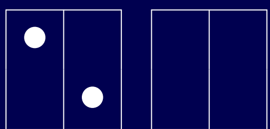
$T_k(xy)$ = Pr(random chr from gen k has x at locus 1 and random chr from other individual has y at locus 2)

Kimura, Biometrics 19:1–17, 1963

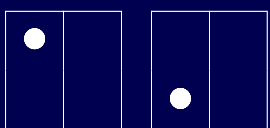
Kimura (1963)



$C_k(xy)$ = Pr(random chr from gen k has x at locus 1 and y at locus 2)



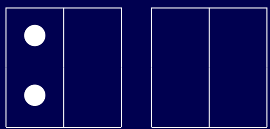
$S_k(xy)$ = Pr(random chr from gen k has x at locus 1 and opposite chr in that ind'l has y at locus 2)



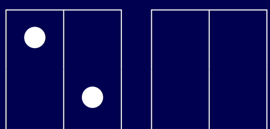
$T_k(xy)$ = Pr(random chr from gen k has x at locus 1 and random chr from other individual has y at locus 2)

Kimura, Biometrics 19:1–17, 1963

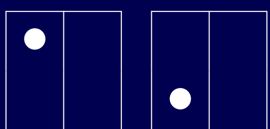
Kimura (1963)



$C_k(xy)$ = Pr(random chr from gen k has x at locus 1 and y at locus 2)



$S_k(xy)$ = Pr(random chr from gen k has x at locus 1 and opposite chr in that ind'l has y at locus 2)



$T_k(xy)$ = Pr(random chr from gen k has x at locus 1 and random chr from other individual has y at locus 2)

$$C_{k+1} = (1 - r)C_k + rS_k$$

$$S_{k+1} = T_k$$

$$T_{k+1} = \frac{1}{4}C_k + \frac{1}{4}S_k + \frac{1}{2}T_k$$

Kimura, Biometrics 19:1–17, 1963

2-locus hap, autosome

$$\Pr(AA) = \frac{1}{4}\left\{\frac{1}{1+6r} + \left[\frac{1-2r-z}{4}\right]^k \left[\frac{6r^2-7r+3rz}{(1+6r)z}\right] - \left[\frac{1-2r+z}{4}\right]^k \left[\frac{6r^2-7r-3rz}{(1+6r)z}\right]\right\}$$

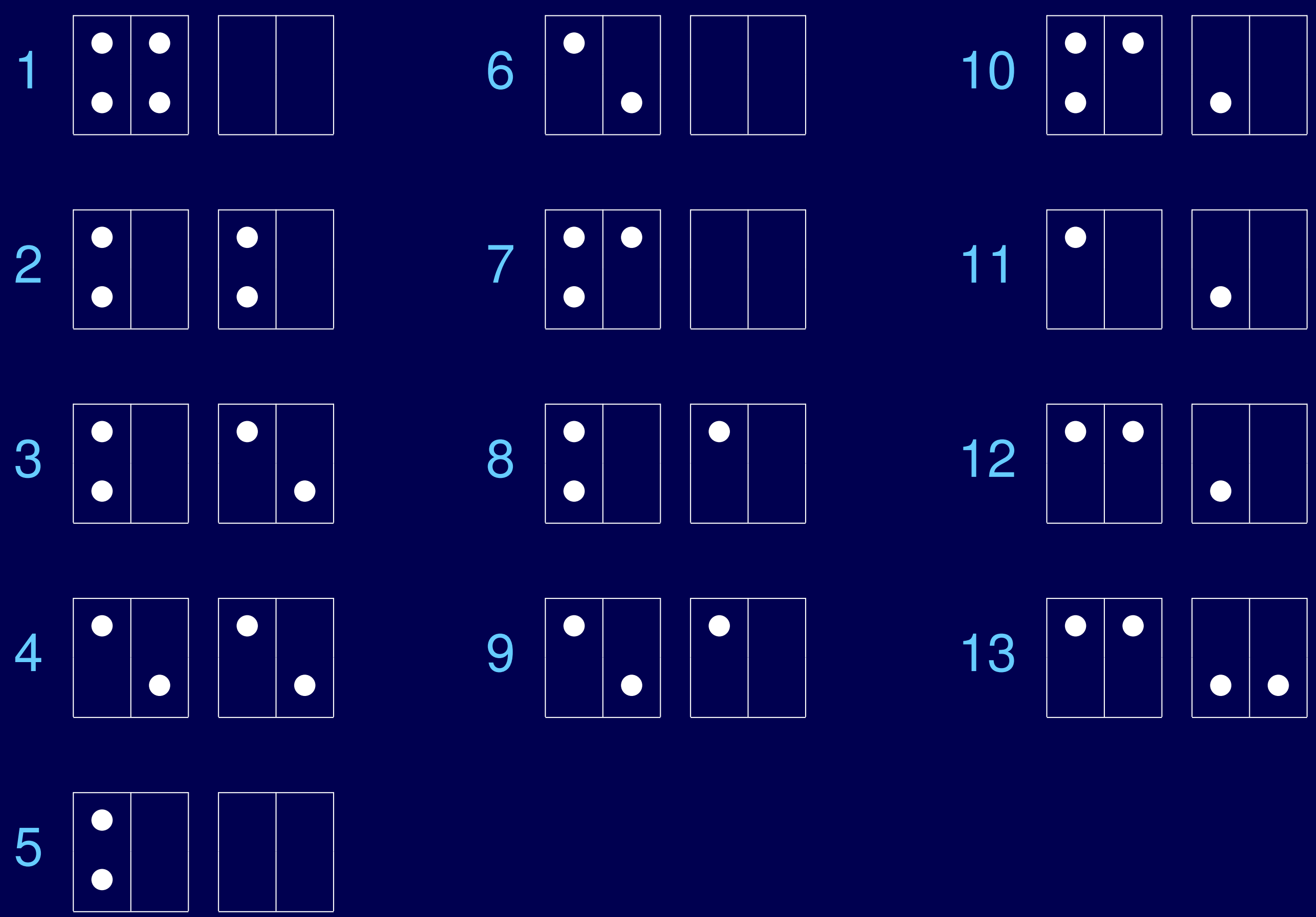
$$\Pr(AB) = \frac{1}{4}\left\{\frac{2r}{1+6r} - \left[\frac{1-2r-z}{4}\right]^k \left[\frac{10r^2-r+rz}{(1+6r)z}\right] + \left[\frac{1-2r+z}{4}\right]^k \left[\frac{10r^2-r-rz}{(1+6r)z}\right]\right\}$$

$$\Pr(AC) = \frac{1}{8}\left\{\frac{4r}{1+6r} + \left[\frac{1-2r-z}{4}\right]^k \left[\frac{4r^2+6r-2rz}{(1+6r)z}\right] - \left[\frac{1-2r+z}{4}\right]^k \left[\frac{4r^2+6r+2rz}{(1+6r)z}\right]\right\}$$

where $z = \sqrt{(1-2r)(5-2r)}$

2-locus gen, autosome

We seek to extend the idea from Kimura (1963) to get probabilities at G_2F_k for patterns like $\begin{array}{c|c} A & A \\ A & A \end{array}$, $\begin{array}{c|c} A & A \\ B & B \end{array}$, $\begin{array}{c|c} A & A \\ C & C \end{array}$



2-locus gen, autosome

The “transition” matrix (really, the matrix that defines the recursion):

	1	2	3	4	5	6	7	8	9	10	11	12	13
1		$(1-r)^2$	$2r(1-r)$	r^2									
2	$\frac{r^2+(1-r)^2}{4}$	$\frac{(1-r)^2}{2}$	$r(1-r)$	$\frac{r^2}{2}$	$\frac{(1-r)^2}{4}$	$\frac{r^2}{4}$	$r(1-r)$						
3								$\frac{1-r}{2}$	$\frac{r}{2}$	$\frac{1}{2}$			
4		$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$							$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$
5					$(1-r)$	r							
6											1		
7								$(1-r)$	r				
8					$\frac{1-r}{4}$	$\frac{r}{4}$	$\frac{1}{4}$	$\frac{1-r}{2}$	$\frac{r}{2}$				
9								$\frac{1}{4}$	$\frac{1}{4}$		$\frac{1}{4}$	$\frac{1}{4}$	
10		$\frac{1-r}{4}$	$\frac{1}{4}$	$\frac{r}{4}$				$\frac{1-r}{4}$	$\frac{r}{4}$	$\frac{1}{4}$			
11					$\frac{1}{4}$	$\frac{1}{4}$					$\frac{1}{2}$		
12								$\frac{1}{2}$	$\frac{1}{2}$				
13		$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$									

What are we doing?

We have a Markov chain with many states (say n)

Transition matrix P; starting distribution π_0

We want to calculate $\pi_k = \pi_0 P^k$

Really, we just need $\pi_k b = \pi_0 P^k b$ for some $n \times 1$ vector b

Suppose we can expand b to an $n \times m$ matrix B in a way that $PB = BA$ for some $m \times m$ matrix A

Then $P^k B = P^{k-1}(PB) = P^{k-1}(BA) = \cdots = BA^k$

And so $\pi_0 P^k B = \pi_0 B A^k$

$\longrightarrow$ work with A $[m \times m]$ in place of P $[n \times n]$

Counting states

Two loci, X chromosome

3 alleles: $3^6 = 729$ types

Swap $A \leftrightarrow B$, two hapl in female, 2 loci $\rightarrow$ 116 types

To get at $\Pr(AA|AA)$ in females, consider just 2 alleles ($2^6 = 64$ types)

After symmetries: 24 types

Two loci, autosome

4 alleles: $4^8 = 65,536$ types

After symmetries: 700 types

To get at $\Pr(AA|AA)$, consider just 2 alleles ($2^8 = 256$ types)

After symmetries: 34 types

2-locus gen, autosome

Returning to that 13×13 matrix, we need the eigenvalues and eigenvectors.

The first 7 eigenvalues:

$$1$$

$$\frac{1\pm\sqrt{5}}{4}$$

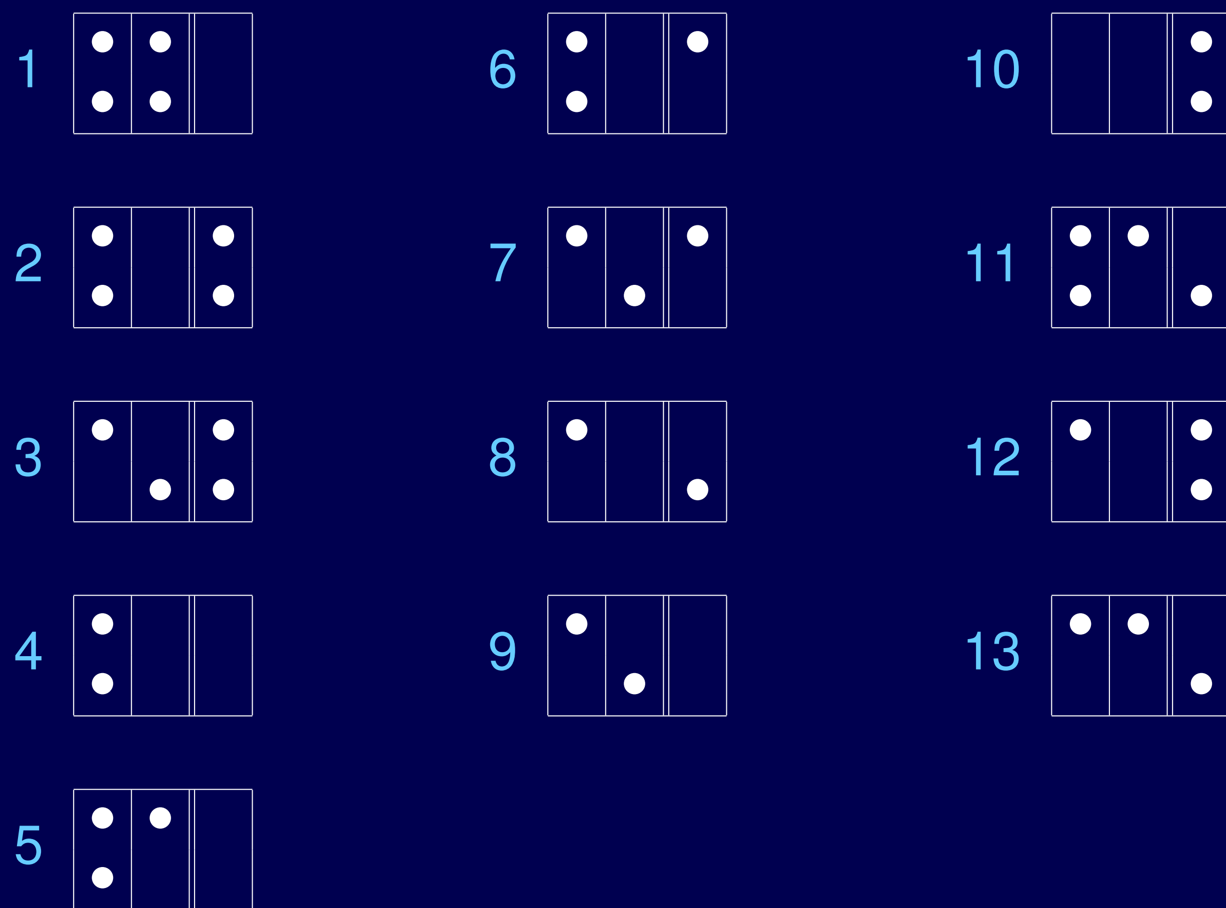
$$\frac{1-2r\pm\sqrt{4r^2-12r+5}}{4}$$

$$\frac{1-2r\pm\sqrt{(1-2r)(9-2r)}}{8}$$

The other 6 eigenvalues are the roots of

$$\begin{aligned}1026x^6 \;&-\; 128(4r^2-8r+7)x^5 - 64(8r^4-24r^3+26r^2-10r+3)x^4\\ &+\; 8(16r^4-80r^3+120r^2-84r+23)x^3 - 8(16r^5-56r^4+72r^3-48r^2+18r-3)x^2\\ &+\; 4(16r^5-40r^4+50r^3-35r^2+13r-2)x + (16r^5-40r^4+44r^3-26r^2+8r-1)\end{aligned}$$

2-locus gen, X chr



2-locus gen, X chr

The first 6 eigenvalues:

$$1, \; -\tfrac{1}{2}$$

$$\tfrac{1 \pm \sqrt{5}}{4}$$

$$\tfrac{1 - r \pm \sqrt{r^2 - 10r + 5}}{4}$$

The other 7 eigenvalues are the roots of

$$16x^3 + 4x^2 - (6r + 4)x + 2r - 1$$

and

$$\begin{aligned} 32x^4 \; &+ \; (16r - 24)x^3 + (16r^3 - 32r^2 + 20r - 8)x^2 \\ &- \; (4r^3 - 8r^2 + 10r - 4)x - (4r^3 - 6r^2 + 4r - 1) \end{aligned}$$

Summary

		1 locus	2 loc, hap	2 loc, gen
X chr	Male	✓	✓	—
	Female	✓	✓	×
	Pair	✓	×	×
Autosome	Ind'l	✓	✓	×
	Pair	✓	×	×

Acknowledgments

Timo Seppäläinen Mathematics, UW–Madison

Bret Larget Statistics & Botany, UW–Madison