

Two-point Linkage Analysis

Jiankang Wang

E-mail: jkwang@cgiar.org; wangjiankang@caas.cn

Web: <http://www.isbreeding.net>

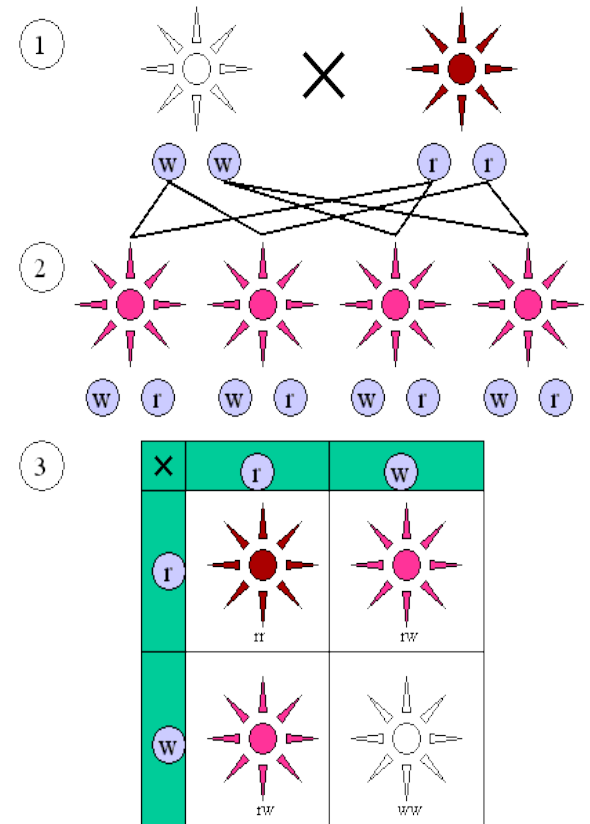
Outlines of the presentation

- Basic Genetics Principles (or Laws)
- Estimation of Recombination Frequency in DH and RIL populations
- Estimation of Recombination Frequency in F_2 populations
- Generation Matrix and the Expected Frequency
- Several Rounds of Random Mating before Repeated Selfing

Basic Genetics Principles (or Laws)

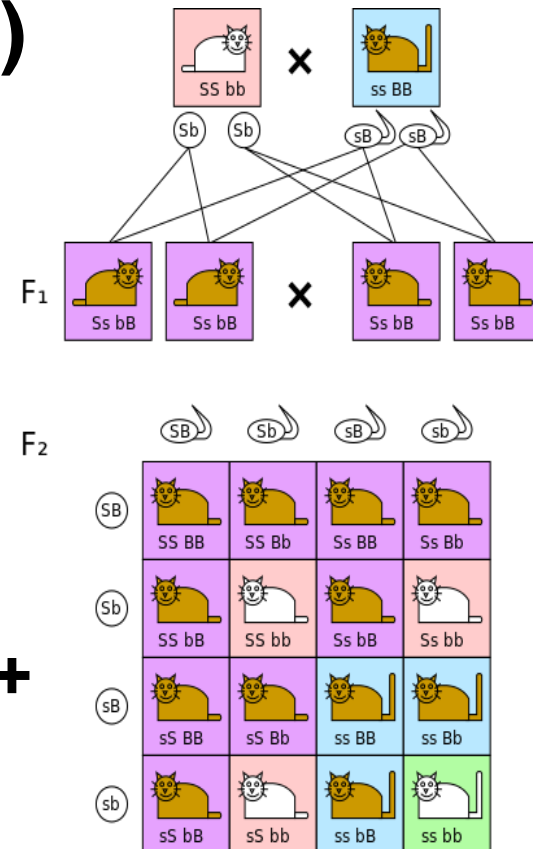
The First Mendelian Law

- The Principle of Segregation (The “First Law”).
- For genotype Aa:
 - From zygote to gamete:
 $\frac{1}{2}A + \frac{1}{2}a$ (1:1)
 - From gamete to zygote:
 $(\frac{1}{2}A + \frac{1}{2}a)^2 = \frac{1}{4}AA + \frac{1}{2}Aa + \frac{1}{4}aa$ (1:2:1)

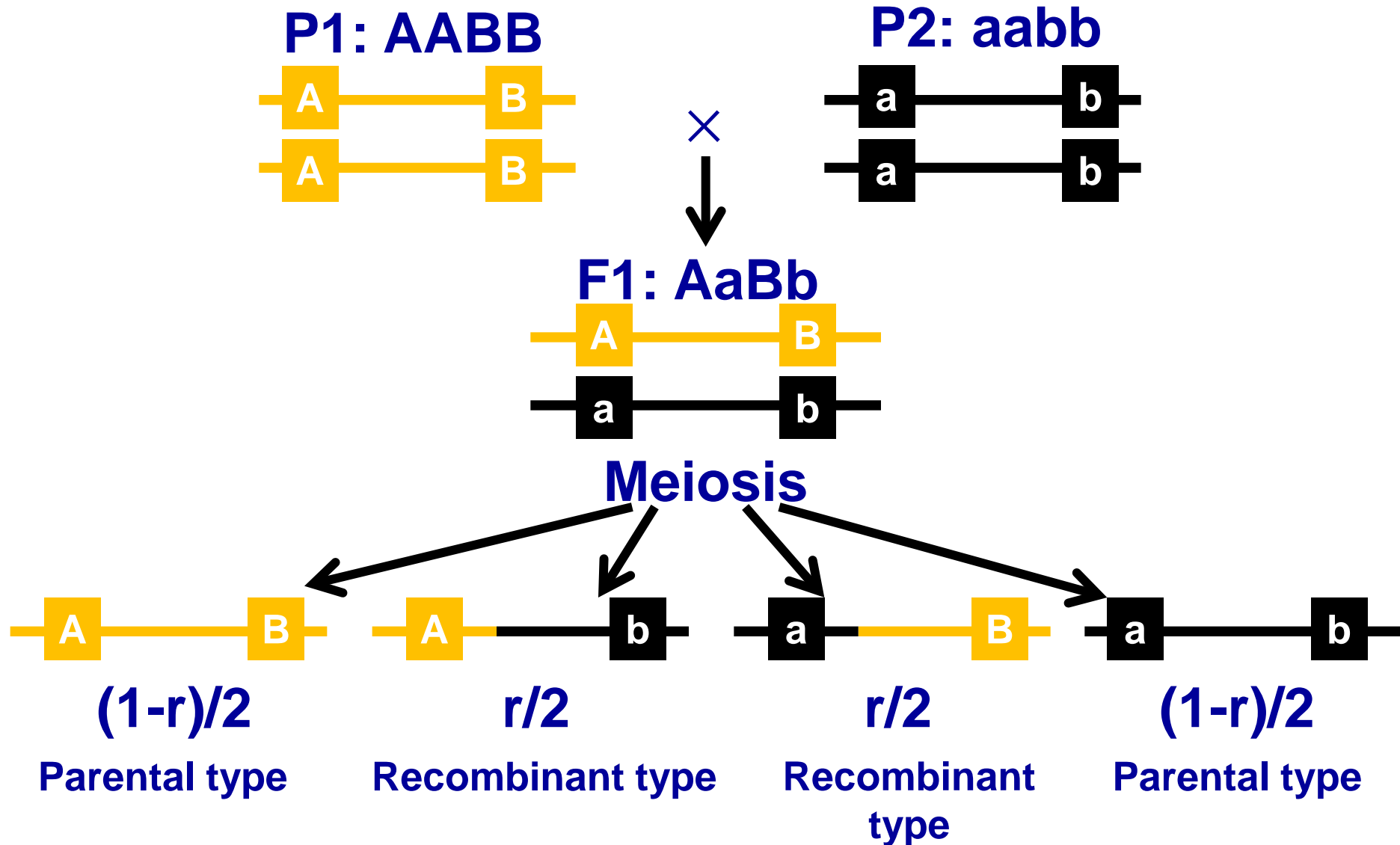


The Second Mendelian Law

- For two independent loci, i.e. no linkage
- The Principle of Independent Assortment (The “Second Law”)
- For genotype AaBb:
 - From zygote to gamete: $\frac{1}{4}AB + \frac{1}{4}Ab + \frac{1}{4}aB + \frac{1}{4}ab$ (1:1:1:1)
 - From gamete to zygote: $(\frac{1}{4}AB + \frac{1}{4}Ab + \frac{1}{4}aB + \frac{1}{4}ab)^2 = _AABB + _AABb + _AAbb + _AaBB + _AaBb + _Aabb + _aaBB + _aaBb + _aabb$ (1:2:1:2:4:2:1:2:1)

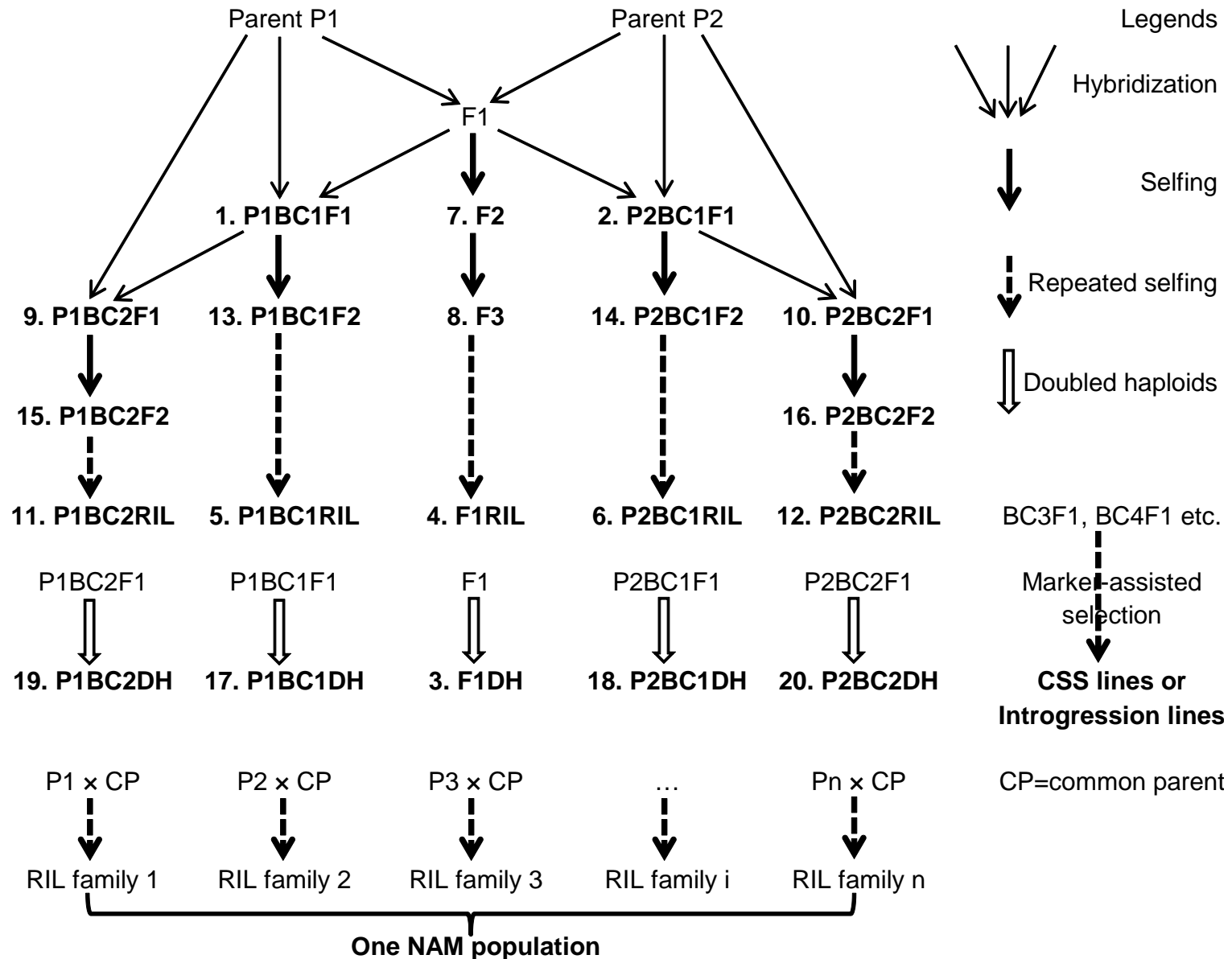


The Third Genetics Law: Linkage and Recombination



Estimation of Recombination Frequency in DH and RIL populations

Populations handled in QTL IciMapping



Example: 10 RILs in a rice population (Linkage map of Chr. 5)

Marker	C263	R830	R3166	XNpb387	R569	R1553	C128	C1402	XNpb81	C246	R2953	C1447	Grain width (mm)
Position (cM)	0.0	3.5	8.5	19.5	32.0	66.6	74.1	78.6	81.8	91.9	92.7	96.8	
RIL1	0	0	0	0	0	0	0	0	0	0	0	0	2.33
RIL2	2	2	2	2	2	0	0	0	0	2	2	2	1.99
RIL3	0	2	2	2	2	2	2	2	2	2	2	2	2.24
RIL4	0	0	0	0	0	0	2	2	2	2	2	2	1.94
RIL5	0	0	0	0	0	2	2	0	0	0	0	0	2.76
RIL6	0	0	0	2	2	2	2	2	2	2	2	2	2.32
RIL7	0	0	0	0	0	0	0	0	0	0	0	0	2.32
RIL8	2	2	0	2	2	0	0	0	0	2	2	2	2.08
RIL9	0	0	0	0	2	2	0	0	0	0	0	0	2.24
RIL10	0	0	0	0	2	2	0	0	0	0	0	0	2.45

Genetic markers in linkage analysis

- **Morphological traits**
 - **Qualitative traits used in Mendel's hybridization experiments**
- **Cytogenetic and bio-chemistry markers (e.g. isozyme)**
- **DNA molecular markers**
 - **RFLP, SSR, SNP etc.**

Expected genotypic frequency in backcross and DH populations

P1: AABB; P2: aabb

P1BC1	P2BC1	DH	Samples	Theoretical frequency
AB/AB	AB/ab	AB/AB	n_1	$f_1=(1-r)/2$
Ab/AB	Ab/ab	Ab/Ab	n_2	$f_2=r/2$
aB/AB	aB/ab	aB/aB	n_3	$f_3=r/2$
ab/AB	ab/ab	ab/ab	n_4	$f_4=(1-r)/2$

AB, **Ab**, **aB**, and **ab** are the 4 haplotypes of F1 hybrid

MLE of recombination frequency

- **Likelihood function**

$$L = \frac{n!}{n_1!n_2!n_3!n_4!} \left[\frac{1}{2}(1-r) \right]^{n_1} \left[\frac{1}{2}r \right]^{n_2} \left[\frac{1}{2}r \right]^{n_3} \left[\frac{1}{2}(1-r) \right]^{n_4} = C(1-r)^{n_1+n_4} (r)^{n_2+n_3}$$

- **Logarithm of likelihood**

$$\ln L = \ln C + (n_1 + n_4) \ln(1-r) + (n_2 + n_3) \ln r$$

- **MLE of r** $\hat{r} = \frac{n_2 + n_3}{n_1 + n_2 + n_3 + n_4} = \frac{n_2 + n_3}{n}$

- **Fisher information**

$$I = -E\left(\frac{d^2 \ln L}{d^2 r}\right) = -E\left[-\frac{n_1 + n_4}{(1-r)^2} - \frac{n_2 + n_3}{r^2}\right] = \frac{n}{r(1-r)}$$

- **Variance of estimated r** $V_{\hat{r}} = \frac{1}{I} = \frac{\hat{r}(1-\hat{r})}{n}$

Significance test of linkage

- Null hypothesis $H_0: r = 0.5$ (no genetic linkage, or locus A-a and B-b are independent)
- Alternative hypothesis $H_A: r \neq 0.5$
- Likelihood ratio test (LRT) or LOD score

$$LRT = -2 \ln \left[\frac{L(r = 0.5)}{L(\hat{r})} \right] \sim \chi^2 (df = 1)$$

$$LOD = \log \frac{L(\hat{r})}{L(r = 0.5)}$$

An example P1BC1 population

- **Genotypes of two inbred parents P1 and P2 are AABB and aabb**
- **Observed samples of the four genotypes in P1BC1**
 - **AABB: 162; AABb: 40; AaBB: 41; AaBb: 158**

$$\hat{r} = \frac{40 + 41}{162 + 40 + 41 + 158} = \frac{81}{401} = 20.20\%$$

$$V_{\hat{r}} = \frac{\hat{r}(1 - \hat{r})}{n} \approx 4.02 \times 10^{-4}$$

Test of linkage

- Null hypothesis $H_0: r = 0.5$
- Alternative hypothesis $H_A: r \neq 0.5$
- Likelihood ratio test (LRT) ($P < 0.0001$) and LOD score

$$LRT = -2 \ln \left[\frac{L(r = 0.5)}{L(\hat{r})} \right] = 152.37$$

$$LOD = \log \left[\frac{L(\hat{r})}{L(r = 0.5)} \right] = 33.09$$

Genotypic frequencies in RIL populations, compared with DH

DH population	Theoretical frequency	RIL population	Theoretical frequency
AABB	$f_1=(1-r)/2$	AABB	$f_1=(1-R)/2$
AAbb	$f_2=r/2$	AAbb	$f_2=R/2$
aaBB	$f_3=r/2$	aaBB	$f_3=R/2$
aabb	$f_4=(1-r)/2$	aabb	$f_4=(1-R)/2$

$$R = 2r / (1 + 2r)$$

10 RILs in a rice population

P1: 0 or A; P2: 2 or B; F1: 1 or H

RIL	Marker 1	Marker 2	Parent type or recombinant
	C263	XNpb387	
RIL1	0 or A	0 or A	P1 type
RIL2	2 or B	2 or B	P2 type
RIL3	0 or A	2 or B	Recombinant
RIL4	0 or A	0 or A	P1 type
RIL5	0 or A	0 or A	P1 type
RIL6	0 or A	2 or B	Recombinant
RIL7	0 or A	0 or A	P1 type
RIL8	2 or B	2 or B	P2 type
RIL9	0 or A	0 or A	P1 type
RIL10	0 or A	0 or A	P1 type

$$n_1 = 6$$

$$n_2 = 2$$

$$n_3 = 0$$

$$n_4 = 2$$

$$R = 2/10 = 0.2$$

$$r = 0.125$$

$$LRT = 3.8549, \\ (P = 0.0496)$$

$$LOD = 0.8371$$

Estimation of Recombination Frequency in F_2 populations

Expected genotypic frequencies in F2 populations

Co-dominant markers		Dominant markers	
Marker type	Frequency	Marker type	Frequency
AABB	$(1-r)^2/4$	A_B_	$[2+(1-r)^2]/4$
AABb	$r(1-r)/2$		
AAbb	$r^2/4$	A_bb	$[1-(1-r)^2]/4$
AaBB	$r(1-r)/2$		
AaBb	$(1-2r+2r^2)/2$		
Aabb	$r(1-r)/2$		
aaBB	$r^2/4$	aaB_	$[1-(1-r)^2]/4$
aaBb	$r(1-r)/2$		
aabb	$(1-r)^2/4$	aabb	$(1-r)^2/4$

Co-dominant markers in other populations

Marker type	Population						
	F2	P1B1F1	P2B1F1	F1DH	P1BC1DH	P2BC1DH	F1-RIL
AABB	$(1-r)^2/4$	$(1-r)/2$		$(1-r)/2$	$\frac{1}{2}+(1-r)^2/4$	$(1-r)^2/4$	$(1-R)/2$
AABb	$r(1-r)/2$	$r/2$					
AAbb	$r^2/4$			$r/2$	$r/2-r^2/4$	$r/2-r^2/4$	$R/2$
AaBB	$r(1-r)/2$	$r/2$					
AaBb	$(1-2r+2r^2)/2$	$(1-r)/2$	$(1-r)/2$				
Aabb	$r(1-r)/2$		$r/2$				
aaBB	$r^2/4$			$r/2$	$r/2-r^2/4$	$r/2-r^2/4$	$R/2$
aaBb	$r(1-r)/2$		$r/2$				
aabb	$(1-r)^2/4$		$(1-r)/2$	$(1-r)/2$	$(1-r)^2/4$	$\frac{1}{2}+(1-r)^2/4$	$(1-R)/2$

$$R = 2r / (1 + 2r)$$

MLE of r in F2: dominant markers

- **Logarithm of the likelihood ratio** $k = (1 - r)^2$

$$\begin{aligned}\ln L &= C + n_1 \ln(3 - 2r + r^2) + (n_3 + n_7) \ln(2r - r^2) + n_9 \ln(1 - 2r + r^2) \\ &= C + n_1 \ln(2 + k) + (n_3 + n_7) \ln(1 - k) + n_9 \ln k\end{aligned}$$

- **MLE of r**

$$k = (1 - r)^2 = \frac{-(2n - 3n_1 - n_9) \pm \sqrt{(2n - 3n_1 - n_9)^2 + n \times n_9}}{2n}$$

- **Variance of the estimated r**

$$V_{\hat{r}} = \frac{(1 - k)(2 - k)}{2n(1 + 2k)} = \frac{(2r - r^2)(3 - 2r + r^2)}{2n(3 - 4r + 2r^2)}$$

MLE of r in F2: co-dominant markers (Newton-Raphson algorithm)

- **Log-likelihood function**

$$\ln L = \ln C + (2n_1 + 2n_9 + n_2 + n_4 + n_6 + n_8) \ln(1-r) \\ + (n_2 + n_4 + n_6 + n_8 + 2n_3 + 2n_7) \ln r + n_5 \ln(1-2r+2r^2)$$

- **The first-order derivative of LogL**

- $f'(r) = \frac{d \ln L}{dr} = \frac{2n_1 + 2n_9 + n_2 + n_4 + n_6 + n_8}{r-1} + \frac{n_2 + n_4 + n_6 + n_8 + 2n_3 + 2n_7}{r} + \frac{n_5(4r-2)}{1-2r+2r^2}$

- **The second-order derivative of LogL**

- $f''(r) = \frac{d^2 \ln L}{d^2 r} = -\frac{2n_1 + 2n_9 + n_2 + n_4 + n_6 + n_8}{(r-1)^2} - \frac{n_2 + n_4 + n_6 + n_8 + 2n_3 + 2n_7}{r^2} + \frac{n_5(4r-4r^2)}{(1-2r+2r^2)^2}$

- **The iteration algorithm:**

$$r_{i+1} = r_i - f'(r_i)/f''(r_i)$$

MLE of r in F2: co-dominant markers (EM algorithm)

- EM for expectation and maximization
- E-step: for an initial r_0 , calculate the probability of crossover in each marker type
- M-step: Update r , and repeat from the E-step

$$r' = \frac{1}{n} \sum_k n_k P_k(R | G)$$

Expected probability of crossover

Marker type	Frequency	Expected sample size	P(R G)
AABB	$f_1 = (1-r)^2/4$	$n_1 = nf_1$	0
AABb	$f_2 = r(1-r)/2$	$n_2 = nf_2$	0.5
AAbb	$f_3 = r^2/4$	$n_3 = nf_3$	1
AaBB	$f_4 = r(1-r)/2$	$n_4 = nf_4$	0.5
AaBb	$f_5 = (1-2r+2r^2)/2$	$n_5 = nf_5$	$r^2/(1-2r+2r^2)$
Aabb	$f_6 = r(1-r)/2$	$n_6 = nf_6$	0.5
aaBB	$f_7 = r^2/4$	$n_7 = nf_7$	1
aaBb	$f_8 = r(1-r)/2$	$n_8 = nf_8$	0.5
aabb	$f_9 = (1-r)^2/4$	$n_9 = nf_9$	0

$$r = [n_1 \times 0 + n_2 \times 0.5 + n_3 \times 1 + \dots + n_8 \times 0.5 + n_9 \times 0] / n$$

Estimated r after 3 EM iterations ($r_0=0.5$)

Geno.	Size	r_0	Exp. Freq.	P(R G)	r_1	Exp. Freq.	P(R G)	r_2	Exp. Freq.	P(R G)	r_3
AABB	30	0.5	0.063	0	0.313	0.118	0	0.198	0.161	0	0.159
AABb	7	0.5	0.125	0.5	0.313	0.107	0.5	0.198	0.080	0.5	0.159
AAbb	1	0.5	0.063	1	0.313	0.024	1	0.198	0.010	1	0.159
AaBB	9	0.5	0.125	0.5	0.313	0.107	0.5	0.198	0.080	0.5	0.159
AaBb	50	0.5	0.250	0.5	0.313	0.285	0.1712	0.198	0.341	0.0577	0.159
Aabb	12	0.5	0.125	0.5	0.313	0.107	0.5	0.198	0.080	0.5	0.159
aaBB	0	0.5	0.063	1	0.313	0.024	1	0.198	0.010	1	0.159
aaBb	10	0.5	0.125	0.5	0.313	0.107	0.5	0.198	0.080	0.5	0.159
aabb	25	0.5	0.063	0	0.313	0.118	0	0.198	0.161	0	0.159
	144		1			1			1		

Estimated r after 3 EM iterations ($r_0=0.25$)

Geno.	Size	r_0	Exp. Freq.	P(R G)	r_1	Exp. Freq.	P(R G)	r_2	Exp. Freq.	P(R G)	r_3
AABB	30	0.25	0.141	0	0.174	0.171	0	0.154	0.179	0	0.150
AABb	7	0.25	0.094	0.5	0.174	0.072	0.5	0.154	0.065	0.5	0.150
AAbb	1	0.25	0.016	1	0.174	0.008	1	0.154	0.006	1	0.150
AaBB	9	0.25	0.094	0.5	0.174	0.072	0.5	0.154	0.065	0.5	0.150
AaBb	50	0.25	0.313	0.1	0.174	0.357	0.0423	0.154	0.370	0.0319	0.150
Aabb	12	0.25	0.094	0.5	0.174	0.072	0.5	0.154	0.065	0.5	0.150
aaBB	0	0.25	0.016	1	0.174	0.008	1	0.154	0.006	1	0.150
aaBb	10	0.25	0.094	0.5	0.174	0.072	0.5	0.154	0.065	0.5	0.150
aabb	25	0.25	0.141	0	0.174	0.171	0	0.154	0.179	0	0.150
	144			1			1			1	

Estimated r after 3 EM iterations ($r_0=0.0$)

Geno.	Size	r_0	Exp. Freq.	P(R G)	r_1	Exp. Freq.	P(R G)	r_2	Exp. Freq.	P(R G)	r_3
AABB	30	0	0.250	0	0.139	0.185	0	0.148	0.182	0	0.149
AABb	7	0	0.000	0.5	0.139	0.060	0.5	0.148	0.063	0.5	0.149
AAbb	1	0	0.000	1	0.139	0.005	1	0.148	0.005	1	0.149
AaBB	9	0	0.000	0.5	0.139	0.060	0.5	0.148	0.063	0.5	0.149
AaBb	50	0	0.500	0	0.139	0.380	0.0253	0.148	0.374	0.0292	0.149
Aabb	12	0	0.000	0.5	0.139	0.060	0.5	0.148	0.063	0.5	0.149
aaBB	0	0	0.000	1	0.139	0.005	1	0.148	0.005	1	0.149
aaBb	10	0	0.000	0.5	0.139	0.060	0.5	0.148	0.063	0.5	0.149
aabb	25	0	0.250	0	0.139	0.185	0	0.148	0.182	0	0.149
	144		1			1			1		

MLE of r in F2: between one co-dominant marker and one dominant marker

- **Likelihood function**

$$L(r) = C(1 - r^2)^{n_1} (r^2)^{n_2} (1 - r + r^2)^{n_3} [r(1 - r)]^{n_4} [r(2 - r)]^{n_5} (1 - r)^{2n_6}$$

- **Log-likelihood function**

$$L(r) = C(1 - r^2)^{n_1} (r^2)^{n_2} (1 - r + r^2)^{n_3} [r(1 - r)]^{n_4} [r(2 - r)]^{n_5} (1 - r)^{2n_6} \\ + (n_1 + n_4 + 2n_6) \ln(1 - r) + n_5 \ln(2 - r) + n_3 \ln(1 - r + r^2)$$

- **The first-order derivative of LogL**

$$\ln L'(r) = \frac{d \ln L(r)}{dr} = \frac{2n_2 + n_4 + n_5}{r} + \frac{n_1}{1+r} - \frac{n_1 + n_4 + 2n_6}{1-r} - \frac{n_5}{2-r} - \frac{n_3(1-2r)}{1-r+r^2}$$

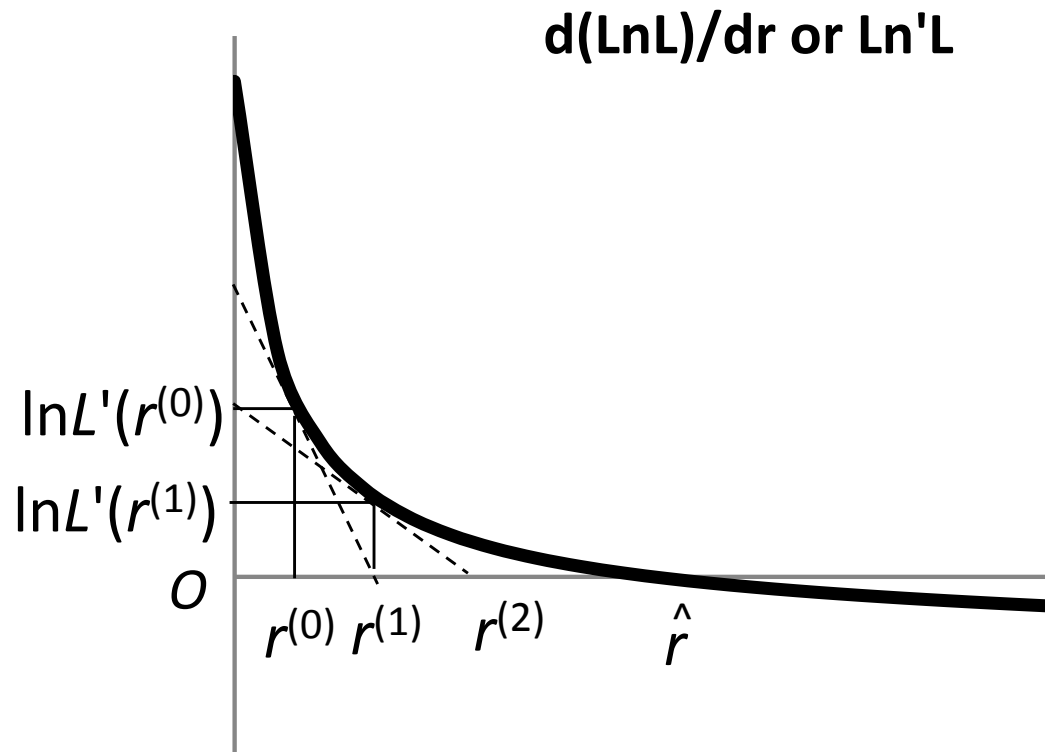
- **The second-order derivative of LogL**

$$\ln L''(r) = \frac{d^2 \ln L(r)}{dr^2} = -\frac{2n_2 + n_4 + n_5}{r^2} - \frac{n_1}{(1+r)^2} + \frac{n_1 + n_4 + 2n_6}{(1-r)^2} + \frac{n_5}{(2-r)^2} + \frac{n_3(1+2r-2r^2)}{(1-r+r^2)^2}$$

- **The iteration algorithm: $r_{i+1} = r_i - f'(r_i)/f''(r_i)$**

Principle of the Newton-Raphson algorithm

- To find the maxima of $\ln L$ is equal to find the solution in equation $\ln L' = 0$



An example in wheat

	Resistant F2			Susceptible F2		
Marker type	A	H	B	A	H	B
Size	572	1161	14	3	22	569

Iteration	1	2	3	4	5
Recom. Freq. r	0.0010	0.0019	0.0037	0.0066	0.0108
LnL	-282.75	-257.02	-234.28	-216.51	-205.69
d(LnL)/dr	39670	19268	9081.59	4018.63	1548.44
d ² (LnL)/dr ²	-4.12E7	-1.11E7	-3.10E6	-9.5E5	-3.58E5
Iteration	5	6	7	8	9
Recom. Freq. r	0.0108	0.0151	0.0175	0.0179	0.0179
LnL	-205.69	-201.67	-201.12	-201.11	-201.11
d(LnL)/dr	1548.44	430.81	50.88	-0.26	0.0071
d ² (LnL)/dr ²	-3.58E5	-1.81E5	-1.35E5	-1.29E5	-1.29E5

Distortion has little effect on linkage analysis!

DH pop	Theo. Freq.	Distortion	Freq. in distortion
AABB	$f_1=(1-r)/2$	$(1-r)/2$	$(1-r)/(1+s)$
AAbb	$f_2=r/2$	$r/2$	$r/(1+s)$
aaBB	$f_3=r/2$	$s \times r/2$	$r \times s/(1+s)$
aabb	$f_4=(1-r)/2$	$s \times (1-r)/2$	$(1-r) \times s/(1+s)$
Sum	1	$(1+s)/2$	1

$$\hat{r} = r/(1+s) + r \times s/(1+s) = r(1+s)/(1+s) = r$$

Generation Matrix and the Expected Frequency

Expected genotype frequencies calculated from the transmission matrix (T)

$$\mathbf{f}^{(t)} = \left[f_{AABB}^{(t)} \quad f_{AABb}^{(t)} \quad f_{AAbb}^{(t)} \quad f_{AaBB}^{(t)} \quad f_{AB/ab}^{(t)} \quad f_{Ab/aB}^{(t)} \quad f_{Aabb}^{(t)} \quad f_{aaBB}^{(t)} \quad f_{aaBb}^{(t)} \quad f_{aabb}^{(t)} \right]$$

$$\mathbf{f}^{(t+1)} = \left[f_{AABB}^{(t+1)} \quad f_{AABb}^{(t+1)} \quad f_{AAbb}^{(t+1)} \quad f_{AaBB}^{(t+1)} \quad f_{AB/ab}^{(t+1)} \quad f_{Ab/aB}^{(t+1)} \quad f_{Aabb}^{(t+1)} \quad f_{aaBB}^{(t+1)} \quad f_{aaBb}^{(t+1)} \quad f_{aabb}^{(t+1)} \right]$$

$$\mathbf{f}^{(t+1)} = \mathbf{f}^{(t)} \mathbf{T}$$

The generation matrix (T) for the backcrossing to P1

$$\mathbf{f}^{(t)} = \begin{bmatrix} f_{AABB}^{(t)} & f_{AABb}^{(t)} & f_{AAbb}^{(t)} & f_{AaBB}^{(t)} & f_{AB/ab}^{(t)} & f_{Ab/aB}^{(t)} & f_{Aabb}^{(t)} & f_{aaBB}^{(t)} & f_{aaBb}^{(t)} & f_{aabb}^{(t)} \end{bmatrix}$$

$$\mathbf{T}_{P1B} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2}(1-r) & \frac{1}{2}r & 0 & \frac{1}{2}r & \frac{1}{2}(1-r) & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2}r & \frac{1}{2}(1-r) & 0 & \frac{1}{2}(1-r) & \frac{1}{2}r & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The generation matrix (T) for the backcrossing to P2

$$\mathbf{f}^{(t)} = \begin{bmatrix} f_{AABB}^{(t)} & f_{AABb}^{(t)} & f_{AAbb}^{(t)} & f_{AaBB}^{(t)} & f_{AB/ab}^{(t)} & f_{Ab/aB}^{(t)} & f_{Aabb}^{(t)} & f_{aaBB}^{(t)} & f_{aaBb}^{(t)} & f_{aabb}^{(t)} \end{bmatrix}$$

$$\mathbf{T}_{P2B} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2}(1-r) & 0 & \frac{1}{2}r & 0 & \frac{1}{2}r & \frac{1}{2}(1-r) \\ 0 & 0 & 0 & 0 & \frac{1}{2}r & 0 & \frac{1}{2}(1-r) & 0 & \frac{1}{2}(1-r) & \frac{1}{2}r \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

The generation matrix (T) for the selfing pollination

$$\mathbf{f}^{(t)} = \begin{bmatrix} f_{AABB}^{(t)} & f_{AABb}^{(t)} & f_{AAbb}^{(t)} & f_{AaBB}^{(t)} & f_{AB/ab}^{(t)} & f_{Ab/aB}^{(t)} & f_{Aabb}^{(t)} & f_{aaBB}^{(t)} & f_{aaBB}^{(t)} & f_{aaBB}^{(t)} \end{bmatrix}$$

$$\mathbf{T}_s = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{4} & 0 & 0 \\ \frac{1}{4}(1-r)^2 & \frac{1}{2}r(1-r) & \frac{1}{4}r^2 & \frac{1}{2}r(1-r) & \frac{1}{2}(1-r)^2 & \frac{1}{2}r^2 & \frac{1}{2}r(1-r) & \frac{1}{4}r^2 & \frac{1}{2}r(1-r) & \frac{1}{4}(1-r)^2 \\ \frac{1}{4}r^2 & \frac{1}{2}r(1-r) & \frac{1}{4}(1-r)^2 & \frac{1}{2}r(1-r) & \frac{1}{2}r^2 & \frac{1}{2}(1-r)^2 & \frac{1}{2}r(1-r) & \frac{1}{4}(1-r)^2 & \frac{1}{2}r(1-r) & \frac{1}{2}r^2 \\ 0 & 0 & \frac{1}{4} & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & \frac{1}{4} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

The generation matrix (T) for the doubled haploids

$$\mathbf{f}^{(t)} = \left[\begin{array}{cccccccccc} f_{AABB}^{(t)} & f_{AABb}^{(t)} & f_{AAbb}^{(t)} & f_{AaBB}^{(t)} & f_{AB/ab}^{(t)} & f_{Ab/aB}^{(t)} & f_{Aabb}^{(t)} & f_{aaBB}^{(t)} & f_{aaBB}^{(t)} & f_{aaBB}^{(t)} \end{array} \right]$$

$$\mathbf{T}_D = \left[\begin{array}{cccccccccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 \\ \frac{1}{2}(1-r) & 0 & \frac{1}{2}r & 0 & 0 & 0 & 0 & 0 & \frac{1}{2}r & 0 & \frac{1}{2}(1-r) \\ \frac{1}{2}r & 0 & \frac{1}{2}(1-r) & 0 & 0 & 0 & 0 & 0 & \frac{1}{2}(1-r) & 0 & \frac{1}{2}r \\ 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array} \right]$$

The generation matrix (T) for the repeated selfing pollination

$$\mathbf{f}^{(t)} = \begin{bmatrix} f_{AABB}^{(t)} & f_{AABb}^{(t)} & f_{AAbb}^{(t)} & f_{AaBB}^{(t)} & f_{AB/ab}^{(t)} & f_{Ab/aB}^{(t)} & f_{Aabb}^{(t)} & f_{aaBB}^{(t)} & f_{aabb}^{(t)} & f_{aabb}^{(t)} \end{bmatrix}$$

$$\mathbf{T}_R = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 \\ \frac{1}{2}(1-R) & 0 & \frac{1}{2}R & 0 & 0 & 0 & 0 & \frac{1}{2}R & 0 & \frac{1}{2}(1-R) \\ \frac{1}{2}R & 0 & \frac{1}{2}(1-R) & 0 & 0 & 0 & 0 & \frac{1}{2}(1-R) & 0 & \frac{1}{2}R \\ 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{f}^{(0)} = [0 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0]$$

Theoretical frequencies of the 20 biparental populations

No.	Population	Theoretical frequencies
1	P1BC1F1	$\mathbf{f}^{(0)} \times \mathbf{T}_{P1B}$
2	P2BC1F1	$\mathbf{f}^{(0)} \times \mathbf{T}_{P2B}$
3	F1DH	$\mathbf{f}^{(0)} \times \mathbf{T}_D$
4	F1RIL	$\mathbf{f}^{(0)} \times \mathbf{T}_R$
5	P1BC1RIL	$\mathbf{f}^{(0)} \times \mathbf{T}_{P1B} \times \mathbf{T}_R$
6	P2BC1RIL	$\mathbf{f}^{(0)} \times \mathbf{T}_{P1B} \times \mathbf{T}_R$
7	F2	$\mathbf{f}^{(0)} \times \mathbf{T}_S$
8	F3	$\mathbf{f}^{(0)} \times \mathbf{T}_S \times \mathbf{T}_S$
9	P1BC2F1	$\mathbf{f}^{(0)} \times \mathbf{T}_{P1B} \times \mathbf{T}_{P1B}$
10	P2BC2F1	$\mathbf{f}^{(0)} \times \mathbf{T}_{P2B} \times \mathbf{T}_{P2B}$
11	P1BC2RIL,	$\mathbf{f}^{(0)} \times \mathbf{T}_{P1B} \times \mathbf{T}_{P1B} \times \mathbf{T}_R$
12	P2BC2RIL,	$\mathbf{f}^{(0)} \times \mathbf{T}_{P2B} \times \mathbf{T}_{P2B} \times \mathbf{T}_R$
13	P1BC1F2	$\mathbf{f}^{(0)} \times \mathbf{T}_{P1B} \times \mathbf{T}_S$
14	P2BC1F2	$\mathbf{f}^{(0)} \times \mathbf{T}_{P2B} \times \mathbf{T}_S$
15	P1BC2F2	$\mathbf{f}^{(0)} \times \mathbf{T}_{P1B} \times \mathbf{T}_{P1B} \times \mathbf{T}_S$
16	P2BC2F2	$\mathbf{f}^{(0)} \times \mathbf{T}_{P2B} \times \mathbf{T}_{P2B} \times \mathbf{T}_S$
17	P1BC1DH	$\mathbf{f}^{(0)} \times \mathbf{T}_{P1B} \times \mathbf{T}_D$
18	P2BC1DH	$\mathbf{f}^{(0)} \times \mathbf{T}_{P2B} \times \mathbf{T}_D$
19	P1BC2DH	$\mathbf{f}^{(0)} \times \mathbf{T}_{P1B} \times \mathbf{T}_{P1B} \times \mathbf{T}_D$
20	P2BC2DH	$\mathbf{f}^{(0)} \times \mathbf{T}_{P2B} \times \mathbf{T}_{P2B} \times \mathbf{T}_D$

Several Rounds of Random Mating before Repeated Selfing

How about several rounds of random mating before the repeated selfing?

- The IBM population of maize:
 - B73 and Mo17 are the two parental lines
 - RILs, but there are 4 rounds of random mating before the repeated selfing, therefore was named IBM
- Random mating can enlarge the recombination frequency, so that close linkage may be separated.

Accumulated recombination frequency

$$D_t = \frac{1}{4} (1 - 2r)(1 - r)^{t-1}$$

Rounds of random mating (t)	Mapping distance (cM)		
1 (=F2)	1.00	2.00	5.00
2	1.50	2.99	7.44
3	2.00	3.98	9.88
4	2.49	4.97	12.31
5	2.99	5.96	14.75
6	3.49	6.95	17.19
7	3.99	7.94	19.63
8	4.48	8.93	22.06
9	4.98	9.92	24.50
10	5.48	10.91	26.94
F1-RIL	1.98	3.92	9.55

Frequencies of AABB, AAbb, aaBB and aabb in RILs?!

- From genotype frequencies (10 genotypes), work out haplotype frequencies (4 haplotypes)
- Work out the haplotype frequencies (4 haplotypes) after several rounds of random mating from

$$D_t = D_1(1-r)^{t-1}, \text{ i.e.}$$

$$f_{AB}^{(t)} = f_A f_B + D_t \quad f_{Ab}^{(t)} = f_A f_b - D_t \quad f_{aB}^{(t)} = f_a f_B - D_t \quad f_{ab}^{(t)} = f_a f_b + D_t$$

- Work out genotype frequencies
- Use generation matrix T_R to find out the genotype frequencies in RILs
- **The genetic analysis can be very complicated even with biparental populations!**

A dominant resistance gene is linked with a co-dominant molecular marker

F2 population	Resistant			Susceptible		
Marker type	A	H	B	A	H	B
Sample size	572	1161	14	3	22	569

Marker types A and B are parental types; H is the type of F1 hybrid

Resistant and susceptible can be fitted by the 3:1 ratio (one dominance gene locus): $\chi^2=0.17$ ($P=0.68$). Marker types A, H, and B can be fitted by the 1:2:1 ratio (one co-dominance gene locus) : $\chi^2=0.32$ ($df=2, P=0.85$)

But Resistance and Marker are not independent, i.e. can not be fitted by the 3:6:3:1:2:1 ratio.

The genetic distance between the gene and marker is estimated at 1.8 cM

	Resistant F2			Susceptible F2		
Marker type	A	H	B	A	H	B
Size	572	1161	14	3	22	569

of recombination frequency

character **P1:AABB** **P2:aabb**

e selfing generation of F1

a are co-dominant, B is dominant to b

marker classes and observed samples

AAB_	AAbb	AaB_	Aabb	aaB_	aabb
572	3	1161	22	14	569

Run

Clear

Save

Results

Recombination Frequency	0.017886
Variance	0.000008
Standard Deviation	0.002748
LOD Score	488.549400
P-Value	0.000000
cM (Haldane)	1.82
cM (Kosambi)	1.79

The genetic distance is at 1.8 cM

	Resistant F2			Susceptible F2		
Marker type	A	H	B	A	H	B
Size	572	1161	14	3	22	569
1/3 of Susceptible						
Marker type	A	H	B	A	H	B
Size	572	1161	14	1	7	190
Dominant marker						
Marker type	AH		B	AH		B
Size	1733		14	25		569
Recessive marker						
Marker type	A	HB		A	HB	
Size	572	1175		3	591	