连锁分析和遗传图谱构建

王建康

中国农业科学院作物科学研究所

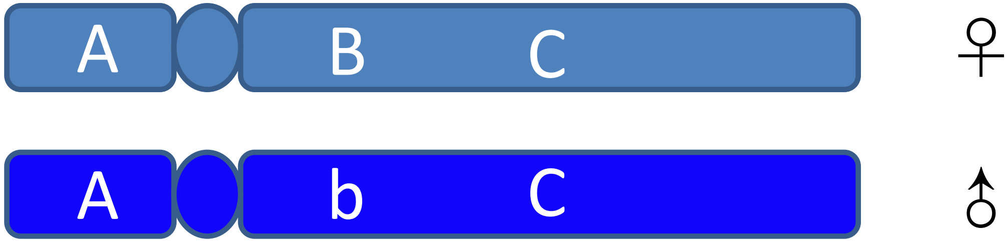
wangjiankang@caas.cn

http://www.isbreeding.net

性状、基因与遗传

Genetics is the science of genes!

- Locus, gene, allele
- Genotype
  - homozygous, homozygousity, e.g. AA, CC
  - heterozygous, heterozygousity, e.g. Bb
- Phenotype: Different genotypes may result in the same phenotype. The same genotype may have different phenotypes in different environments.



Two homologous chromosomes

Qualitative traits in genetics

- Seed shape: round vs wrinkled
- Cotyledon color: yellow vs green
- Seed coat color: grey-brown vs white
- Pod shape: inflated vs constricted
- Unripe pod color: green vs yellow
- Flower position: axial vs terminal
- Stem length: long vs short

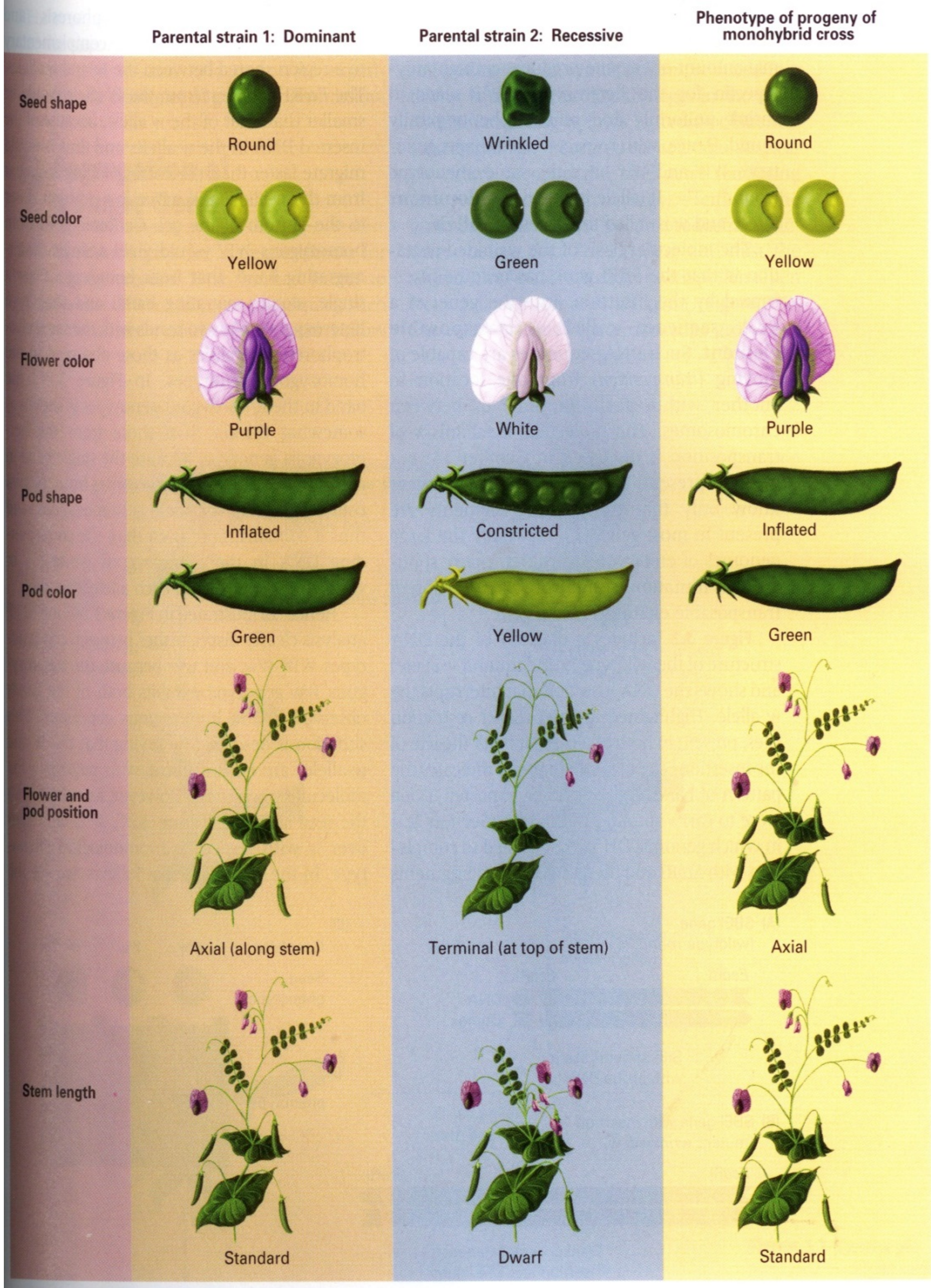


Figure 3.1 The seven different traits in peas studied by Mendel. The phenotype shown at the far right is the dominant trait, which appears in the hybrid produced by crossing.

Rediscovery of Mendel's hybridization experiments in garden pea

- *Experiments with Plant Hybrids* (1866)
  - Seed shape: 5474 round vs 1850 wrinkled
  - Cotyledon color: 6022 yellow vs 2001 green
  - Seed coat color: 705 grey-brown vs 224 white
  - Pod shape: 882 inflated vs 299 constricted
  - Unripe pod color: 428 green vs 152 yellow
  - Flower position: 651 axial vs 207 terminal
  - Stem length: 787 long (20-50cm) vs 277 short (185-230cm)
- Rediscovered in 1900; Era of modern genetics
- Principle of Segregation
- Principle of Independent Assortment

Gregor J. Mendel (1822-1884)

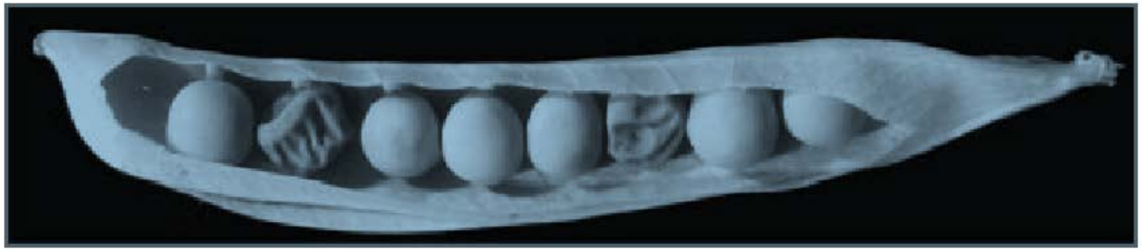
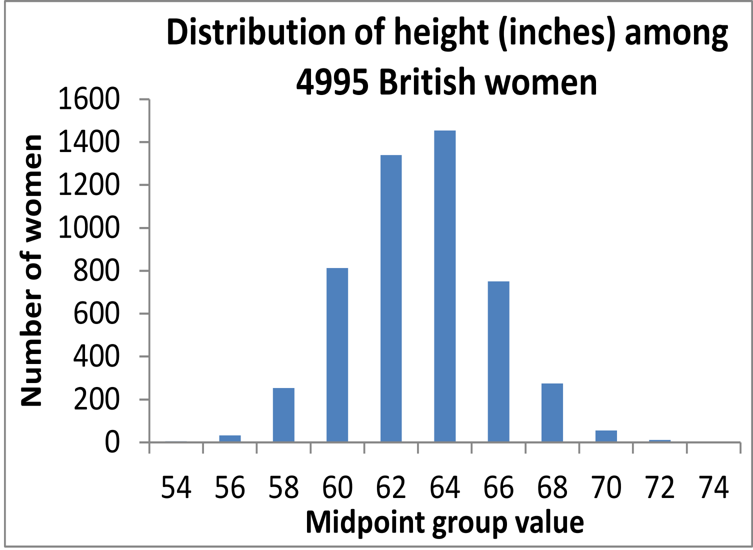
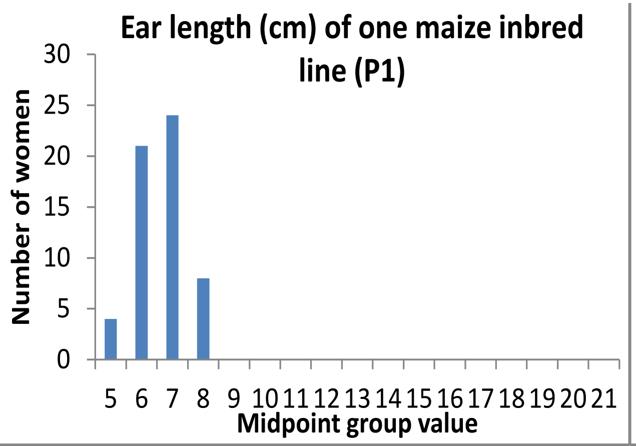
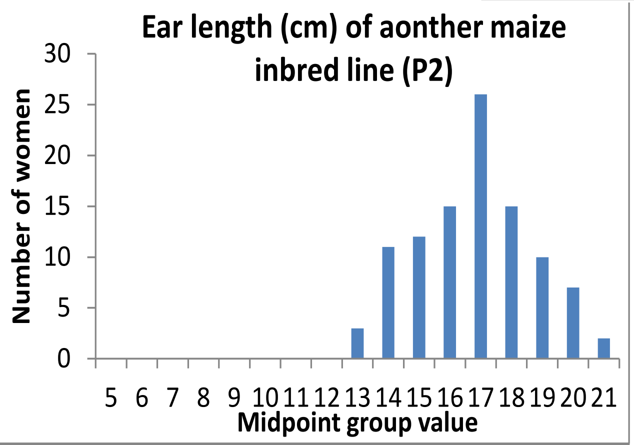


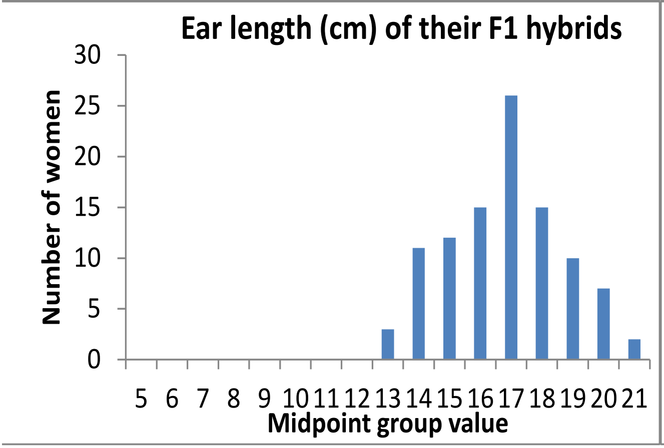
Fig. 1.1. Segregation for wrinkled and smooth seeds within the pea fruits heterozygous for the gene.

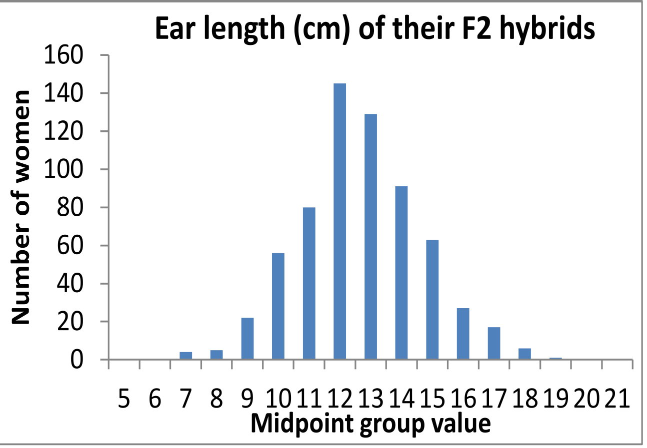
Quantitative traits in genetics


Distribution of height (inches) among 4995 British women


Ear length (cm) of one maize inbred line (P1)


Ear length (cm) of aonther maize inbred line (P2)


Ear length (cm) of their F1 hybrids


Ear length (cm) of their F2 hybrids

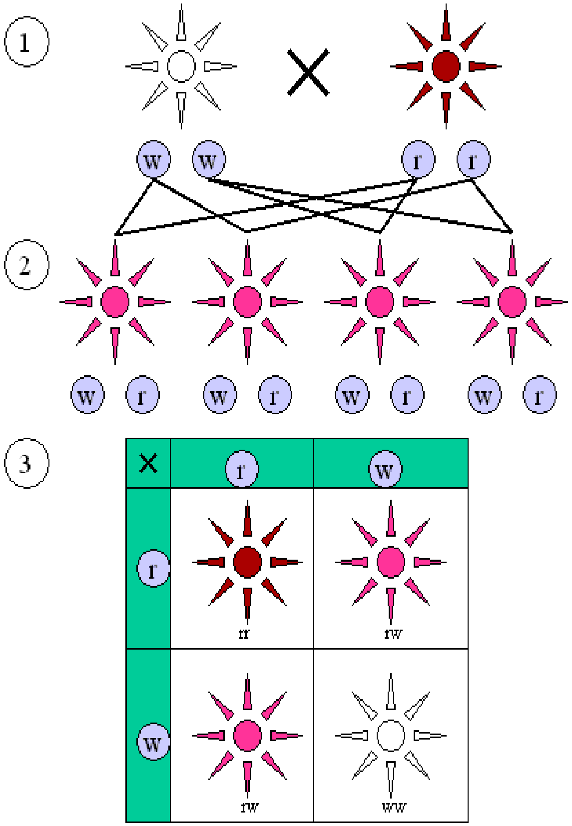
数量遗传研究主线

- 纯系理论：P = G + E + e
- 经典数量遗传学：G = A + D + I
- 混和遗传模型：G =主基因 + 微效多基因
- QTL作图：G =单个QTL效应之和

- 遗传效应的分解
- 遗传方差的分解

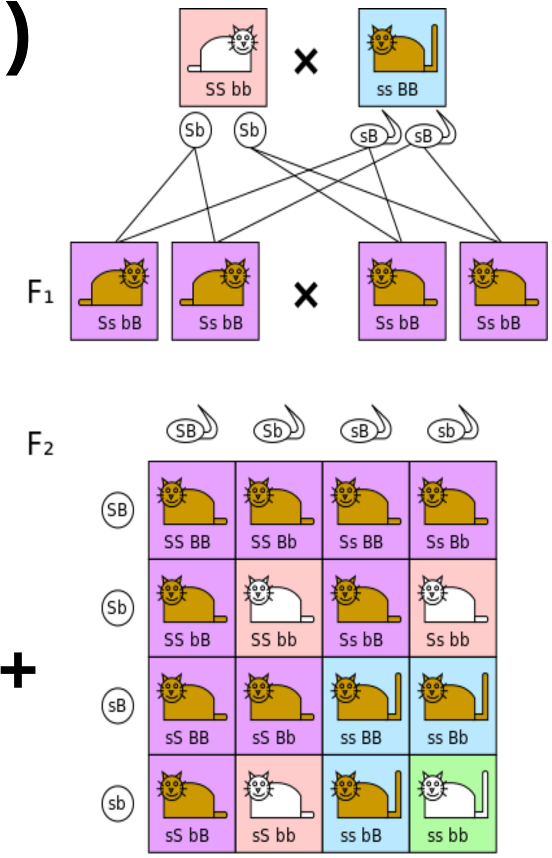
The First Mendelian Law

- The Principle of Segregation (The "First Law").
- For genotype Aa:
  - From zygote to gamete: ½A + ½ 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)



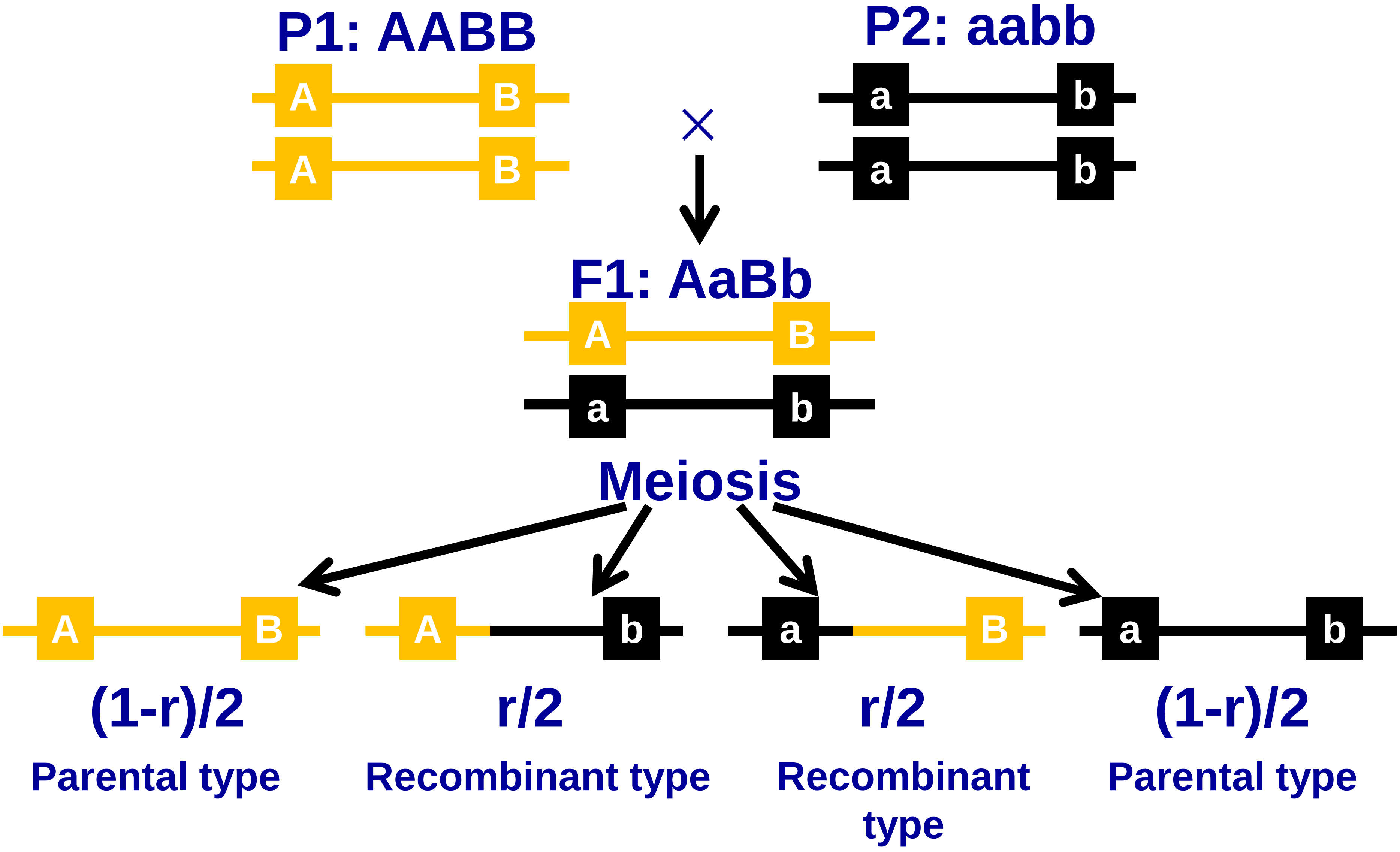
The Second Mendelian Law

- For two independent loci, i.e. no linkage
- The Principle of Independent Assortment (The "Second Law")
- For genotype AaBb:
  - From zygote to gamete: ¼AB + ¼Ab + ¼aB + ¼ab (1:1:1:1)
  - From gamete to zygote: (¼AB + ¼Ab + ¼aB + ¼ab)^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



P1: AABB

P2: aabb

F1: AaBb

Meiosis

(1-r)/2
Parental type

r/2
Recombinant type

r/2
Recombinant type

(1-r)/2
Parental type

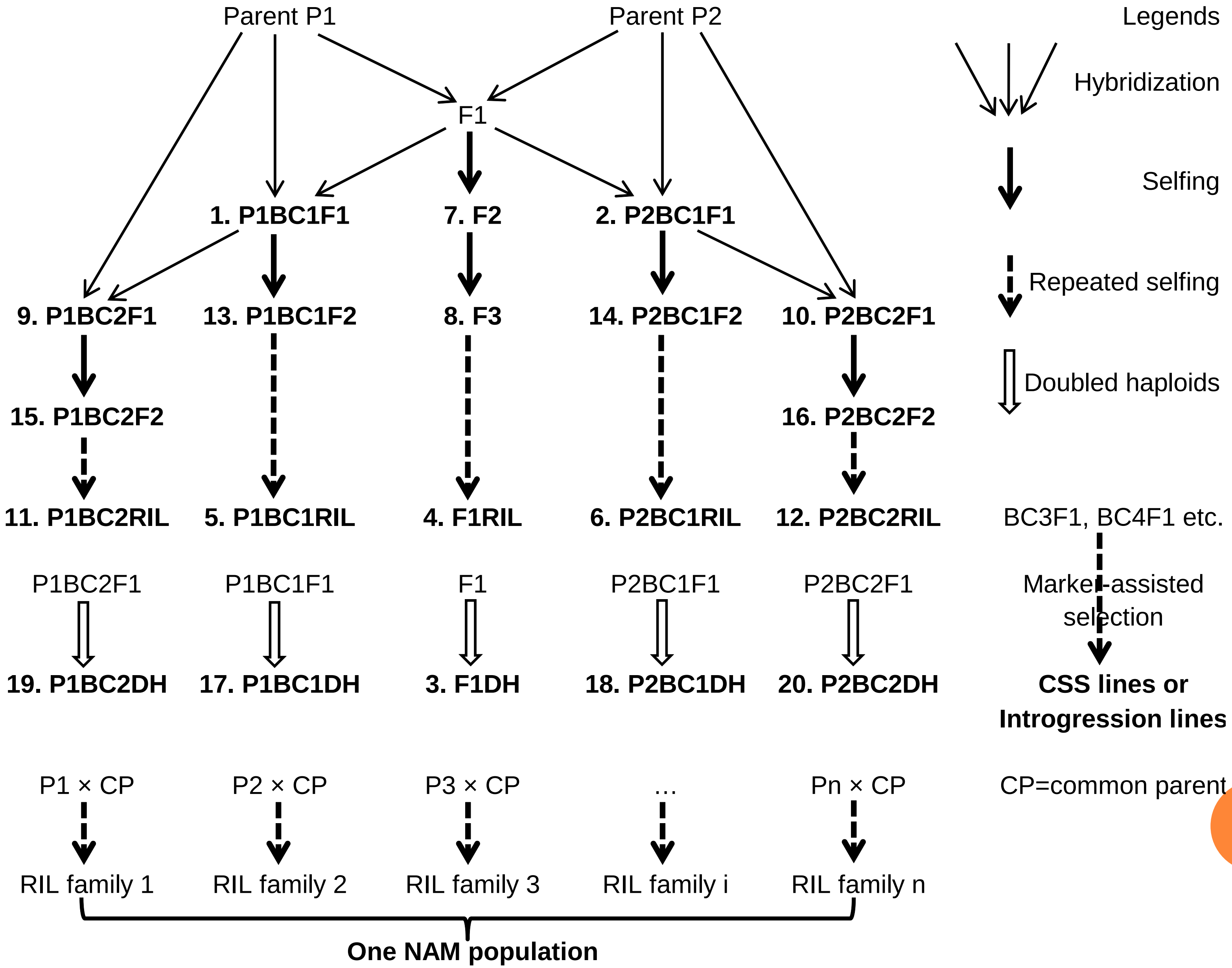
数量性状基因定位的目的

- 控制数量性状的基因
  - 有多少?
  - 在什么地方?
  - 表型效应有多大?
  - 和其它基因间有无互作?
  - 和环境间有无互作?

Dataset of QTL Mapping

- Mapping population
- Linkage map
- Marker genotype
- Phenotypic data

常见的遗传群体



Parent P1 Parent P2

F1

1. **P1BC1F1** 7. **F2** 2. **P2BC1F1**

9. **P1BC2F1** 13. **P1BC1F2** 8. **F3** 14. **P2BC1F2** 10. **P2BC2F1**

15. **P1BC2F2** 16. **P2BC2F2**

11. **P1BC2RIL** 5. **P1BC1RIL** 4. **F1RIL** 6. **P2BC1RIL** 12. **P2BC2RIL**

P1BC2F1 P1BC1F1 F1 P2BC1F1 P2BC2F1

19. **P1BC2DH** 17. **P1BC1DH** 3. **F1DH** 18. **P2BC1DH** 20. **P2BC2DH**

P1 × CP P2 × CP P3 × CP … Pn × CP

RIL family 1 RIL family 2 RIL family 3 RIL family i RIL family n

One NAM population

Legends

Hybridization

Selfing

Repeated selfing

Doubled haploids

BC3F1, BC4F1 etc.

Marker-assisted selection

CSS lines or Introgression lines

CP=common parent

QTL作图群体

- F_2群体
- 回交 (BC, backcross) 群体
- 加倍单倍体 (DH, doubled haploids) 群体
- 重组近交家系 (RIL, recombination inbred lines) 群体
- 导入系（染色体片断置换系）
- 自然群体

作图群体的分类

- 按基因型是否纯合分
  - 暂时群体（Temporary population）
  - 永久群体（Permanent population）
  - 自然群体（Natural population）
- 按群体间的亲缘关系
  - 初级作图群体（Primary mapping population）
  - 次级作图群体（Secondary mapping population）

Example: 10 RILs of Rice (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

连锁分析

遗传标记的种类

- 形态标记
- 细胞学和生化标记
- 分子标记

遗传标记的主要用途

- 建立遗传图谱
- 定位未知基因
- 研究基因间的连锁关系
- 遗传标记应用的两个阶段
  - 自然标记阶段
  - 分子标记阶段

回交群体和DH群体中的期望基因型频率

BC1	BC2	DH群体	观测次数	理论频率
AABB	AaBb	AABB	n_1	$f_1=(1-r)/2$
AABb	Aabb	AAbb	n_2	$f_2=r/2$
AaBB	aaBb	aaBB	n_3	$f_3=r/2$
AaBb	aabb	aabb	n_4	$f_4=(1-r)/2$

重组率的极大似然估计

- **建立似然函数**

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

- **建立对数似然函数**

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

- **求解重组率的极大似然估计** $\hat{r}=\frac{n_2+n_3}{n_1+n_2+n_3+n_4}=\frac{n_2+n_3}{n}$

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

- **应用估计公式求重组率的估计值和它的方差**

$$V_{\hat{r}}=\frac{1}{I}=\frac{\hat{r}(1-\hat{r})}{n}$$

连锁的显著性检验

- H_0: r = 0.5, 即两个标记间无连锁 (独立遗传)
- H_A: r ≠ 0.5, 即两个标记间存在连锁
- 似然比检验 (LRT) 或计算LOD得分

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

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

实　例

- 某回交试验中
  - P1和P2的基因型分别为AABB和aabb
- 回交BC1世代中4种基因型的植株数
  - AABB：162；AABb：40；AaBB：41；AaBb：158

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

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

连锁对独立的检验

- **零假设H_0：r=0.5**
- **备择假设H_A：r<0.5**
- **两种假设下似然函数的比值**

$$\frac{L(\hat{r})}{L(r=0.5)} = \frac{(1-r)^{n_1+n_4} r^{n_2+n_3}}{(\frac{1}{4})^{n_1+n_2+n_3+n_4}} = 6.3\times 10^{153}$$

- **似然比检验统计量(LRT), (P<0.001)**

$$LRT = 2*\ln[\frac{L(\hat{r})}{L(r=0.5)}] = 708.27$$

- **LOD值**

$$LOD = \log[\frac{L(\hat{r})}{L(r=0.5)}] = 153.80$$

DH和RIL群体中的期望基因型频率比较

DH群体	理论频率	RIL群体	理论频率
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个水稻RIL家系群体

亲本P1：MM或A、 P2： mm或B

家系	标记1	标记2	标记类型
	C263	XNpb387	
RIL1	0 或 A	0 或 A	P1亲本型
RIL2	2 或 B	2 或 B	P2亲本型
RIL3	0 或 A	2 或 B	交换型I
RIL4	0 或 A	0 或 A	P1亲本型
RIL5	0 或 A	0 或 A	P1亲本型
RIL6	0 或 A	2 或 B	交换型I
RIL7	0 或 A	0 或 A	P1亲本型
RIL8	2 或 B	2 或 B	P2亲本型
RIL9	0 或 A	0 或 A	P1亲本型
RIL10	0 或 A	0 或 A	P1亲本型

$n_1=6$

$n_2=2$

$n_3=0$

$n_4=2$

$R=2/10=0.2$

$r=0.125$

LRT=17.72, ($P=2.56\times10^{-5}$)

LOD=3.85

F2群体中各种标记型的频率

共显性标记		显性标记	
标记型	频率	标记型	频率
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	$(1-r)^2/4$

F2群体中重组率的估计: 显性标记

- 对数似然函数 $k=(1-r)^2$

$$\ln L = C + n_1\ln(3-2r+r^2) + (n_3+n_7)\ln(2r-r^2) + n_9\ln(1-2r+r^2)$$

$$= C + n_1\ln(2+k) + (n_3+n_7)\ln(1-k) + n_9\ln k$$

- 重组率的估计值

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

- 重组率估计值的方差

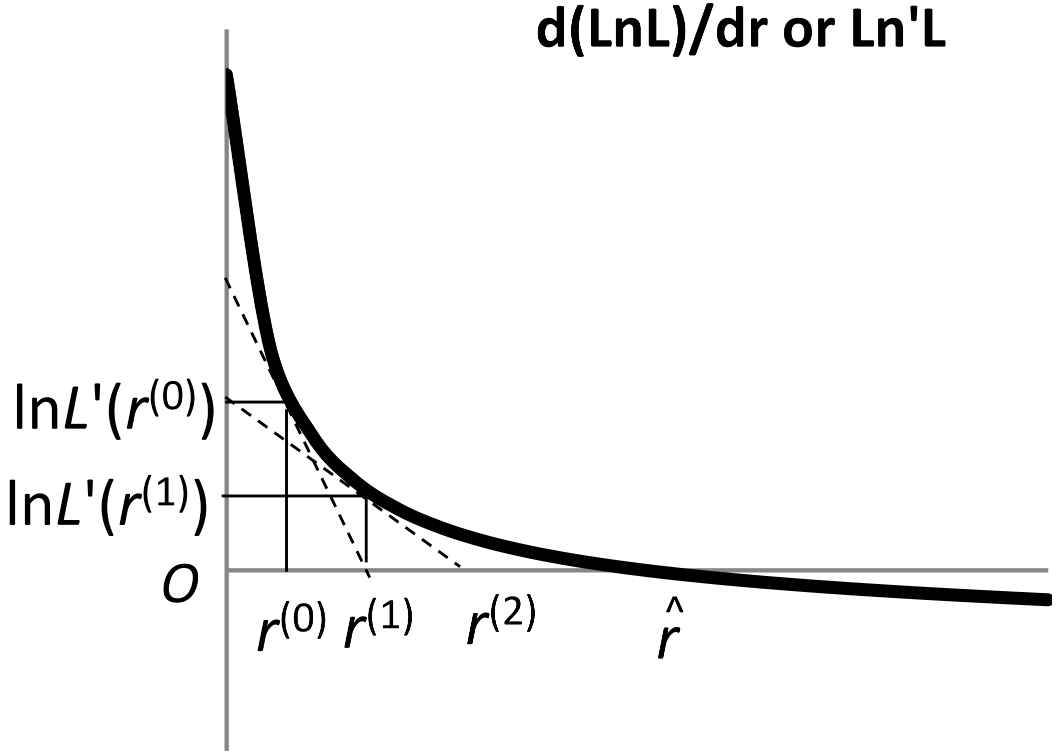
$$V_{\hat{r}}=\frac{(1-k)(2-k)}{2n(1+2k)}=\frac{(2r-r^2)(3-2r+r^2)}{2n(3-4r+2r^2)}$$

F2群体中共显性标记重组率估计的牛顿迭代算法

- 对数似然函数 $$\ln L = \ln C + (2n_1 + 2n_9 + n_2 + n_4 + n_6 + n_8)\ln(1-r) + (n_2 + n_4 + n_6 + n_8 + 2n_3 + 2n_7)\ln r + n_5 \ln(1-2r+2r^2)$$
- 一阶导数 $$f'(r) = \frac{d\ln L}{dr} = \frac{2n_1 + 2n_9 + n_2 + n_4 + n_6 + n_8}{r-1} + \frac{n_2 + n_4 + n_6 + n_8 + 2n_3 + 2n_7}{r} + \frac{n_5(4r-2)}{1-2r+2r^2}$$
- 二阶导数 $$f''(r) = \frac{d^2\ln L}{d^2 r} = -\frac{2n_1 + 2n_9 + n_2 + n_4 + n_6 + n_8}{(r-1)^2} - \frac{n_2 + n_4 + n_6 + n_8 + 2n_3 + 2n_7}{r^2} + \frac{n_5(4r-4r^2)}{(1-2r+2r^2)^2}$$
- 迭代算法 $$r_{i+1} = r_i - f'(r_i)/f''(r_i)$$

Principe of the Newton-Raphson algorithm

- To find the maxima of LnL is equal to find the solution in equation Ln’L=0



F2群体中共显性标记重组率估计的EM迭代算法

- **E-步骤：根据重组率的初始值计算各种标记基因型属于重组基因型的概率。**
- **M-步骤：重新计算重组率**

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

F_2群体中共显性标记间重组率的EM估计

标记型	频率	期望观测值	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更新为: $r' = [n_1 \times 0 + n_2 \times 0.5 + n_3 \times 1 + \ldots + n_8 \times 0.5 + n_9 \times 0] / n$

EM算法迭代3次（初始重组率设为0.5）

基因型	样本量	重组率	期望频率	P(R|G)	重组率	期望频率	P(R|G)	重组率	期望频率	P(R|G)	重组率
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**		

EM算法迭代3次（初始重组率设为0.25）

基因型	样本量	重组率	期望频率	P(R|G)	重组率	期望频率	P(R|G)	重组率	期望频率	P(R|G)	重组率
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**		

EM算法迭代3次（初始重组率设为0.00）

基因型	样本量	重组率	期望频率	P(R|G)	重组率	期望频率	P(R|G)	重组率	期望频率	P(R|G)	重组率
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**	**0.025355**	**0.148**	**0.374**	**0.029153**	**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**		

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

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

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

The generation matrix (T) for the backcrossing to P2

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

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

The generation matrix (T) for the selfing pollination

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

$$\mathbf{T}_{\mathrm{S}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \tfrac{1}{4} & \tfrac{1}{2} & \tfrac{1}{4} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \tfrac{1}{4} & 0 & 0 & \tfrac{1}{2} & 0 & 0 & 0 & \tfrac{1}{4} & 0 & 0 \\ \tfrac{1}{4}(1-r)^2 & \tfrac{1}{2}r(1-r) & \tfrac{1}{4}r^2 & \tfrac{1}{2}r(1-r) & \tfrac{1}{2}(1-r)^2 & \tfrac{1}{2}r^2 & \tfrac{1}{2}r(1-r) & \tfrac{1}{4}r^2 & \tfrac{1}{2}r(1-r) & \tfrac{1}{4}(1-r)^2 \\ \tfrac{1}{4}r^2 & \tfrac{1}{2}r(1-r) & \tfrac{1}{4}(1-r)^2 & \tfrac{1}{2}r(1-r) & \tfrac{1}{2}r^2 & \tfrac{1}{2}(1-r)^2 & \tfrac{1}{2}r(1-r) & \tfrac{1}{4}(1-r)^2 & \tfrac{1}{2}r(1-r) & \tfrac{1}{2}r^2 \\ 0 & 0 & \tfrac{1}{4} & 0 & 0 & 0 & \tfrac{1}{2} & 0 & 0 & \tfrac{1}{4} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \tfrac{1}{4} & \tfrac{1}{2} & \tfrac{1}{4} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

The generation matrix (T) for the doubled haploids

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

$$\mathbf{T}_{\mathrm{D}} = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\tfrac{1}{2} & 0 & \tfrac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\tfrac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & \tfrac{1}{2} & 0 & 0 \\
\tfrac{1}{2}(1-r) & 0 & \tfrac{1}{2}r & 0 & 0 & 0 & 0 & \tfrac{1}{2}r & 0 & \tfrac{1}{2}(1-r) \\
\tfrac{1}{2}r & 0 & \tfrac{1}{2}(1-r) & 0 & 0 & 0 & 0 & \tfrac{1}{2}(1-r) & 0 & \tfrac{1}{2}r \\
0 & 0 & \tfrac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & \tfrac{1}{2} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \tfrac{1}{2} & 0 & \tfrac{1}{2} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}$$

The generation matrix (T) for the repeated selfing pollination

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

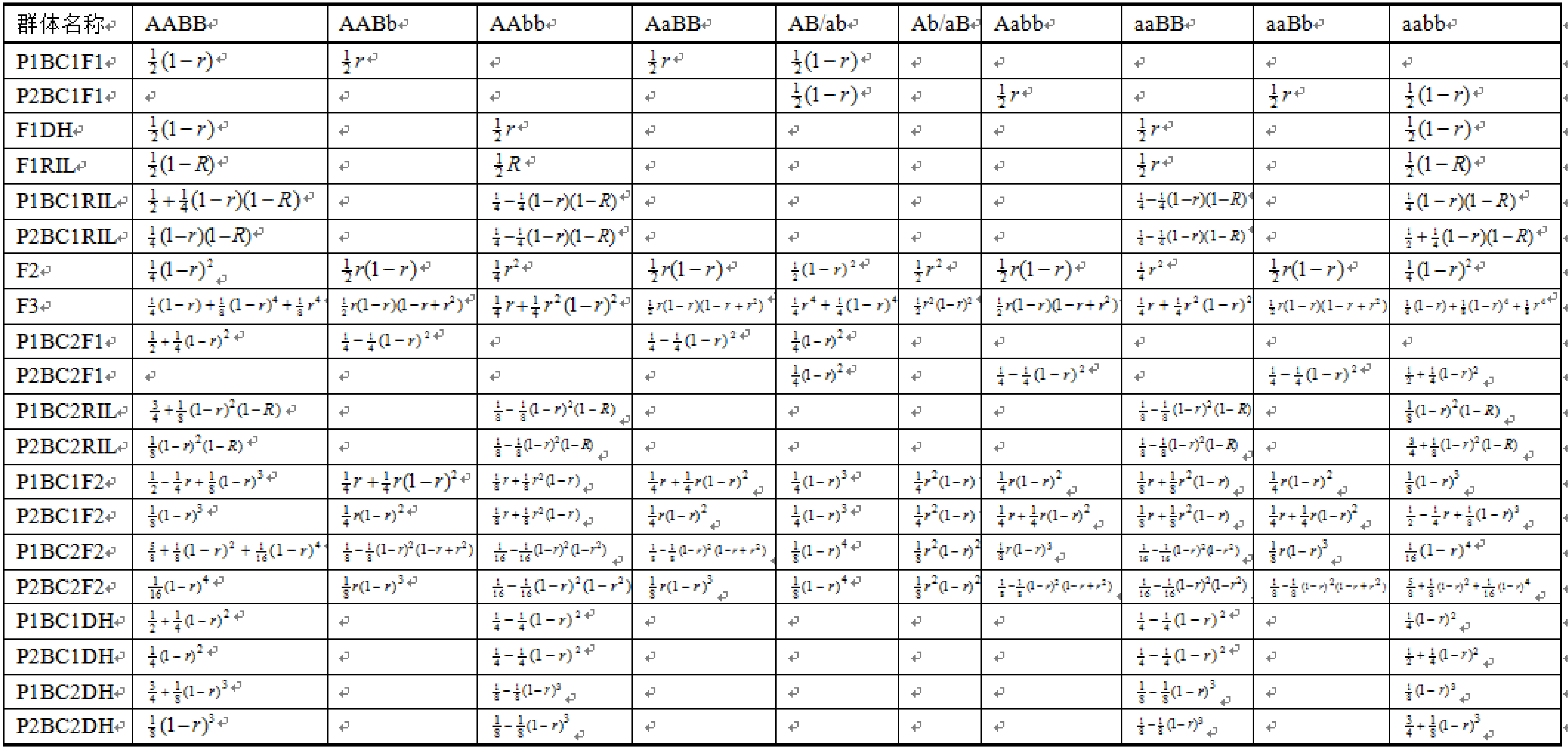
$$\mathbf{T}_{\mathrm{R}}=\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
\tfrac{1}{2} & 0 & \tfrac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
\tfrac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & \tfrac{1}{2} & 0 & 0\\
\tfrac{1}{2}(1-R) & 0 & \tfrac{1}{2}R & 0 & 0 & 0 & 0 & \tfrac{1}{2}R & 0 & \tfrac{1}{2}(1-R)\\
\tfrac{1}{2}R & 0 & \tfrac{1}{2}(1-R) & 0 & 0 & 0 & 0 & \tfrac{1}{2}(1-R) & 0 & \tfrac{1}{2}R\\
0 & 0 & \tfrac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & \tfrac{1}{2}\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \tfrac{1}{2} & 0 & \tfrac{1}{2}\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}$$

Theoretical frequencies of the 20 biparental populations

$$\mathbf{f}^{(0)} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

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	$\mathbf{f}^{(0)} \times \mathbf{T}_{P2B} \times \mathbf{T}_{P2B} \times \mathbf{T}_{D}$

In formulas, …



群体名称	AABB	AABb	AAbb	AaBB	AB/ab	Ab/aB	Aabb	aaBB	aaBb	aabb
P1BC1F1	$\frac{1}{2}(1-r)$	$\frac{1}{2}r$		$\frac{1}{2}r$	$\frac{1}{2}(1-r)$					
P2BC1F1					$\frac{1}{2}(1-r)$		$\frac{1}{2}r$		$\frac{1}{2}r$	$\frac{1}{2}(1-r)$
F1DH	$\frac{1}{2}(1-r)$		$\frac{1}{2}r$					$\frac{1}{2}r$		$\frac{1}{2}(1-r)$
F1RIL	$\frac{1}{2}(1-R)$		$\frac{1}{2}R$					$\frac{1}{2}r$		$\frac{1}{2}(1-R)$
P1BC1RIL	$\frac{1}{2}+\frac{1}{4}(1-r)(1-R)$		$\frac{1}{4}-\frac{1}{4}(1-r)(1-R)$					$\frac{1}{4}-\frac{1}{4}(1-r)(1-R)$		$\frac{1}{4}(1-r)(1-R)$
P2BC1RIL	$\frac{1}{4}(1-r)(1-R)$		$\frac{1}{4}-\frac{1}{4}(1-r)(1-R)$					$\frac{1}{2}-\frac{1}{2}(1-r)(1-R)$		$\frac{1}{2}+\frac{1}{4}(1-r)(1-R)$
F2	$\frac{1}{4}(1-r)^2$	$\frac{1}{2}r(1-r)$	$\frac{1}{4}r^2$	$\frac{1}{2}r(1-r)$	$\frac{1}{2}(1-r)^2$	$\frac{1}{2}r^2$	$\frac{1}{2}r(1-r)$	$\frac{1}{4}r^2$	$\frac{1}{2}r(1-r)$	$\frac{1}{4}(1-r)^2$
F3	$\frac{1}{4}(1-r)+\frac{1}{8}(1-r)^4+\frac{1}{8}r^4$	$\frac{1}{2}r(1-r)(1-r+r^2)$	$\frac{1}{4}r+\frac{1}{4}r^2(1-r)^2$	$\frac{1}{2}r(1-r)(1-r+r^2)$	$\frac{1}{4}r^4+\frac{1}{4}(1-r)^4$	$\frac{1}{2}r^2(1-r)^2$	$\frac{1}{2}r(1-r)(1-r+r^2)$	$\frac{1}{4}r+\frac{1}{4}r^2(1-r)^2$	$\frac{1}{2}r(1-r)(1-r+r^2)$	$\frac{1}{4}(1-r)+\frac{1}{8}(1-r)^4+\frac{1}{8}r^4$
P1BC2F1	$\frac{1}{2}+\frac{1}{4}(1-r)^2$	$\frac{1}{4}-\frac{1}{4}(1-r)^2$		$\frac{1}{4}-\frac{1}{4}(1-r)^2$	$\frac{1}{4}(1-r)^2$					
P2BC2F1					$\frac{1}{4}(1-r)^2$		$\frac{1}{4}-\frac{1}{4}(1-r)^2$		$\frac{1}{4}-\frac{1}{4}(1-r)^2$	$\frac{1}{2}+\frac{1}{4}(1-r)^2$
P1BC2RIL	$\frac{3}{4}+\frac{1}{8}(1-r)^2(1-R)$		$\frac{1}{8}-\frac{1}{8}(1-r)^2(1-R)$					$\frac{1}{8}-\frac{1}{8}(1-r)^2(1-R)$		$\frac{1}{8}(1-r)^2(1-R)$
P2BC2RIL	$\frac{1}{8}(1-r)^2(1-R)$		$\frac{1}{8}-\frac{1}{8}(1-r)^2(1-R)$					$\frac{1}{8}-\frac{1}{8}(1-r)^2(1-R)$		$\frac{3}{4}+\frac{1}{8}(1-r)^2(1-R)$
P1BC1F2	$\frac{1}{2}-\frac{1}{4}r+\frac{1}{8}(1-r)^3$	$\frac{1}{4}r+\frac{1}{4}r(1-r)^2$	$\frac{1}{8}r+\frac{1}{8}r^2(1-r)$	$\frac{1}{4}r+\frac{1}{4}r(1-r)^2$	$\frac{1}{4}(1-r)^3$	$\frac{1}{4}r^2(1-r)$	$\frac{1}{4}r(1-r)^2$	$\frac{1}{8}r+\frac{1}{8}r^2(1-r)$	$\frac{1}{4}r(1-r)^2$	$\frac{1}{8}(1-r)^3$
P2BC1F2	$\frac{1}{8}(1-r)^3$	$\frac{1}{4}r(1-r)^2$	$\frac{1}{8}r+\frac{1}{8}r^2(1-r)$	$\frac{1}{4}r(1-r)^2$	$\frac{1}{4}(1-r)^3$	$\frac{1}{4}r^2(1-r)$	$\frac{1}{4}r+\frac{1}{4}r(1-r)^2$	$\frac{1}{8}r+\frac{1}{8}r^2(1-r)$	$\frac{1}{4}r+\frac{1}{4}r(1-r)^2$	$\frac{1}{2}-\frac{1}{4}r+\frac{1}{8}(1-r)^3$
P1BC2F2	$\frac{5}{8}+\frac{1}{8}(1-r)^2+\frac{1}{16}(1-r)^4$	$\frac{1}{8}-\frac{1}{8}(1-r)^2(1-r+r^2)$	$\frac{1}{16}-\frac{1}{16}(1-r)^2(1-r^2)$	$\frac{1}{8}-\frac{1}{8}(1-r)^2(1-r+r^2)$	$\frac{1}{8}(1-r)^4$	$\frac{1}{8}r^2(1-r)^2$	$\frac{1}{8}r(1-r)^3$	$\frac{1}{16}-\frac{1}{16}(1-r)^2(1-r^2)$	$\frac{1}{8}r(1-r)^3$	$\frac{1}{16}(1-r)^4$
P2BC2F2	$\frac{1}{16}(1-r)^4$	$\frac{1}{8}r(1-r)^3$	$\frac{1}{16}-\frac{1}{16}(1-r)^2(1-r^2)$	$\frac{1}{8}r(1-r)^3$	$\frac{1}{8}(1-r)^4$	$\frac{1}{8}r^2(1-r)^2$	$\frac{1}{8}-\frac{1}{8}(1-r)^2(1-r+r^2)$	$\frac{1}{16}-\frac{1}{16}(1-r)^2(1-r^2)$	$\frac{1}{8}-\frac{1}{8}(1-r)^2(1-r+r^2)$	$\frac{5}{8}+\frac{1}{8}(1-r)^2+\frac{1}{16}(1-r)^4$
P1BC1DH	$\frac{1}{2}+\frac{1}{4}(1-r)^2$		$\frac{1}{4}-\frac{1}{4}(1-r)^2$					$\frac{1}{4}-\frac{1}{4}(1-r)^2$		$\frac{1}{4}(1-r)^2$
P2BC1DH	$\frac{1}{4}(1-r)^2$		$\frac{1}{4}-\frac{1}{4}(1-r)^2$					$\frac{1}{4}-\frac{1}{4}(1-r)^2$		$\frac{1}{2}+\frac{1}{4}(1-r)^2$
P1BC2DH	$\frac{3}{4}+\frac{1}{8}(1-r)^3$		$\frac{1}{8}-\frac{1}{8}(1-r)^3$					$\frac{1}{8}-\frac{1}{8}(1-r)^3$		$\frac{1}{8}(1-r)^3$
P2BC2DH	$\frac{1}{8}(1-r)^3$		$\frac{1}{8}-\frac{1}{8}(1-r)^3$					$\frac{1}{8}-\frac{1}{8}(1-r)^3$		$\frac{3}{4}+\frac{1}{8}(1-r)^3$

不同群体中各种共显性标记型的理论频率

标记型	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	$(1-r)^2/4$		$(1-r)/2$	$(1-r)/2$	$(1-r)^2/4$	$\frac{1}{2}+(1-r)^2/4$	$(1-R)/2$

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 meting 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

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

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

$$f_{\mathrm{AB}}^{(t)} = f_{\mathrm{A}} f_{\mathrm{B}} + D_t \quad f_{\mathrm{Ab}}^{(t)} = f_{\mathrm{A}} f_{\mathrm{b}} - D_t \quad f_{\mathrm{aB}}^{(t)} = f_{\mathrm{a}} f_{\mathrm{B}} - D_t \quad f_{\mathrm{ab}}^{(t)} = f_{\mathrm{a}} f_{\mathrm{b}} + D_t$$

- 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 bi-parental populations!**

奇异分离对重组率估计的影响？

DH群体	理论频率	奇异分离	奇异分离理论频率
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×r/2	r×s/(1+s)
aabb	f_4=(1-r)/2	s×(1-r)/2	(1-r)×s/(1+s)
总和	1	(1+s)/2	1

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

遗传图谱构建

3个标记间的重组率

$$r_{13}=r_{12}+r_{23}-2(1-\delta)\ r_{12}r_{23}$$

- ❋ **当 $\delta=0$ (即两个区间上的交换是独立的) 时, 有**

  $$(1-r_{13})=(1-r_{12})(1-r_{23})+r_{12}r_{23}$$ **或**

  $$r_{13}=r_{12}(1-r_{23})+(1-r_{12})r_{23}=r_{12}+r_{23}-2r_{12}r_{23}$$

- ❋ **当时 $\delta=1$(即完全干涉, 一个区间上的交换完全阻止另外一个区间上的交换), 有** $r_{13}=r_{12}+r_{23}$

作图函数

- **图距(Mapping distance)**

$$m_{13} = m_{12} + m_{23}$$

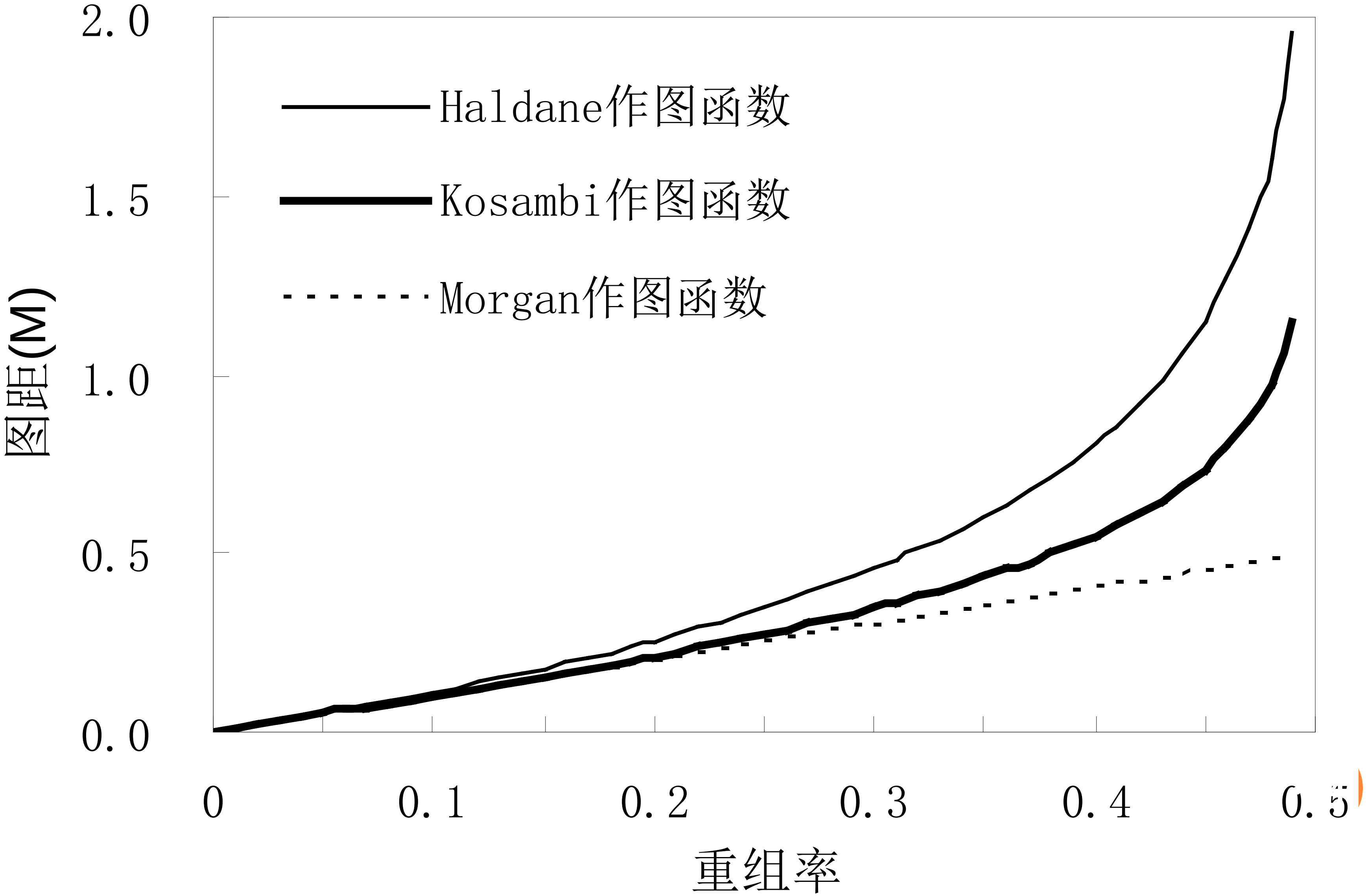
- **图距的单位：摩尔根（M, Morgan）或厘摩（cM，centi-Morgan），1M=100cM**

- **图距*m*是交换率*r*的函数，即：**$m = f(r)$**，称*f* 为作图函数 (Mapping function)。**

常见作图函数

- **Morgan 作图函数**
  - ✓**以M为单位** $m = r$ **(M)**
  - ✓**以cM为单位** $m = r \times 100$ **(cM)**
- **Haldane 作图函数 没有考虑干涉的情况下，即M1-M2间的　交换和M2-M3间的交换相互独立**
  - ✓**以M为单位** $m = f(r) = -\frac{1}{2}\ln(1-2r)$ $r = \frac{1}{2}(1-e^{-2m})$
  - ✓**以cM为单位** $m = f(r) = -50\ln(1-2r)$ $r = \frac{1}{2}(1-e^{-m/50})$
- **Kosambi作图函数 考虑干涉的情况下，即M1-M2间的交换和M2-M3间的交换不独立，干涉系数应重组率的函数**
  - ✓**以M为单位** $m = \frac{1}{4}\ln\frac{1+2r}{1-2r}$ $r = \frac{1}{2}\frac{e^{4m}-1}{e^{4m}+1}$
  - ✓**以cM为单位** $m = 25\ln\frac{1+2r}{1-2r}$ $r = \frac{1}{2}\frac{e^{m/25}-1}{e^{m/25}+1}$

三种作图函数的比较


2.0
1.5
1.0
0.5
0.0
图距(M)
Haldane作图函数
Kosambi作图函数
Morgan作图函数
0
0.1
0.2
0.3
0.4
0.5
重组率

构建连锁图谱步骤

- 估计标记间的重组率

- 标记分群 (Grouping)：包括Anchor信息的利用
- 标记排序 (Ordering)：多种排序算法
- 序列调整 (Rippling)：根据多种标准

- 目的：长度最短的连锁图谱

Three steps in map construction

- **Step 1: Grouping.** Grouping can be based on
  - (i) a threshold of LOD score
  - (ii) a threshold of marker distance (cM)
  - (iii) anchor information

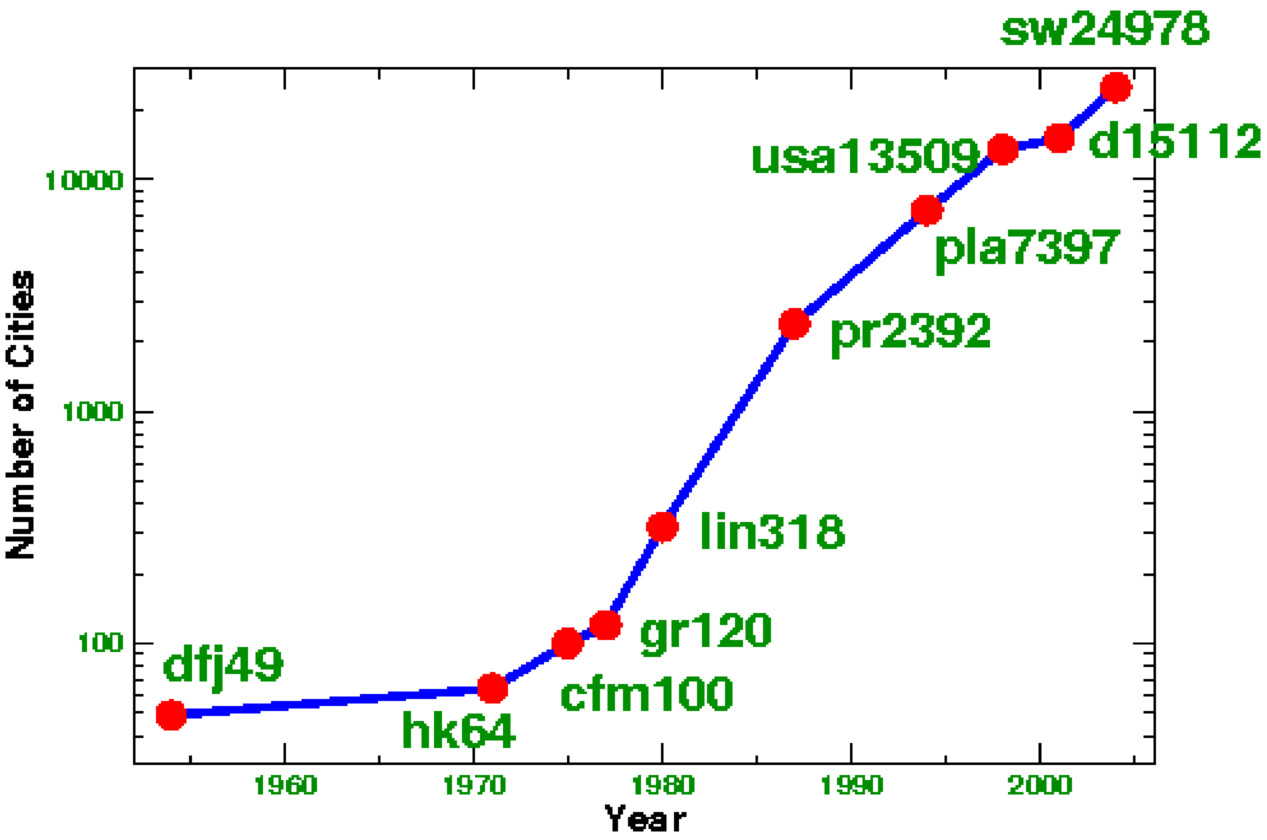
Three steps in map construction

- **Step 2: Ordering.** Three ordering algorithms are
  - (i) SER: SERiation (Buetow and Chakravarti, 1987. Am J Hum Genet 41:180–188)
  - (ii) RECORD: REcombination Counting and ORDering (Van Os et al., 2005. Theor Appl Genet 112: 30–40)
  - (iii) nnTwoOpt: nearest neighbor was used for tour construction, and two-opt was used for tour improvement, similar to Travelling Salesman Problem (TSP) (Lin and Kernighan, 1973. Oper. Res. 21: 498–516.
  - Or by the input order, say we know the order from physical map for GBS markers

Travelling Salesman Problem (TSP)

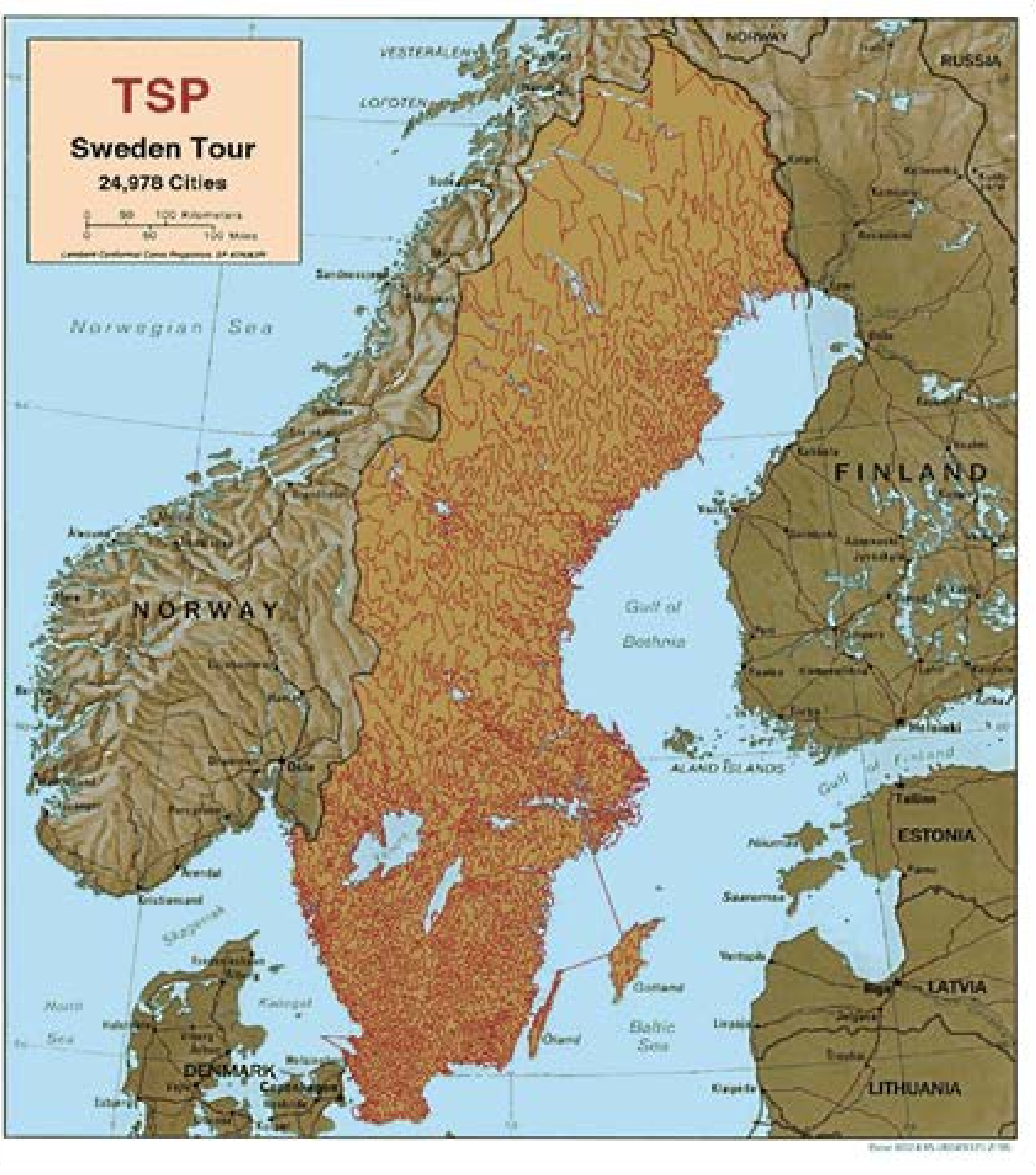
- A salesman is required to visit each of n given cities once and only once, starting from any city and returning to the original place of departure. What tour he choose in order to minimize his total travel distance?
- The distance between any pair of cities are assumed to be known by the salesman. Distance can be replaced by another notion, such as time or money.
- TSP is one of the most widely studied problems in combinatorial optimization. It is easy to state, but hard to solve! **TSP is an NP-hard problem, i.e. non-deterministic polynomial-time hard.**

Finding the solutions



- TSP is represented by some letters plus the number of cities. For example, there are 24978 Sweden cities in TSP “sw24978”.

The solution for TSP “sw24978”



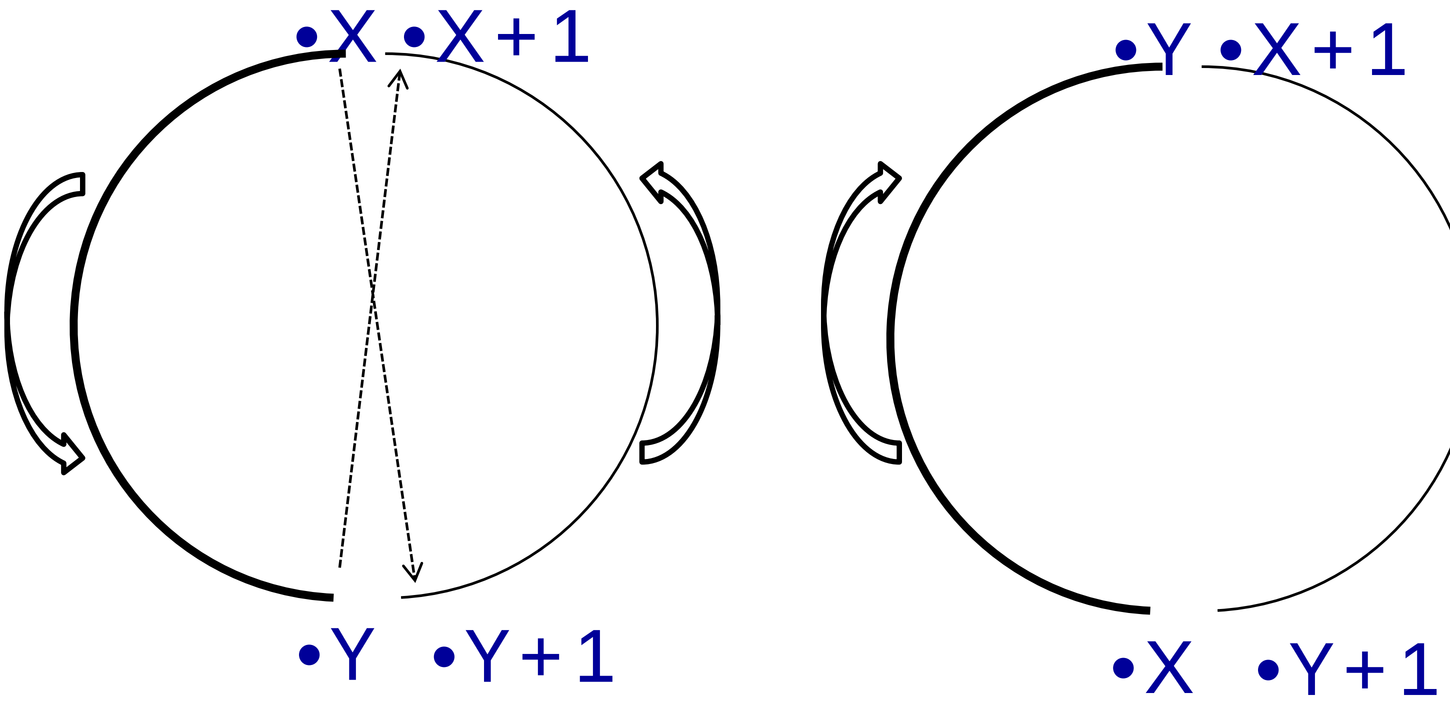
Approximate algorithm of TSP

- Tour construction algorithms
- Tour improvement algorithms
- Composite algorithms

Nearest-neighbor (nn) algorithm for tour construction

- A simple algorithm for tour construction
- Start in an arbitrary city. As long as there are cities, that have not yet been visited, visit the nearest city that still not appeared in the tour. Finally, return to the first city.
- This approach is simple, but often too greedy!
- The first distances in the construction process are reasonable short, whereas the distance at the end of the process usually will be rather long.

Two-Opt algorithm for tour improvement (Lin and Kernighan, 1973)



•X •X+1

•Y •Y+1

•Y •X+1

•X •Y+1

Three steps in map construction

- Due to the large number of markers (n), it is impossible to compare all possible orders (say n=50, possible orders are $n!/2=1.52\times10^{64}$). Orders from the above algorithms are regional optimizations.
- **Step 3: Rippling.** Five rippling criteria are
  - (i) SARF (Sum of Adjacent Recombination Frequencies)
  - (ii) SAD (Sum of Adjacent Distances)
  - (iii) SALOD (Sum of Adjacent LOD scores)
  - (iv) COUNT (number of recombination events)
- **Universe criteria: shortest map!**

Linkage map and physical map

Species	Size of haploid genome (kb)	Size of linkage map (cM)	kb/cM
Yeast	2.2×10^4	3700	6
Neurospora	4.2×10^4	500	80
Arabidopsis	7.0×10^4	500	140
Drosophila	2.0×10^5	290	700
Tomato	7.2×10^5	1400	510
Human	3.0×10^6	2710	1110
Wheat	1.6×10^7	2575	6214
Rice	4.4×10^5	1575	279
Corn	3.0×10^6	1400	2140

The MAP functionality in QTL IciMapping

Coding of co-dominant marker (P1 and P2 bands are both present in F1)

	Two parental lines P1 (MM), P2 (mm) and their F1 hybrid (Mm)			Populations with heterozygousity, i.e., 1, 2, 7, 8, 9, 10, 13, 14, 15, 16			Populations with no heterozygousity, i.e., 3, 4, 5, 6, 11, 12, 17, 18, 19, 20	
	P1: MM	P2: mm	F1: Mm	MM	mm	Mm	MM	mm
Code	2	0	1	2	0	1	2	0

	Two parental lines P1 (MM), P2 (mm) and their F1 hybrid (Mm)			Populations with heterozygousity, i.e., 1, 2, 7, 8, 9, 10, 13, 14, 15, 16			Populations with no heterozygousity, i.e., 3, 4, 5, 6, 11, 12, 17, 18, 19, 20	
	P1: MM	P2: mm	F1: Mm	MM	mm	Mm	MM	mm
Code	2	0	1	2	0	1	2	0

When heterozygote F1 is present, P1 is coded as 2, F1 is coded as 1, and P2 is coded as 0. When heterozygote F1 is absent, P1 is coded as 2, and P2 is coded as 0.

Coding of dominant marker (F1 band is the same as the P1 band)

Two parental lines P1 (MM), P2 (mm) and their F1 hybrid (Mm)

Populations with heterozygousity, i.e., 1, 2, 7, 8, 9, 10, 13, 14, 15, 16

Populations with no heterozygousity, i.e., 3, 4, 5, 6, 11, 12, 17, 18, 19, 20

	P1: MM	P2: mm	F1: Mm	MM	mm	Mm	MM	mm
Code	2	0	1	12	0	12	2	0

Two parental lines P1 (MM), P2 (mm) and their F1 hybrid (Mm)

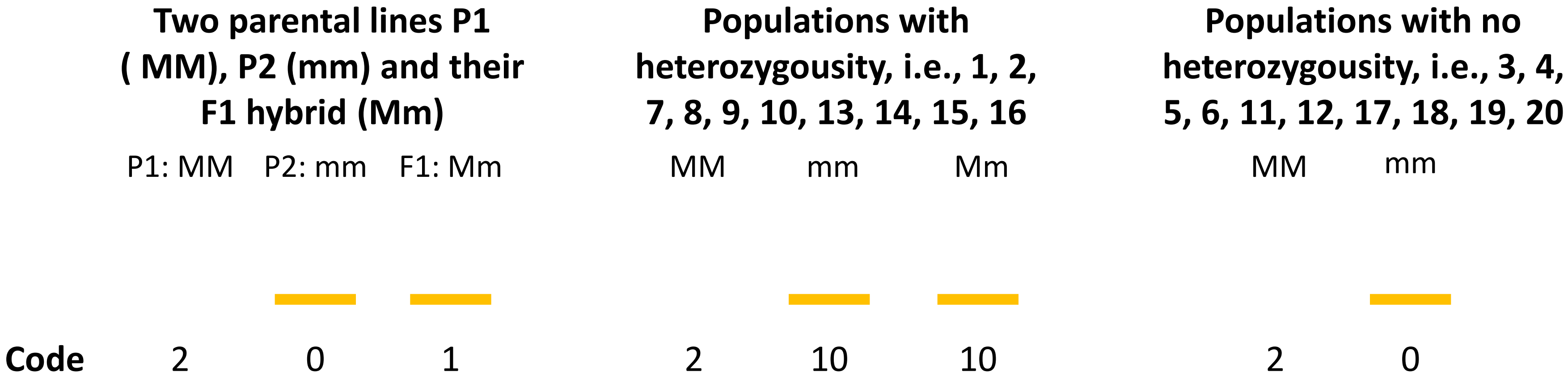
Populations with heterozygousity, i.e., 1, 2, 7, 8, 9, 10, 13, 14, 15, 16

Populations with no heterozygousity, i.e., 3, 4, 5, 6, 11, 12, 17, 18, 19, 20

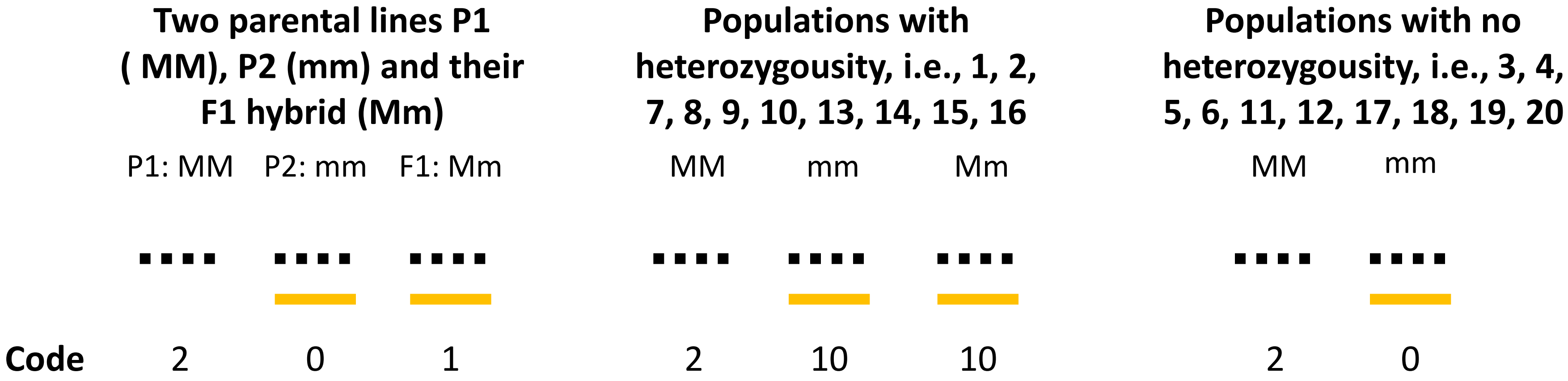
	P1: MM	P2: mm	F1: Mm	MM	mm	Mm	MM	mm
Code	2	0	1	12	0	12	2	0

When heterozygote F1 is present, P1 and F1 are coded as 12, and P2 is coded as 0.
When heterozygote F1 is absent, P1 is coded as 2, and P2 is coded as 0.

Coding of recessive marker
(F1 band is the same as P2 band)



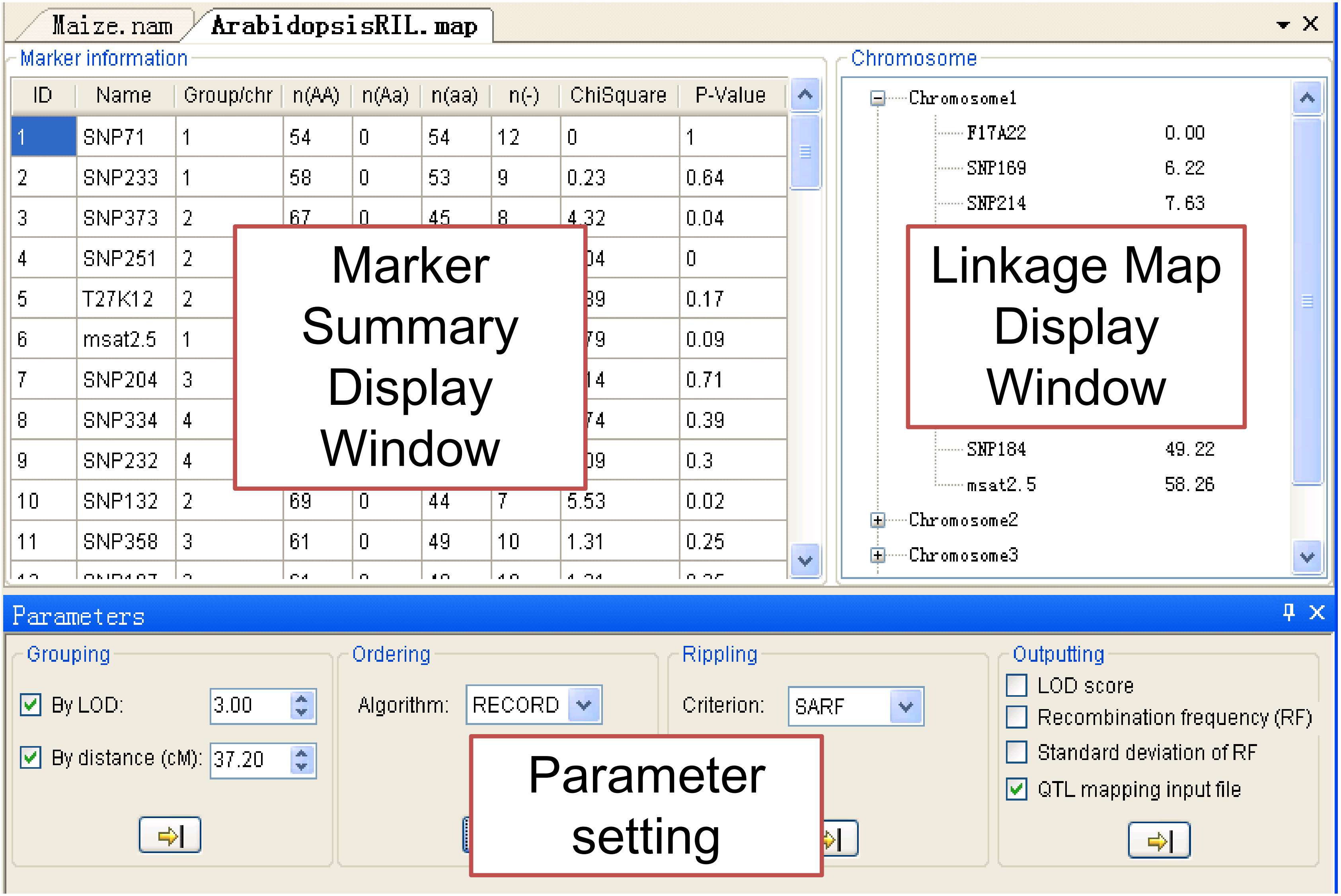
	Two parental lines P1 (MM), P2 (mm) and their F1 hybrid (Mm)			Populations with heterozygousity, i.e., 1, 2, 7, 8, 9, 10, 13, 14, 15, 16			Populations with no heterozygousity, i.e., 3, 4, 5, 6, 11, 12, 17, 18, 19, 20	
	P1: MM	P2: mm	F1: Mm	MM	mm	Mm	MM	mm
		▬	▬		▬	▬		▬
Code	2	0	1	2	10	10	2	0



	Two parental lines P1 (MM), P2 (mm) and their F1 hybrid (Mm)			Populations with heterozygousity, i.e., 1, 2, 7, 8, 9, 10, 13, 14, 15, 16			Populations with no heterozygousity, i.e., 3, 4, 5, 6, 11, 12, 17, 18, 19, 20	
	P1: MM	P2: mm	F1: Mm	MM	mm	Mm	MM	mm
	▪▪▪▪	▪▪▪▪	▪▪▪▪	▪▪▪▪	▪▪▪▪	▪▪▪▪	▪▪▪▪	▪▪▪▪
		▬	▬		▬	▬		▬
Code	2	0	1	2	10	10	2	0

When heterozygote F1 is present, P1 is coded as 2, and F1 and P2 are coded as 10.
When heterozygote F1 is absent, P1 is coded as 2, and P2 is coded as 0.

Interface of the MAP functionality



Marker information

ID	Name	Group/chr	n(AA)	n(Aa)	n(aa)	n(-)	ChiSquare	P-Value
1	SNP71	1	54	0	54	12	0	1
2	SNP233	1	58	0	53	9	0.23	0.64
3	SNP373	2	67	0	45	8	4.32	0.04
4	SNP251	2						0
5	T27K12	2						0.17
6	msat2.5	1						0.09
7	SNP204	3						0.71
8	SNP334	4						0.39
9	SNP232	4						0.3
10	SNP132	2	69	0	44	7	5.53	0.02
11	SNP358	3	61	0	49	10	1.31	0.25

Marker Summary Display Window

Chromosome

- Chromosome1
  - F17A22 0.00
  - SNP169 6.22
  - SNP214 7.63
  - SNP184 49.22
  - msat2.5 58.26
- Chromosome2
- Chromosome3

Linkage Map Display Window

Grouping

☑ By LOD: 3.00

☑ By distance (cM): 37.20

Ordering

Algorithm: RECORD

Rippling

Criterion: SARF

Outputting

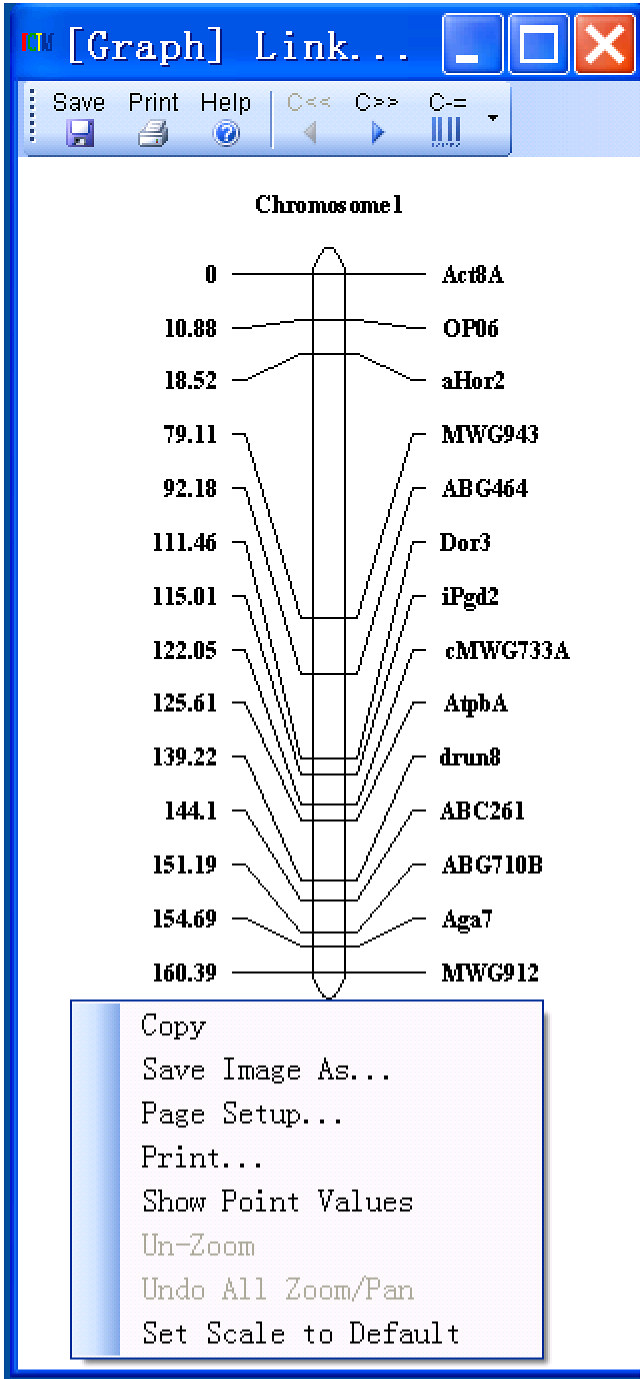
☐ LOD score
☐ Recombination frequency (RF)
☐ Standard deviation of RF
☑ QTL mapping input file

Parameter setting

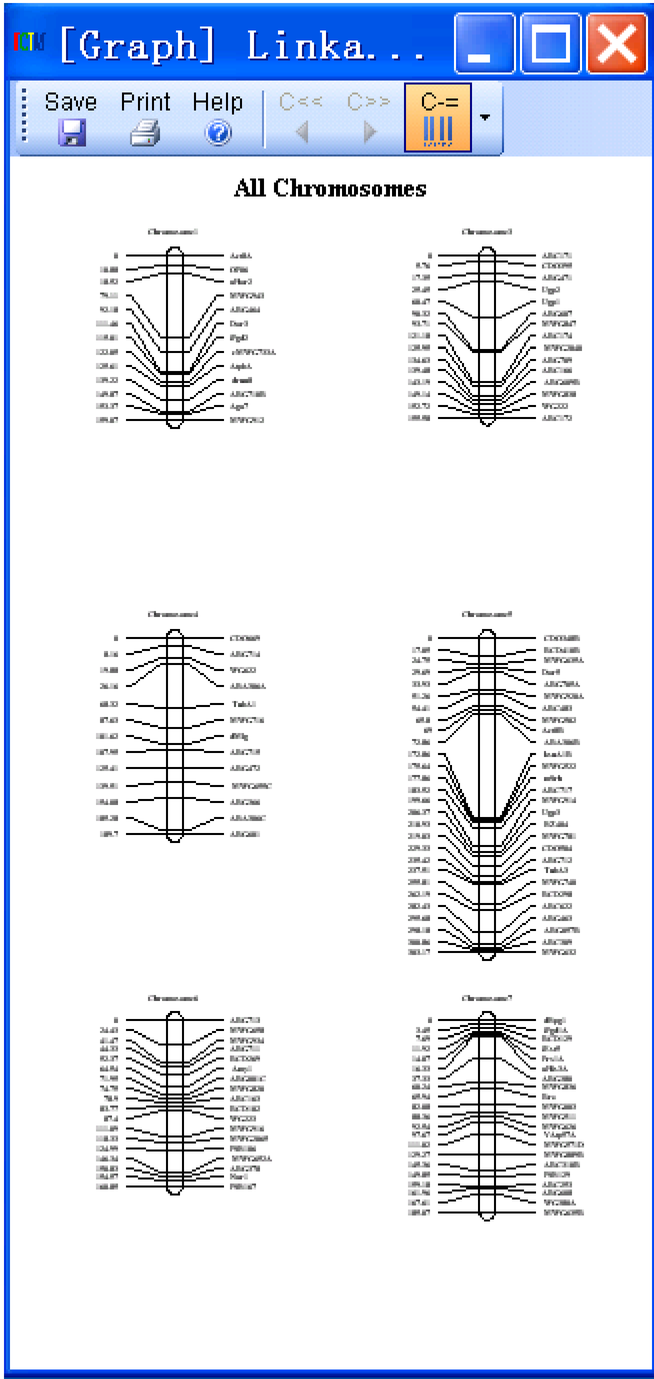
Map outputs:

Linkage map for each chromosome (A) or all chromosomes (B)

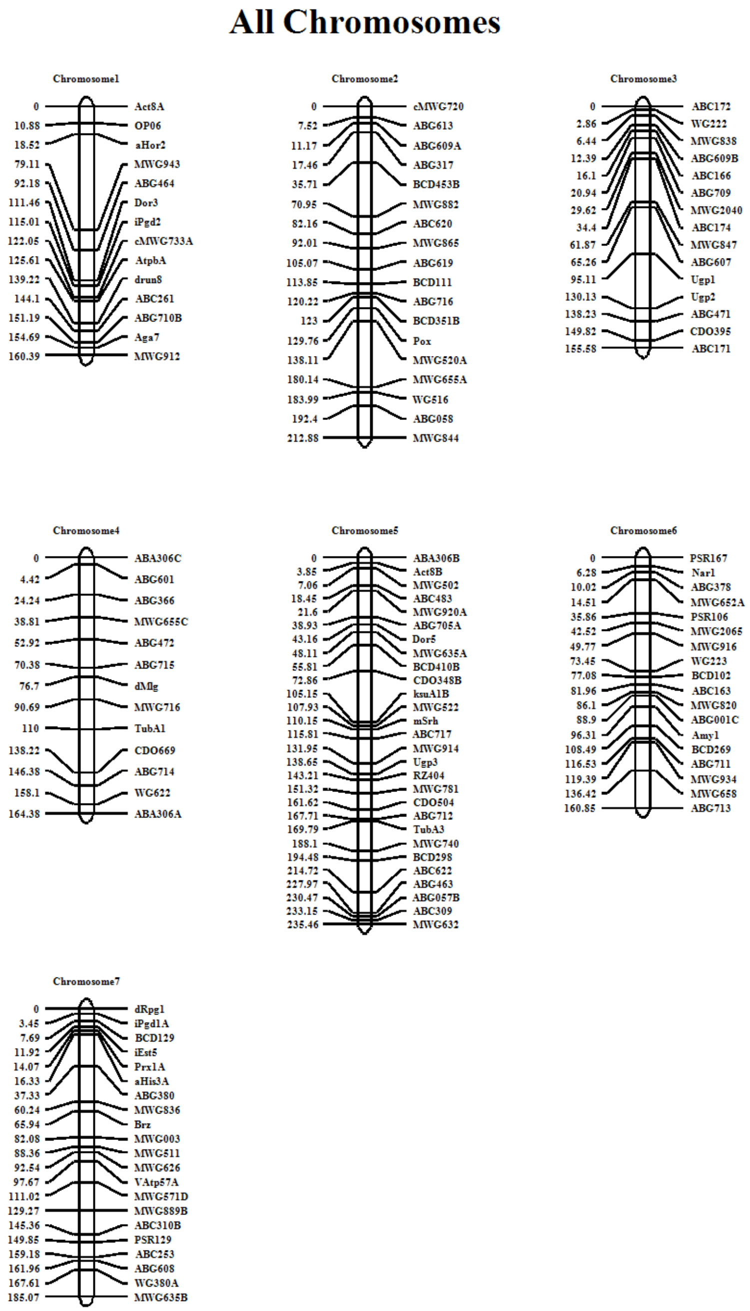
A. Map of one chromosome



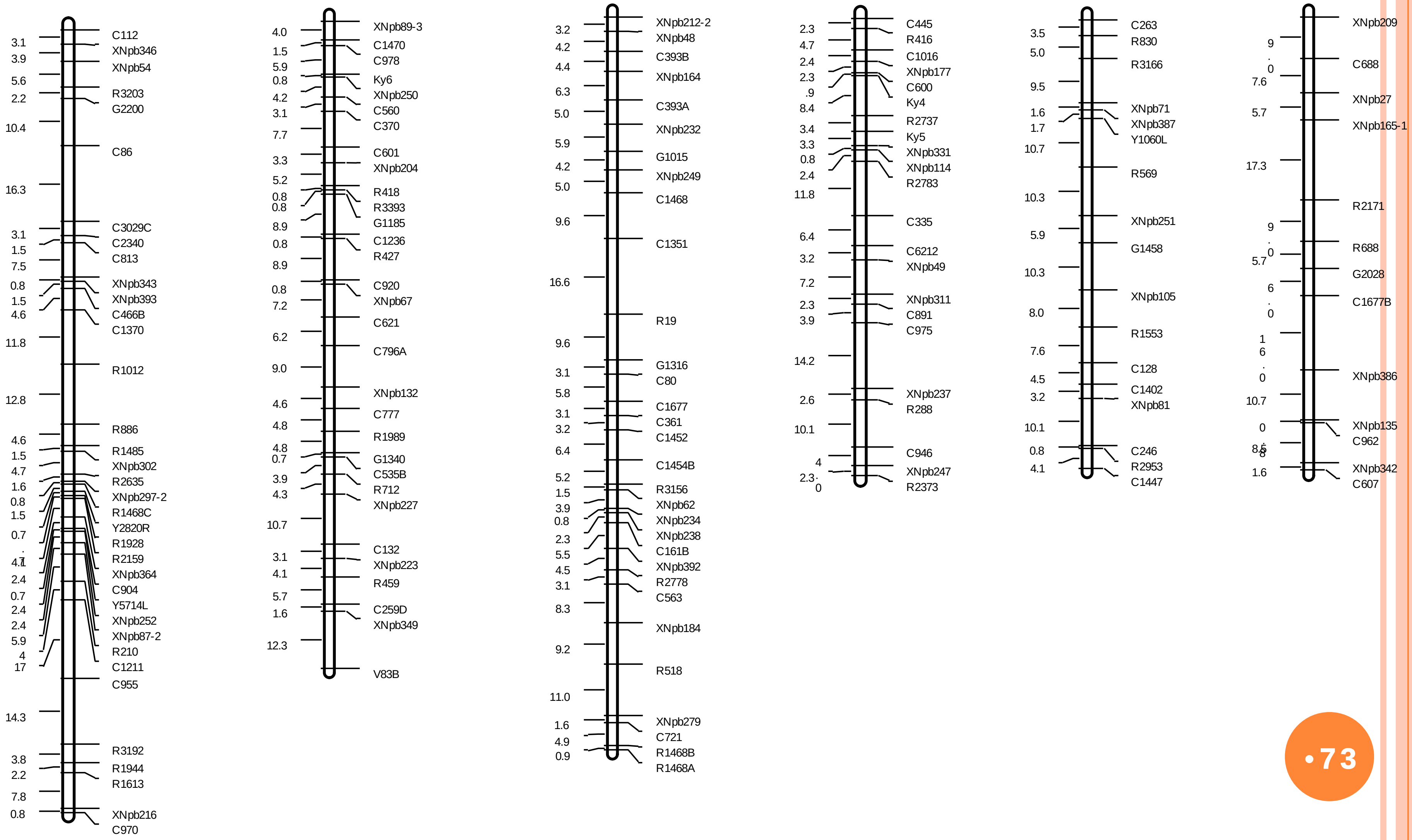
B. Map of all chromosomes



An example map of seven chromosomes or groups



构建连锁图谱


3.1
3.9
5.6
2.2
10.4
16.3
3.1
1.5
7.5
0.8
1.5
4.6
11.8
12.8
4.6
1.5
4.7
1.6
0.8
1.5
0.7
4.1
2.4
0.7
2.4
2.4
5.9
4
17
14.3
3.8
2.2
7.8
0.8
C112
XNpb346
XNpb54
R3203
G2200
C86
C3029C
C2340
C813
XNpb343
XNpb393
C466B
C1370
R1012
R886
R1485
XNpb302
R2635
XNpb297-2
R1468C
Y2820R
R1928
R2159
XNpb364
C904
Y5714L
XNpb252
XNpb87-2
R210
C1211
C955
R3192
R1944
R1613
XNpb216
C970
4.0
1.5
5.9
0.8
4.2
3.1
7.7
3.3
5.2
0.8
0.8
8.9
0.8
8.9
0.8
7.2
6.2
9.0
4.6
4.8
4.8
0.7
3.9
4.3
10.7
3.1
4.1
5.7
1.6
12.3
XNpb89-3
C1470
C978
Ky6
XNpb250
C560
C370
C601
XNpb204
R418
R3393
G1185
C1236
R427
C920
XNpb67
C621
C796A
XNpb132
C777
R1989
G1340
C535B
R712
XNpb227
C132
XNpb223
R459
C259D
XNpb349
V83B
3.2
4.2
4.4
6.3
5.0
5.9
4.2
5.0
9.6
16.6
9.6
3.1
5.8
3.1
3.2
6.4
5.2
1.5
3.9
0.8
2.3
5.5
4.5
3.1
8.3
9.2
11.0
1.6
4.9
0.9
XNpb212-2
XNpb48
C393B
XNpb164
C393A
XNpb232
G1015
XNpb249
C1468
C1351
R19
G1316
C80
C1677
C361
C1452
C1454B
R3156
XNpb62
XNpb234
XNpb238
C161B
XNpb392
R2778
C563
XNpb184
R518
XNpb279
C721
R1468B
R1468A
2.3
4.7
2.4
2.3
.9
8.4
3.4
3.3
0.8
2.4
11.8
6.4
3.2
7.2
2.3
3.9
14.2
2.6
10.1
4
2.3
0
C445
R416
C1016
XNpb177
C600
Ky4
R2737
Ky5
XNpb331
XNpb114
R2783
C335
C6212
XNpb49
XNpb311
C891
C975
XNpb237
R288
C946
XNpb247
R2373
3.5
5.0
9.5
1.6
1.7
10.7
10.3
5.9
10.3
8.0
7.6
4.5
3.2
10.1
0.8
4.1
C263
R830
R3166
XNpb71
XNpb387
Y1060L
R569
XNpb251
G1458
XNpb105
R1553
C128
C1402
XNpb81
C246
R2953
C1447
9.0
7.6
5.7
17.3
9.0
5.7
6.0
16.0
10.7
0
8.5
8
1.6
XNpb209
C688
XNpb27
XNpb165-1
R2171
R688
G2028
C1677B
XNpb386
XNpb135
C962
XNpb342
C607

不同物种的遗传图距和物理图距间的关系

物种	单倍体基因组大小（kb）	遗传图谱的长度（cM）	碱基对（kb)/cM
酵母（Yeast）	2.2×10^4	3700	6
Neurospora	4.2×10^4	500	80
Arabidopsis	7.0×10^4	500	140
Drosophila	2.0×10^5	290	700
西红柿（Tomato）	7.2×10^5	1400	510
人类（Human）	3.0×10^6	2710	1110
小麦（Wheat）	1.6×10^7	2575	6214
水稻（Rice）	4.4×10^5	1575	279
玉米（Corn）	3.0×10^6	1400	2140

课堂练习：EM算法

Type	A	B	AB	O	Total
N_i	2162	738	228	2876	6004

- 上表是某人群体中4种血型的调查数据。在HWE假定下，计算三种等位基因的频率
- 假定三个等位基因用A、B、O表示，
  - 血型A=基因型AA+基因型AO
  - 血型B=基因型BB+基因型BO
  - 血型AB=基因型AB
  - 血型O=基因型OO

课堂练习：EM算法

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

- **某座位上等位基因为A和a，另一座位上等位基因为B和b。亲本AABB与aabb杂交产生的F2群体中，这两个座位上6种可识别基因型的调查数据如上表。试计算两个座位间的重组率**
- **AAB_ = AABB + AABb**
- **AaB_ = AaBB + AB/ab + Ab/ab**
- **aaB_ = aaBB + aaBb**

实习6 QTL IciMapping软件 (I)

- 绘制遗传连锁图谱（MapShow）
- 两个位点间重组率的估计（2pointREC）
- 表型数据方差分析 （ANOVA）
- 连锁图谱构建功能（MAP）
  - Grouping：分群、设定Anchor信息
  - Ordering：标记排序
  - Rippling：图谱调整
  - 连锁图绘制
  - 输入和输出