QTL mapping 的流程

試 驗 設 計

(Experimental Design)

⇓

標識生產

(Marker Generation)

⇓

連鎖圖建立

(Linkage Map Construction)

⇓

QTL 估計

(QTL Estimation)

⇓

QTL 使用

(QTL Manipulation)

與數量性狀基因座定位(QTL mapping)相關的一些遺傳名詞

- 質性狀（qualitative trait)：豌豆的花色，人類的血型、cystic fibrosis, Huntington disease……。
- 數量性狀（quantitative trait)：人類的身高、體重、高血壓、糖尿病，水稻的株高、產量及對疾病的控病程度，牛乳的產量……。
- 主基因（major gene）
- 微效基因（polygenes）
- 基因座（locus, loci）
- 等位基因（allele）
- 數量性狀基因座（QTL；quantitative trait loci）
- 遺傳率（heritability）
- 上位性（epistasis，即交感作用）

- 基因型與外表型（genotype & phenotype）
- 基因型值與外表型值（genotypic value & phenotypic value）
- 連鎖（linkage）
- 交換與重組（crossing over & recombination）
- 重組體與重組比例（recombinant & recombination fraction）
- 物理距離和遺傳距離（physical & genetic distances）
- 物理圖譜和遺傳圖譜（physical & genetic map）
- 染色體與連鎖群（chromosomes & linkage groups）
- 分子遺傳標識（molecular genetic marker）：如 RFLP, RAPD, VNTR, SSR, SNP 等。

試驗族群的遺傳結構

● 回交和 F_2 族群的產生流程和遺傳結構

P_1 × P_2

$\frac{mn}{mn}$ $\frac{MN}{MN}$

F_1

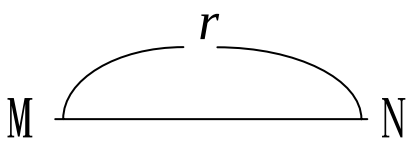
$\frac{MN}{mn}$

	BC_1				BC_2			
基因型	$\frac{mn}{MN}$	$\frac{mn}{Mn}$	$\frac{mn}{mN}$	$\frac{mn}{mn}$	$\frac{MN}{MN}$	$\frac{MN}{Mn}$	$\frac{MN}{mN}$	$\frac{MN}{mn}$
代號	11	10	01	00	22	21	12	11
期望比例	$\frac{1-r}{2}$	$\frac{r}{2}$	$\frac{r}{2}$	$\frac{1-r}{2}$	$\frac{1-r}{2}$	$\frac{r}{2}$	$\frac{r}{2}$	$\frac{1-r}{2}$

F_2

基因型	$\frac{MN}{MN}$	$\frac{MN}{Mn}$	$\frac{Mn}{Mn}$	$\frac{MN}{mN}$	$\frac{MN}{mn}$	$\frac{Mn}{mN}$	$\frac{Mn}{mn}$	$\frac{mN}{mN}$	$\frac{mN}{mn}$	$\frac{mn}{mn}$
代號	22	21	20	12	11	11	10	02	01	00
期望比例	$\frac{(1-r)^2}{4}$	$\frac{r(1-r)}{2}$	$\frac{r^2}{4}$	$\frac{r(1-r)}{2}$	$\frac{(1-r)^2}{2}$	$\frac{r^2}{2}$	$\frac{r(1-r)}{2}$	$\frac{r^2}{4}$	$\frac{r(1-r)}{2}$	$\frac{(1-r)^2}{4}$

● $\frac{MN}{mn}$個體的配子基因型分布



r：重組比例

$0 < r < 0.5$

$\frac{MN}{mn}$個體所產生的四種不同配子基因型分布如下：

recombinant

配子基因型	MN	Mn	mN	mn
期望比例	$\frac{1-r}{2}$	$\frac{r}{2}$	$\frac{r}{2}$	$\frac{1-r}{2}$

$\frac{MN}{mn}$個體自交或互交後之子代基因型分布為

	MN	Mn	mN	mn
MN	$\frac{(1-r)^2}{4}$	$\frac{r(1-r)}{4}$	$\frac{r(1-r)}{4}$	$\frac{(1-r)^2}{4}$
Mn	$\frac{r(1-r)}{4}$	$\frac{r^2}{4}$	$\frac{r^2}{4}$	$\frac{r(1-r)}{4}$
mN	$\frac{r(1-r)}{4}$	$\frac{r^2}{4}$	$\frac{r^2}{4}$	$\frac{r(1-r)}{4}$
mn	$\frac{(1-r)^2}{4}$	$\frac{r(1-r)}{4}$	$\frac{r(1-r)}{4}$	$\frac{(1-r)^2}{4}$

● 回交和 F_2 族群的特點：

1. 各基因座內的兩種不同等位基因的分離比為 1：1。
2. 一基因座內之不同基因型的分離比在對回交族群中為 1：1（$MM:Mm$ 或 $mm:Mm$），而在 F_2 族群中為 1：2：1（$MM:Mm:mm$）。
3. 當同時考慮 m 個基因座時，回交族群有 2^m 種不同基因型，而 F_2 族群有 3^m 種不同基因型。

● 偽試交（pseudo testcross）族群

$$P_1 \quad \times \quad P_2$$

$$\frac{A_1}{A_2} \qquad \downarrow \qquad \frac{A_3}{A_4}$$

基 因 型	$\frac{A_1}{A_3}$	$\frac{A_1}{A_4}$	$\frac{A_2}{A_3}$	$\frac{A_2}{A_4}$
期望比例	$1/4$	$1/4$	$1/4$	$1/4$

帶有 A_1 等位基因的子代比例為 $1/2$

遺傳連鎖圖的建立

- 連鎖圖的建立所遵循的原則雖然相當簡單，但其中所牽涉到的數學、統計問題卻至為廣泛，包括假設檢定(hypothesis testing)、介量估計(parameter estimation)，數值最佳化(numerical optimization)、機率論(probability)和組合學(combinatorics)等問題(Week 1991)。

- 重組現象的發生與基因間的距離有關。一般而言，距離越近，重組機會（頻度）越低；距離越遠，重組機會（頻度）越高。

- 重組：染色體在減數分裂中的四分子期若發生交換，染色體（基因）會有重組的現象，而產生重組型的配子。

- 基因重組頻度的高低與它們之間距離的遠近並無簡單關係。

- 數量遺傳學家利用遺傳圖譜函數（genetic map function）令定義遺傳距離，使重組頻度與(遺傳)距離有一簡單的函數關係。

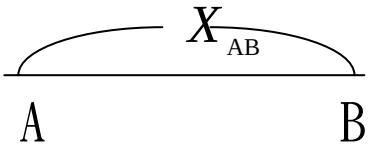
- 遺傳圖譜函數（genetic map function)：將易測量但不具有累加性的交換比例r轉換成不易測量、但具有累加性的遺傳距離X的函數，稱為遺傳圖譜函數。

$$M\left(X_{\mathrm{AB}}\right)= r_{\mathrm{AB}} \quad \text{or} \quad M^{-1}\left(r_{\mathrm{AB}}\right)= X_{\mathrm{AB}}$$

- Haldane map function （Haldane 1919）

Morgan（M）：一個世代一個配子發生一次交換的距離稱為一個 Morgan。

1M = 100cM



假設基因 A 和 B 相距 X_{AB} M，且這段染色體可被切為 N 個等長的小片段。如果這 N 個小片段內之交換彼此不互相干擾（互相獨立），則這 N 個片段中正好有 t 個片段發生交換的機率為

$$P(T=t)=\frac{N!}{t!(N-t)!}\left(\frac{X_{AB}}{N}\right)^{t}\left(1-\frac{X_{AB}}{N}\right)^{N-t}$$

當 $N\to\infty$，$P(T=t)\approx\dfrac{X_{AB}^{t}\,e^{-X_{AB}}}{t!}$

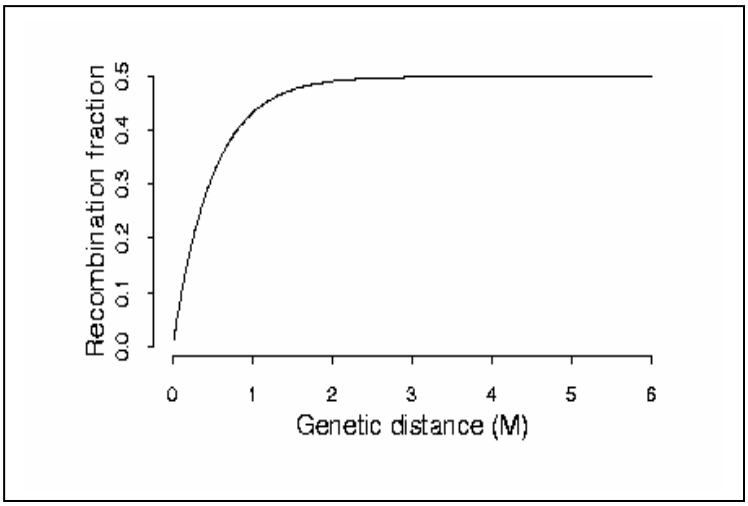
則 A 與 B 發生重組的機率為

$$r_{AB} = \sum_{t\text{爲奇數}} P(T=t) = \frac{1}{2}\left(1-e^{-2X_{AB}}\right)$$

或是 $X_{AB} = -\dfrac{1}{2}\ln\left(1-2r_{AB}\right)$

● Haldane map function 的特性

1.



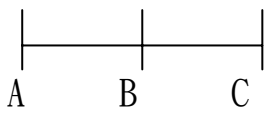
$0 < r < 0.5$

X =1 cM, r=0.0099

X =70cM, r=0.37

$X = \infty$, r=0.5

2.



$$X_{\mathrm{AC}} = X_{\mathrm{AB}} + X_{\mathrm{BC}}$$

$$(1-2r_{AC}) = (1-2r_{AB})(1-2r_{BC})$$

或 $r_{AC} = r_{AB} + r_{BC} - 2r_{AB}r_{BC}$

或 $r_{AC} = r_{AB}(1-r_{BC}) + (1-r_{AB})r_{BC}$

- Kosambi（1944）在允許干擾(非獨立)的假設下推導出另一 map function

$$X_{\mathrm{AB}} = \frac{1}{4}\ln\left(\frac{1+2r_{\mathrm{AB}}}{1-2r_{\mathrm{AB}}}\right)$$

- 人類染色體組約有 33M，或是 3.3×10^{6}Kb。一般認為，人類之染色體組中 1cM≈1000 Kb。不同的生物有不同的 Kb/cM 比，且一同種生物之不同染色體區域之 Kb/cM 比也不同。

● 兩基因座間之重組比例r的估計及檢驗

$$r=\frac{\text{重組體數}}{\text{重組體數}+\text{非重組體數}}$$

	回交族群子代				
基因型	$\frac{AB}{AB}$	$\frac{AB}{Ab}$	$\frac{AB}{aB}$	$\frac{AB}{ab}$	總數
代號	22	21	12	11	
觀測數	a	b	c	d	n
期望數	$\frac{n}{2}(1-r_{\mathrm{AB}})$	$\frac{n}{2}r_{\mathrm{AB}}$	$\frac{n}{2}r_{\mathrm{AB}}$	$\frac{n}{2}(1-r_{\mathrm{AB}})$	n

$$L(r_{\mathrm{AB}})\propto(1-r_{\mathrm{AB}})^{a+d}(r_{\mathrm{AB}})^{b+c}$$

$$S_{r_{\mathrm{AB}}}=-\frac{a+d}{1-r_{\mathrm{AB}}}+\frac{b+c}{r_{\mathrm{AB}}}$$

$$I_{r_{\mathrm{AB}}}=\frac{n}{r_{\mathrm{AB}}(1-r_{\mathrm{AB}})}$$

得

$$\hat{r}_{\mathrm{AB}}=\frac{b+c}{n},\quad Var(\hat{r}_{\mathrm{AB}})=\frac{r_{\mathrm{AB}}(1-r_{\mathrm{AB}})}{n}$$

實際上，$\hat{r}_{\mathrm{AB}} = \min\left(\frac{b+c}{n}, 0.5\right)$

是否連鎖？

$$H_0 : r_{\mathrm{AB}} = 0.5$$

$$H_1 : r_{\mathrm{AB}} \le 0.5$$

$$X^2 = -2\log\frac{L(0.5)}{L(\hat{r})} \sim \chi_1^2$$

若 $X^2 > \chi_{1,\alpha}^2$，則推翻 H_0： $r_{\mathrm{AB}} = 0.5$

● 例子：a=53, b=13, c=9, d=55。

$$\hat{r}_{\mathrm{AB}} = \min\left(\frac{13+9}{130}, 0.5\right) = 0.1692$$

$$Var(\hat{r}_{\mathrm{AB}}) = 1.08 \times 10^{-3}$$

$$X^2 = -2\log\frac{L(0.5)}{L(0.1692)} \approx 627 > \chi_{1,0.05}^2$$

得A與B連鎖的結論

- 建立遺傳連鎖圖的步驟

  先檢查（所有）標識的分離比是否正常，再對標識分（連鎖）群，最後對屬於同一群的標識進行排序。

- 檢查分離比

  測驗 H_0： $P = 0.5$ 以確定分離比是否為 1：1（回交族群）。分離比偏離 1：1 的標識一般不再被進一步用來分群及排序。

- 分（連鎖）群

  任意兩標識（M_i和M_j）間之重組比例r_{ij}若小於0.5，則歸屬為同一（連鎖）群。經由對每一對標識測驗重組比例r_{ij}是否為0.5

$$H_0: \ r_{ij} = 0.5$$

  決定那些標識屬於同一群。分群標準越嚴格（α越小），群數越多。標準越寬鬆（α越大），群數越少。

- 群內標識排序（marker ordering）

  確定了屬於同一群的標識後，必需進一步對各群內之標識進行排序。若群內有 N 個標識，則共有 $N!/2$ 種可能的排列次序(10!/2 等於 1,814,400)。在 $N!/2$ 種次序中，最符合某種標準（例如 likelihood 最大，或 SAR 最小）的次序為最佳次序（best order）。然而，最佳次序並無保證必為真次序（true order）。

排序方法：

1. 多點分析法（multi-point analysis）

		重 組 事 件		
配子型	數目	A-B-C	A-C-B	B-A-C
ABC or abc	n_1	無	無	無
ABc or abC	n_2	B-C	A-C 和 C-B	A-C
Abc or aBC	n_3	A-B	A-C	B-A 和 A-C
AbC or aBc	n_4	A-B 和 B-C	C-B	B-A
總數	n			

次序 ABC 的 likelihood 為

$$L(\mathrm{ABC}) = [(1-r_{\mathrm{AB}})(1-r_{\mathrm{BC}})]^{n_1} \times [(1-r_{\mathrm{AB}})r_{\mathrm{BC}}]^{n_2} \times [r_{\mathrm{AB}}(1-r_{\mathrm{BC}})]^{n_3} \times [r_{\mathrm{AB}} \times r_{\mathrm{BC}}]^{n_4}$$

同理可得 $L(\mathrm{ACB})$ 和 $L(\mathrm{BAC})$。Likelihood 最大的次序為最佳次序。

- 當標識數 N 稍大時，多點分析法在實際的執行上有困難（為什麼？）。遺傳學家被迫利用一些替代性的方法來找最佳次序。

（一）SAL（sum of adjacent log-likelihoods）

$$\mathrm{SAL} = \sum_{i=1}^{N-1} \log[L_i(r)]$$

SAL 最大的次序為最佳次序。

(二)SAR(sum of adjacent recombination fraction)

1. 若 ABC 為正確的次序，則次序 ABC 的 SAR=$r_{AB}+r_{BC}$。
2. 依 Haldane map function r_{AC}，r_{AB}和r_{BC}有以下關係：

$$r_{AC}=r_{AB}(1-r_{BC})+(1-r_{AB})r_{BC}$$

3. 若另一次序 BAC 被誤認為是正確的次序，則此錯誤次序 BAC 的 SAR 為 $r_{AB}+r_{AC}$。
4. 正確次序的 SAR 恆比錯誤次的 SAR 小：

$$\begin{aligned}&\mathrm{SAR}_{BAC}=r_{AB}+r_{AC}=\\&r_{AB}+r_{AB}(1-r_{BC})+(1-r_{AB})r_{BC}\\&=r_{AB}+r_{BC}+r_{AB}(1-2r_{BC})\\&\geq r_{AB}+r_{BC}=\mathrm{SAR}_{ABC}\end{aligned}$$

故在所有次序中，SAR最小的次序為最佳次序（Weir 1996）。

5. 正確次序的 SAR 最小的結論可適用於任意多個標識的序列。

6. 利用 SAL 或 SAR 找最佳次序有兩個問題：

a. 取樣誤差使得 SAR 最小的次序並無保證必為真次序。

b. 有太多的次序。

(三)Seriation(Buetow & Chakravarti 1987)

例：標識 A、B、C 和 D 間有重組比例如下：

$$\begin{array}{c} \\ A \\ B \\ C \\ D \end{array} \begin{array}{c} \begin{array}{cccc} A & B & C & D \end{array} \\ \begin{bmatrix} 0 & 0.09 & 0.19 & 0.17 \\ & 0 & 0.26 & 0.22 \\ & & 0 & 0.32 \\ & & & 0 \end{bmatrix} \end{array}$$

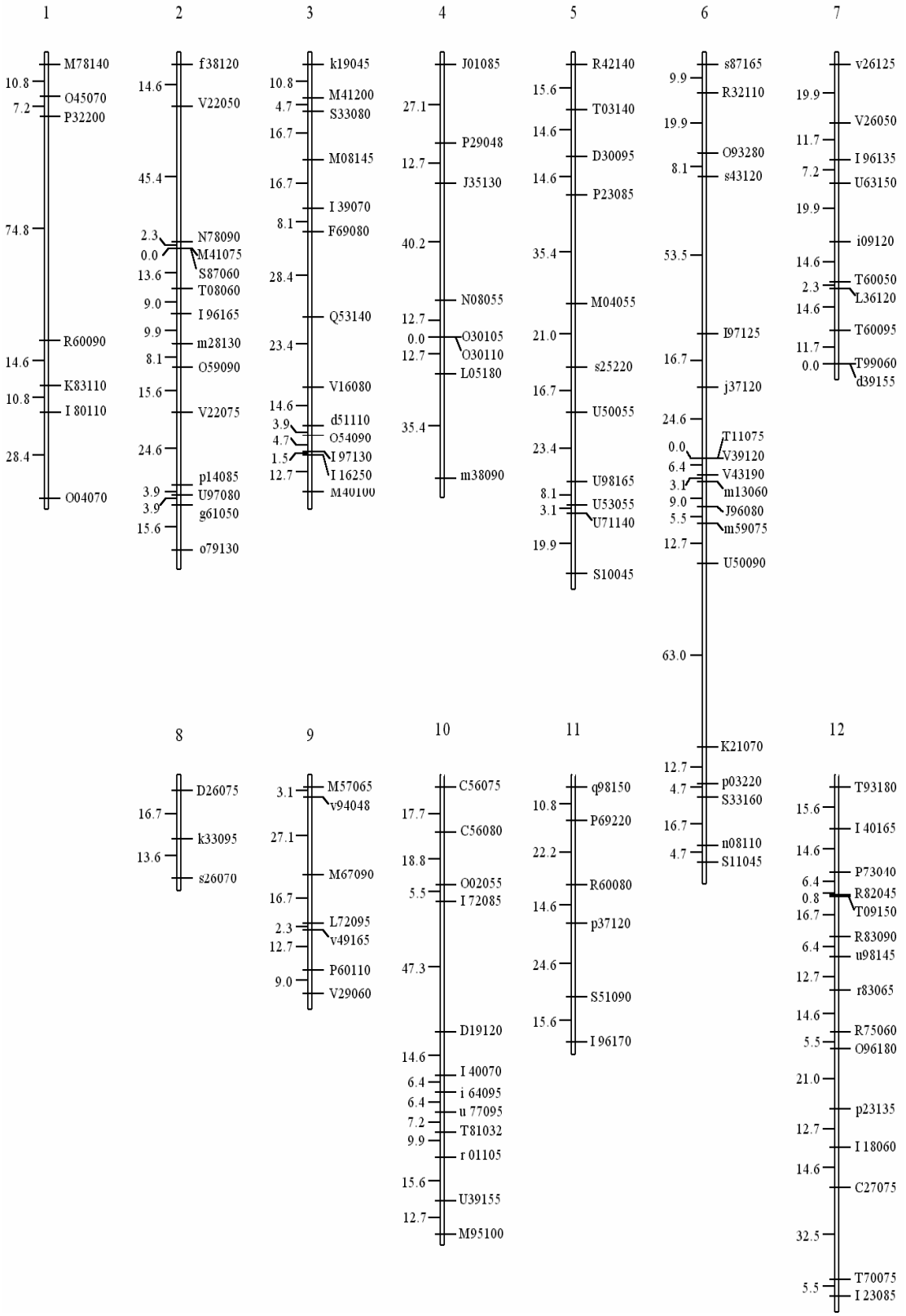
得次序 C-B-A-D，其相對應的 SAR 為 0.52。在所有 4!/2 種次序中，SAR 最小的次序為 DBAC(SAR=0.50)。

（四）RCD (Rapid chain delineation , Doerge1996)

（五）Simulated annealing (Kirkpatrick et al.1983)

（六）Branch and bound (Thompson 1987)

● Radiata pine 十二對染色體遺傳連鎖圖



參考文獻

高振宏 (2000)對試驗族群進行數量性狀基因座定位的統計方法評介。中國統計學報第 38 卷第一期，1-36

Buetow and Chakravarti (1987a), Multipoint gene mapping using seriation I. General methods. Am. J. Hum. Genet. 41:180-188.

Buetow and Chakravarti (1987b), Multipoint gene mapping using seriation II. Analysis of simulated and empirical data. Am. J. Hum. Genet. 41:189-201.

Doerge, R. W. (1996), Construction genetic maps by rapid chain delineation. Journal Of Quantitative Trait Loci 2, Article 6. http://probe.nalusda.gov:8000/otherdocs/jqtl.

Haldane, J.B.S (1919), The combination of linkage values and the calculation of distances between the loci of linked factors. J. Genet. 8: 299-309.

Kirkpatrick, S., C. D. Gelatt, Jr., and M. P. Vecchi (1983), Optimization by simulated annealing. Science 220: 671-680.

Lander, E. S. and D. Botstein (1989), Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121: 185-199.

Thompson, E. A. (1987), Crossover counts and likelihood in multipoint linkage analysis. IMA J of mathematics applied in medicine & biology 4: 93-108.

Weeks, D. E. (1991), Human linkage analysis: strategies for locus ordering.Pp. 297-330 in Advanced Techniques in Chromosome Research. K. W. Adolph (Editor) Dekker, New York.

Weir, B. S. (1996), Genetic Data Analysis II. Sinauer Associates, Inc. Sunderland, Massachusetts, USA.