

两个基因座位的遗传平衡原理

潘沈元, 屈艾, 彭会, 李爱玲

(徐州师范大学生命科学学院, 江苏徐州 221116)

摘要: 由于许多教科书中关于连锁平衡的介绍, 多是引用结论或是推导不太严谨, 使学生在群体遗传学中对理解连锁平衡的原理感到困难。此文从遗传平衡的基本条件出发, 通过较严谨的数学推导, 介绍了两个基因座的连锁平衡条件、平衡过程等原理, 供教师和学生群体遗传学教学中参考。

关键词: 群体遗传学; 教学; 遗传平衡; 连锁平衡

中图分类号: Q347

文献标识码: A

文章编号: 0253-9772(2004)02-0215-04

Principle of Genetic Equilibrium for Two Gene Loci

PAN Shen-Yuan, QU Ai, PENG Hui, LI Ai-Ling

(College of Life Sciences, Xuzhou Normal University, Jiangsu Xuzhou 221116, China)

Abstract: Because linkage equilibrium is introduced by directly quoting the conclusions or imprecise mathematical reasoning in most of textbooks, many students are puzzled with the problem of linkage equilibrium when they learn population genetics. Based on the radical conditions of genetic equilibrium, the principle of linkage equilibrium condition and process, for two gene loci is introduced by precise mathematical reasoning. The article may provide reference to teachers and students in the teaching and learning of population genetics.

Key words: population genetics; teaching; genetic equilibrium; linkage equilibrium

两个基因座的遗传平衡又叫做连锁平衡, 在群体遗传学书籍中均有介绍。由于某些教科书中关于连锁平衡问题的介绍, 多是引用结论或是推导不太严谨, 使学生在群体遗传学中对理解连锁平衡问题的原理感到困难。目前国内关于群体遗传学的介绍多是关于一对基因的遗传平衡^[1], 而关于两个基因座的连锁平衡介绍的较少^[2]。作者根据自己的理解和教学经验, 并参考其他教科书^[3~6], 从遗传平衡的基本条件出发, 通过较严谨的数学推导, 介绍了两个基因座连锁平衡的平衡条件、平衡过程的基本原理, 供教师和学生群体遗传学教学中参考。

1 连锁平衡条件

对一给定群体, 在雌雄性基因型比例相同的条件下, 若分别考虑常染色体上的每个基因座, 随机交配一代后都可以达到平衡, 但是, 我们若同时考虑两个或多个基因座时, 则随机交配一代后群体未必能达到平衡, 因为上下代的基因型频

率在变化。不管两个或多个基因是否连锁, 两个或多个基因座间基因型频率的不平衡状态, 似乎是由于基因座间的连锁关系引起的, 所以称这种基因型频率的不平衡状态为连锁不平衡(linkage disequilibrium)。又由于在随机交配的情况下, 群体内的基因型频率是由配子型频率所决定, 导致连锁不平衡的原因是由于群体产生的配子没有达到预期的比例, 似乎存在连锁关系, 所以这种不平衡也叫做配子相不平衡(gametic phase disequilibrium)。

1.1 基本假定

基因频率和基因型频率在世代间保持不变, 是判断群体是否达到平衡状态的基本标准。根据 Hardy-Wenber 定律, 在群体雌雄性比例相同、大的、随机交配和没有其他因素干扰的群体中, 每个基因座上的基因频率将保持不变, 所以在此只要考察上下代间基因型频率是否变化。

现在考虑两对常染色体上的基因, 第一基因座上的等位基因是 A_1 、 A_2 , 第二基因座上的等位基因是 B_1 、 B_2 , 这两个

收稿日期: 2003-02-12; 修回日期: 2003-07-11

作者简介: 潘沈元(1957-), 男, 江苏宜兴人, 教授, 研究方向: 数量及群体遗传学, 生物统计学。Tel: 0516-3403170, E-mail: PanShenyuan@

基因座可以连锁,也可以不连锁,其重组频率表示为 c , 相应的基因频率分别表示为:

$$p_1 = P(A_1), p_2 = P(A_2), q_1 = P(B_1), q_2 = P(B_2),$$

基因型 $\frac{A_i B_j}{A_k B_l}$ 仅表示 $A_i B_j$ 是来自父方的配子, $A_k B_l$ 是来自母方的配子,并不一定是连锁关系。相应的配子基因型频率表示为

$$P_{ij} = P(A_i B_j), \quad i, j = 1, 2$$

由配子基因型频率可以求出某一基因座上的基因频率

$$p_i = P(A_i B_1) + P(A_i B_2) = P_{i1} + P_{i2} = \sum_{j=1}^2 P_{ij} \quad (1)$$

$$q_i = P(A_1 B_i) + P(A_2 B_i) = P_{1i} + P_{2i} = \sum_{j=1}^2 P_{ij} \quad (2)$$

即 $p_1 = P_{11} + P_{12}, p_2 = P_{21} + P_{22}$

$$q_1 = P_{11} + P_{21}, q_2 = P_{12} + P_{22}$$

设原初群体,其基因型频率在两性中相同,原初群体配子基因型频率表示为 $P_{ij}^{(0)}$,从原初群体开始随机交配(若基因型频率在两性中不相同,随机交配一代后即成为相同,此时定为原初群体)。

1.2 推论 1

要判断第 n 个世代是否处于平衡状态,即基因型频率在上下代之间保持不变,等价于判断第 n 个世代产生各种配子基因型的频率是否等于第 $n-1$ 个世代相应配子基因型的频率,即

$$P_{ij}^{(n)} = P_{ij}^{(n-1)}, \quad i, j = 1, 2 \quad (3)$$

因为在随机交配情况下,第 n 个世代基因型频率是由亲本配子基因型频率决定的,如果亲本配子基因型频率不发生变化,子代基因型频率也不发生变化。

1.3 世代间配子频率的变化

为分析第 n 代群体的遗传组成情况,将第 $n-1$ 代配子基因型及频率和第 n 代群体的遗传组成列于表 1。

表 1 第 $n-1$ 代配子基因型及频率
和第 n 代群体的遗传组成

Table 1 Gametes and its frequencies in generations $n-1$
and genotypes in generations n

$n-1$ 代配子 Gametes (概率 Frequencies)	$A_1 B_1$ (P_{11})	$A_1 B_2$ (P_{12})	$A_2 B_1$ (P_{21})	$A_2 B_2$ (P_{22})
$A_1 B_1$ (P_{11})	$\frac{A_1 B_1}{A_1 B_1}$	$\frac{A_1 B_2}{A_1 B_1}$	$\frac{A_2 B_1}{A_1 B_1}$	$\frac{A_2 B_2}{A_1 B_1}$
$A_1 B_2$ (P_{12})	$\frac{A_1 B_1}{A_1 B_2}$	$\frac{A_1 B_2}{A_1 B_2}$	$\frac{A_2 B_1}{A_1 B_2}$	$\frac{A_2 B_2}{A_1 B_2}$
$A_2 B_1$ (P_{21})	$\frac{A_1 B_1}{A_2 B_1}$	$\frac{A_1 B_2}{A_2 B_1}$	$\frac{A_2 B_1}{A_2 B_1}$	$\frac{A_2 B_2}{A_2 B_1}$
$A_2 B_2$ (P_{22})	$\frac{A_1 B_1}{A_2 B_2}$	$\frac{A_1 B_2}{A_2 B_2}$	$\frac{A_2 B_1}{A_2 B_2}$	$\frac{A_2 B_2}{A_2 B_2}$

由表 1 可见,只有反对角线上的 4 种基因型通过重组才有可能产生不同的重组型配子,从而引起配子频率发生变化,它们分别是相引杂合子 $(\frac{A_1 B_1}{A_2 B_2}$ 和 $\frac{A_2 B_2}{A_1 B_1})$, 相斥杂合子 $(\frac{A_1 B_2}{A_2 B_1}$ 和 $\frac{A_2 B_1}{A_1 B_2})$, 其中,相引杂合子在群体中的频率是 $2P_{11}^{(n-1)} P_{22}^{(n-1)}$, 相斥杂合子在群体中的频率是 $2P_{12}^{(n-1)} P_{21}^{(n-1)}$, 这样,相引型杂合子因发生重组而使相引型配子减少(同时也是相斥型配子增多)的频率是

$$2P_{11}^{(n-1)} P_{22}^{(n-1)} \cdot c;$$

相斥型杂合子因发生重组而使相引型配子增多(同时也是相斥型配子减少)的频率是

$$2P_{12}^{(n-1)} P_{21}^{(n-1)} \cdot c$$

由于是对等交换,两种相引型配子 $A_1 B_1, A_2 B_2$ 的增加和减少量相等;两种相斥型配子 $A_1 B_2, A_2 B_1$ 的增加和减少量相等。所以,两个世代间配子频率的变化量可表示为:

$$P_{ij}^{(n)} - P_{ij}^{(n-1)} = \begin{cases} (P_{12}^{(n-1)} P_{21}^{(n-1)} - P_{11}^{(n-1)} P_{22}^{(n-1)}) \cdot c, & i = j \\ (P_{11}^{(n-1)} P_{22}^{(n-1)} - P_{12}^{(n-1)} P_{21}^{(n-1)}) \cdot c, & i \neq j \end{cases} \quad (4)$$

1.4 推论 2

不论 i 是否等于 j ,要使世代间配子频率不发生变化,即(3)式成立,只需(4)式右边等于 0,即

$$(P_{12}^{(n-1)} P_{21}^{(n-1)} - P_{11}^{(n-1)} P_{22}^{(n-1)}) \cdot c = 0 \quad (5)$$

或写成

$$P_{12}^{(n-1)} P_{21}^{(n-1)} \cdot c = P_{11}^{(n-1)} P_{22}^{(n-1)} \cdot c \quad (6)$$

也就是说,第 n 代群体连锁平衡的条件是:在第 $n-1$ 代群体中,由于重组而使 $n-1$ 代配子基因型 $A_i B_j$ 减少的频率等于由于重组而使 $A_i B_j$ 增加的频率。显然这是一种动态平衡。

1.5 推论 3

(1)当 $c=0$ 时,(6)式成立,即完全连锁的两个基因,原初群体只要随机交配一代后,群体就达到平衡。

(2)当 $c \neq 0$ 时,(6)式可简化为

$$P_{12}^{(n-1)} P_{21}^{(n-1)} = P_{11}^{(n-1)} P_{22}^{(n-1)} \quad (7)$$

即,当不完全连锁时,若群体中相引型杂合子频率等于相斥型杂合子频率,群体达到平衡。

2 离衡差

2.1 定义

$$D^{(n)} = P_{11}^{(n-1)} P_{22}^{(n-1)} - P_{12}^{(n-1)} P_{21}^{(n-1)} \quad (8)$$

为第 n 代的离衡差(difference from equilibrium),用以表示连锁不平衡程度的大小。它是第 n 代群体中相引型杂合子与相斥型杂合子频率差的一半。

2.2 推论 4

当 $D=0$ 时,群体达到平衡。

2.3 离衡差的基因频率表达方式

由(1)、(2)式可得:

$$P_{12} = p_1 - P_{11}$$

$$P_{21} = q_1 - P_{11}$$

$$P_{22} = 1 - p_1 - q_1 + P_{11}$$

于是:

$$P_{11}P_{22} = P_{11}(1 - p_1 - q_1 + P_{11}) = P_{11} - p_1P_{11} - q_1P_{11} + P_{11}^2$$

$$P_{12}P_{21} = (p_1 - P_{11})(q_1 - P_{11}) = p_1q_1 - p_1P_{11} - q_1P_{11} + P_{11}^2$$

代入(8)式得 $D^{(n)} = P_{11}^{(n-1)} - p_1q_1$

$$\text{同理可得} \begin{cases} D^{(n)} = P_{22}^{(n-1)} - p_2q_2 \\ D^{(n)} = -(P_{12}^{(n-1)} - p_1q_2) \\ D^{(n)} = -(P_{21}^{(n-1)} - p_2q_1) \end{cases}$$

以上 4 个式子可合并成

$$D^{(n)} = \begin{cases} P_{ij}^{(n-1)} - p_iq_j, & i = j \\ -(P_{ij}^{(n-1)} - p_iq_j) & i \neq j \end{cases} \quad (9)$$

2.4 世代间配子频率变化的其他表达方式

2.4.1 离衡差表达式

将(8)式代入(4)式,可得两个世代间配子频率变化量的离衡差表现形式

$$P_{ij}^{(n)} - P_{ij}^{(n-1)} = \begin{cases} -D^{(n)}c, & i = j \\ D^{(n)}c & i \neq j \end{cases} \quad (10)$$

2.4.2 基因频率表达式

将(9)式代入(10)式,可得两个世代间配子频率变化量的统一表达式

$$P_{ij}^{(n)} - P_{ij}^{(n-1)} = (p_iq_j - P_{ij}^{(n-1)}) \cdot c \quad (11)$$

2.5 推论 5

由推论 4 和(9)式,可推知第 n 代群体达到平衡的条件又可表示为

$$P_{ij}^{(n-1)} = p_iq_j, \quad i, j = 1, 2 \quad (12)$$

也就是说,当第 $n-1$ 代配子 A_iB_j 的频率就等于相应基因频率 $P(A_i)$ 和 $P(B_j)$ 的乘积时,在随机交配的情况下,第 n 代群体达到平衡。

$P(A_iB_j) = P(A_i)P(B_j)$, 意味着 A_i 基因和 B_j 基因在组合成配子时,是相互独立的,好像不连锁一样,故称连锁平衡(linkage equilibrium)。

3 平衡过程

3.1 配子频率的递推公式

由(11)式,可得到第 n 代群体产生 A_iB_j 配子的频率

$$\begin{aligned} P_{ij}^{(n)} &= (1-c)P_{ij}^{(n-1)} + cp_iq_j \\ &= (1-c)P_{ij}^{(n-1)} + cp_iq_j + p_iq_j - p_iq_j \\ &= (1-c)P_{ij}^{(n-1)} - p_iq_j(1-c) + p_iq_j \\ &= (1-c)(P_{ij}^{(n-1)} - p_iq_j) + p_iq_j \end{aligned} \quad (13)$$

可写成

$$P_{ij}^{(n)} - p_iq_j = (1-c)(P_{ij}^{(n-1)} - p_iq_j) \quad (14)$$

于是有

$$\begin{aligned} P_{ij}^{(1)} - p_iq_j &= (1-c)(P_{ij}^{(0)} - p_iq_j) \\ P_{ij}^{(2)} - p_iq_j &= (1-c)(P_{ij}^{(1)} - p_iq_j) = (1-c)^2(P_{ij}^{(0)} - p_iq_j) \\ &\vdots \\ P_{ij}^{(n)} - p_iq_j &= (1-c)^n(P_{ij}^{(0)} - p_iq_j) \end{aligned} \quad (15)$$

由这种递推关系可得到相对于原初群体配子频率的表达式

$$P_{ij}^{(n)} = (1-c)^n(P_{ij}^{(0)} - p_iq_j) + p_iq_j \quad (16)$$

所以,在已知原初群体的基因频率和配子基因型频率的情况下,其随机交配后代群体的配子频率可由(16)计算,进一步可求得下一世代($n+1$ 代)群体的基因型频率。

3.2 离衡差的递推公式

对照(9)式,(14)和(15)可写成离衡差的递推关系式

$$\begin{aligned} D^{(n+1)} &= (1-c)D^{(n)} \\ D^{(n+1)} &= (1-c)^n D^{(1)} \text{ 或 } D^{(n)} = (1-c)^{(n-1)} D^{(1)} \end{aligned} \quad (17)$$

式中 $D^{(1)}$ 是原初群体随机交配一代后群体的离衡差。因为原初群体不一定是由随机交配而来,故原初群体的离衡差 $D^{(0)}$ 无意义。

因为有 $c \neq 0$ 时, $c < 0.5$, 所以 $(1-c) < 1$, 当 $n \rightarrow \infty$ 时, $(1-c)^n \rightarrow 0$, $D^n \rightarrow 0$, 所以群体一定能达到平衡。

4 应用举例

假定基因型为 $A_1A_1B_1B_1$ 和 $A_2A_2B_1B_2$ 的品系以等数量、等性别混合,这个混合世代为 0 世代,如果这两个基因座是:(1)不连锁;(2)连锁,且重组值为 0.2,那么在随机交配第 3 代的基因型 $A_1A_1B_2B_2$ 在这两种情况下的频率分别为多少?

[解]在第 0 代, $A_1A_1B_1B_1$ 和 $A_2A_2B_1B_2$ 以等数量、等性别混合,所以 A_1, A_2, B_1, B_2 等位基因频率都是 0.5, 即 $p_1 = p_2 = q_1 = q_2 = 0.5$, $P_{ij}^{(0)} = 0$

若求随机交配 3 代的基因型 $A_1A_1B_1B_1$ 频率,只要知道第 2 代配子 A_1B_1 的频率 $P_{12}^{(2)}$ 即可,根据(16)有

$$P_{12}^{(2)} = (1-c)^2(P_{12}^{(0)} - p_1q_2) + p_1q_2$$

(1)不连锁情况, $c=0.5$

$$P_{12}^{(2)} = (1-0.5)^2(0-0.5 \times 0.5) + 0.5 \times 0.5 = 0.1875$$

所以,随机交配第 3 代的基因型 $A_1A_1B_2B_2$ 的频率

$$P^{(3)}(A_1A_1B_2B_2) = P_{12}^{(2)} \times P_{12}^{(2)} = 0.1875 \times 0.1875 = 0.0352$$

(2)连锁,且 $c=0.2$

$$P_{12}^{(2)} = (1-0.2)^2(0-0.5 \times 0.5) + 0.5 \times 0.5 = 0.09$$

所以,随机交配第 3 代的基因型 $A_1A_1B_2B_2$ 的频率

$$P^{(3)}(A_1A_1B_2B_2) = P_{12}^{(2)} \times P_{12}^{(2)} = 0.09 \times 0.09 = 0.0081$$

参 考 文 献 (References):

- [1] WANG Xiao-Long, YUAN Zhi-Fa, GUO Man-Cai, SONG Shi-De, ZHANG Quan-Qi, BAO Zhen-Min. Maximum Entropy Principle and Population Genetic Equilibrium. *Acta Genetica Sinica*, 2002, 6; 562~564.
汪小龙, 袁志发, 郭满才, 宋世德, 张全启, 包振民. 最大信息熵原理与群体遗传平衡. *遗传学报*, 2002, 6; 562~564.
- [2] LUO Jia-Bin, LI Yue-Qiu, WANG Qing-Yuan, WANG Wei-Ren, LÜ Xue-Shen. The Test of Double-Locus Genetic Equilibrium were Carried Out Between Three Unlinked Gene

System Hp, Gc and ABO. *Hereditas(Beijing)*, 1994, 2: 10~14.

罗佳滨, 李月秋, 王庆元, 王维人, 吕学读. Hp, Gc, ABO 三个非连锁基因系统间双位点遗传平衡检验. *遗传*, 1994, 2: 10~14.

[3] GUO Ping-Zhong. *Population Genetics Introduction*. Beijing: China Agriculture Press, 1993, 39~50.

郭平仲. *群体遗传学导论*. 北京: 农业出版社, 1993, 39~50.

[4] MAO Sheng-Xian, HUANG Yuan-Zhang. *Population Genetics and Programmer*. Beijing: Beijing Normal University

Press, 1991, 34~36.

毛盛贤, 黄远樟. *群体遗传及其程序设计*. 北京: 北京师范大学出版社, 1991, 34~36.

[5] Bruce S. Weir. *Data analysis: Methods for discrete population genetic data*. Beijing: China Agriculture Press, 1996, 89~108.

Bruce S. Weir. *遗传学数据分析*. 北京: 中国农业出版社, 1996, 89~108.

[6] Franz Pirchner. *Population Genetics in Animal Breeding*, 2nd ed. New York: Plenum Press, 1983, 12~16.

上海交通大学 Bio-X 生命科学研究中心 招聘交叉学科研究生和研究人员启事

自从进入 21 世纪后, 生命科学已明显地跃到了前沿科学的位置, 其重要性得到世人的共识。然而, 未来生命科学的关键性突破将越来越依赖于生物学与物理学、化学、数学、电子学、信息学、医学、药学、农学、工程学等的交融与交叉。为顺应这种趋势和“创新是一个民族的灵魂, 是国家兴旺发达的不竭动力”的精神, 上海交通大学经过研究和论证于 2000 年成立了世界上第二个 Bio-X 生命科学研究中心(以下简称“中心”), 为百年老校翻开新的一页。从此标志了我校在国内率先正式启动了生命学科与其他多学科间进行交叉研究的研究基地的建设。目前, “中心”已从国内外引进了一批优秀人才, 共建有数个不同学科的核心课题组和多个“卫星”课题组, 在徐汇校区拥有 4,000 平方米建地面积用作“中心”的实验基地。“中心”的建设已列为上海交通大学“211 和 985 工程”重点项目之一, 并得到了国家 973、863、国家自然科学基金、教育部基金、上海科委基金以及国外企业界的资助。“中心”正积极参加国际竞争, 扎扎实实地为人类做几项贡献! 诺贝尔奖获得者朱棣文教授于 2002 年 6 月正式接受作为“中心”的名誉主任; “中心”的现主任是贺林教授。

“中心”目前围绕着“大生物学”的主题思想, 通过多学科交叉的优势, 已完成了中国第一台“DNA 计算机”雏形的研究。根据课题发展的需要, 为加大对 DNA 计算机的研究力度, “中心”为多学科交叉专门提供了 800 平方米的实验室, 并为 DNA 计算研究提供了 100 平方米专门研究场所, 目前希望招聘若干研究生和研究人员加入以下课题的研究。

子课题 1: DNA 数学理论和计算理论

子课题 2: DNA 分子作为信息储存介质的特征研究

子课题 3: 单分子和单磁珠 DNA 计算单元的研制

子课题 4: 高通量并行计算的 SNP 计算模型样机

子课题 5: 超高速 DNA 碱基变换技术的前期研究

子课题 6: 新型 DNA 计算硬件的研究开发

子课题 7: DNA 计算相关的微纳米排布技术

子课题 8: DNA 计算的基本单位的定性定量研究

子课题 9: DNA 计算芯片与电子设备之间的通讯和接口模块的研究

子课题 10: DNA 计算芯片作为双向协处理器与电子计算机 CPU 的整合

子课题 11: 微缩 DNA 计算反应器

所招人员的背景不限, 但要求对以生命科学为主题的交叉学科研究有兴趣, 尤其是对生物大分子相关技术, 生物信息学, 以及/或信息学科(包括计算机科学)的理论和一定的技术有一定的基础。研究生导师将实行多位导师共同负责制。

研究室设在上海交大的徐汇校区。研究生待遇将按国家有关规定执行, 同时根据“中心”的有关规定提供一些额外的研究津贴。新招工作人员除享有国家有关规定的固定工资待遇外, 根据面试的结果, 通过面议决定额外津贴数额; 工作方式为合同制, 但根据贡献情况, 被录用者将有机会成为上海交大的正式职工。

有意加入者, 请寄个人简历至以下联系人和地址:

曹冬梅, 华山路 1954 号上海交通大学 Bio-X 生命科学研究中心(浩然大楼 501 信箱), 200030

E-mail: sheilaback@citiz; zhangzz@sjtu.edu.cn