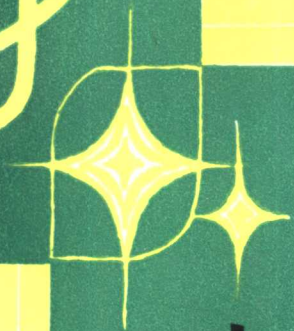




全国高等农业院校教材

全国高等农业院校教材指导委员会审定



人类遗传连锁分析

(修订版)

● 泽格·欧特 著 杨达 译 吴常信 校

● 动物遗传育种 植物遗传育种专业用

北京农业大学出版社

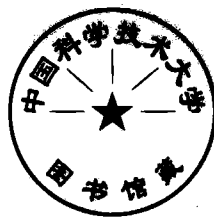
人类遗传连锁分析

(修订版)

泽格·欧特 著

杨 达 译 吴常信 校

动物遗传育种、植物遗传育种专业用



北京农业大学出版社

Analysis of Human Genetic Linkage

Revised Edition

JURG OTT

Professor, Department of Genetics and Development and
Department of Psychiatry, Columbia University

**The Johns Hopkins University Press
Baltimore and London**

图书在版编目(CIP)数据

人类遗传连锁分析/(美)欧特 Ott, J.)著;杨达译.
北京:北京农业大学出版社,1995,9

书名原文:Analysis of Human Genetic Linkage

ISBN 7-81002-722-0

I. 人… I. ①欧… ②杨… II. 人类基因-连锁(遗传)
-研究 N. Q987

中国版本图书馆 CIP 数据核字(95)第 13705 号

图字:01 — 95 — 3961 号

出 版 北京农业大学出版社
发 行
经 销 新华书店
印 刷 北京丰华印刷厂印刷
版 次 1995 年 11 月第 1 版
印 次 1995 年 11 月第 1 次印刷
开 本 16 印张 12.25 千字 300
规 格 787×1092 毫米
印 数 1—2040

定 价 9.60 元

Copyright(©)1991 by the Johns Hopkins University Press
This edition of Analysis of Human Genetic Linkage,
Revised, is published by arrangement with the Johns
Hopkins University Press, Baltimore, Maryland, U. S. A.

本书简介

《人类遗传连锁分析》(修订版)一书是美国哥伦比亚大学泽格·欧特(Jurg Ott)教授所著。经约翰·霍普金斯大学出版社(The Johns Hopkins University Press)同意,由北京农业大学杨达博士翻译成中文。原出版社同意中文版自首次发行日期起,四年内可在中国大陆、台湾和香港发行与销售。

全书系统地介绍了遗传连锁分析的原理与方法,内容深入浅出,引用文献资料丰富,是目前国内生物学、医学、农学、畜牧学等专业高等院校不可多得的教材。主要读者对象是上述有关专业的科技人员、教师、研究生。本书也可作为生命科学中其他专业人员以及大学本科学生的参考书。

作者序言

我很感激对本书第一版的生动热烈的反响。与同事们的讨论，特别是与 Robert Elston 博士的讨论，指出了第一版中需要改进的一些领域。我感谢 Lodewijk Sandkuijl, Daniel Weeks, Neil Risch 等博士对此书有助益的讨论，及 Joseph Terwilliger 对书稿的认真阅读及许多有建设性的讨论。

我一直想用简单与直觉的方法来介绍书中的内容。书中各章节的数学难度不同，理论方面一般集中在每节的结尾部分。每章之后都有习题，它们的答案附在书的结尾。某些习题涉及到一些在一般书内容中可能是太专业性的特别理论。

对于想学习连锁分析理论方面的读者，熟悉方程与简单的代数将是有帮助的。与第一版相同，这一版没有介绍概率微积分。没有概率知识的读者应接受书中的结论并应能读懂书中的大部分内容。但是，书中将介绍一些统计方法，特别是显著性检验。

自从本书第一版问世以来，连锁分析的题目与问题已发生巨大变化。同时，已存在的技术也有了显著的变迁。很自然，现在多位点连锁分析占有更多的篇幅，而少数特别专业化的节段（如变配类型）被缩短或删除。此外，计算机模拟，这个以前没有介绍的方法，正成为一个强有力的工具，并将予以广泛的介绍。

本书没有介绍群体遗传学，因为在这个重要的题目上有杰出的处理（CavalliSforza 和 Bodmer 1971; Hartl 1988; Maynard Smith 1989）。Hartl (1988) 介绍了群体遗传学的分子遗传学方面，而 Maynard Smith (1989) 从进化角度介绍了现代遗传学概念。Weir (1990) 对经典与群体遗传学中的广泛题目做了数学处理，包括 DNA 序列的统计方法。对于人类遗传学更普通的内容，读者可参考某个教科书，特别是由 Vogel 和 Motulsky (1986) 所写的教科书。对于连锁分析的原则，Conneally 和 Rivas (1986) 有一个一般性的简短概括。在系谱分析中的抽样理论可以在 Thompson (1986) 的书中找到。各种基因定位与序列检验的题目，以及各种定位技术的易读指南已由美国国会技术估价办公室 (1988) 发行的书中加以介绍。

在连锁分析这样的领域中，各种不同职业背景的专家互相合作。根据专家们的定义来使用专业术语是重要的但明显是很困难的。在专业术语方面存在相当程度的混淆，而且对某一特定小范围专业不很熟悉的研究者要知道现行技术术语的意思是困难的。在某一特定研究领域的技术术语在被不同领域的不同专业的人使用后可能出现新的涵义。例如，连锁和基因图谱在早期使用时只有纯粹的遗传意义：连锁是指遗传连锁（连锁群）而基因图谱是代表基因线性排列的理论构造，所以实体定位（或物理定位）法是该词以原始意义使用时的一种相互矛盾。现在许多这些术语和它们的新意义在人类遗传学中如此流行，过于咬文嚼字没有必要。

我在定义与使用术语时的标准是使混淆最小化。例如，突变以前是指一个等位基因变成另一个等位基因的过程。但现在，它经常表示突变了等的位基因，甚至只表示致病等位基因。在某些场合，避免混淆意味着否定一些新的涵义。例如，同型交配本来是指两个表现型的交配，但现在有时是指一对配偶祖先的表现型，而与该对配偶的表现型无关。这后一种用法不应继续使用。而且，不幸的是，支持区间经常被称为置信区间。有时要知道一个术语到底意

味什么是困难的。最好的例子似乎是非连锁平衡。该词本来纯粹是用来表达单倍体中等位基因不相互独立出现这一事实。但现在，它经常被用来表示以上现象的原因，即紧密的连锁。

连锁分析有时被理解为应用适当计算机程序的问题，以致于任何有足够计算机知识的人在有家系资料和标记基因后都能“做连锁分析”。这种断言在其它研究领域一般是不正确的，在这里也不正确。由没有必要理论背景的人来执行连锁分析是危险的。为方便只想草率阅读本书的读者，书的目录列出了本书的各章节。

译者序言

我非常感谢本书作者泽格·欧特教授和出版者约翰·霍普金斯大学出版社无偿给予北京农业大学出版社全部翻译出版此书并在中国大陆、香港与台湾发行这个中译版。

这个中译版之所以能与读者见面还在于中国农业部教育司毛达如教授与北京农业大学吴常信教授的大力支持与热忱鼓励。北京农业大学张勤博士对译文的排版与技术问题提供了很大帮助，吴常信教授与张勤博士还为译文的审校付出了辛勤的劳动。本书翻译的部分工作是我在北京农业大学由中国——康奈尔合作项目资助的短期工作中完成的。对所有这些，我谨在此表示衷心感谢！

分子生物学技术的发展提供了丰富的多态标记基因及发现与定位基因的新途径，从而使对人类致病基因及动植物有经济价值基因的定位成了一个新的热门学科。遗传连锁分析是基因定位的一个基本手段。《人类遗传连锁分析》一书总结了现代遗传连锁分析的理论与技术，是人类遗传、医学遗传与动植物遗传育种等学科教学与科研的重要参考书。

杨 达

1995年3月17日于美国

目 录

第一章 普通遗传学与细胞遗传学基础	1
1.1 孟德尔遗传学	1
1.2 重组与遗传连锁	3
1.3 连锁群与同体	4
1.4 交换与图距	6
1.5 图距函数	8
1.6 人类染色体	11
第二章 基因与遗传多态	14
2.1 基因的术语与特征	14
2.2 多态性程度	15
2.3 血液细胞抗原	16
2.4 血液蛋白	20
2.5 DNA 多态性和基因组研究	20
2.6 染色体形态学的变异	21
2.7 人类基因组的标记基因复盖率	22
第三章 统计推断方法	24
3.1 似然率	24
3.2 最大似然率估计	25
3.3 最大似然率估计的统计性质	26
3.4 显著性检验	27
3.5 似然率法	29
3.6 区间估计	30
3.7 贝叶斯定理	31
第四章 连锁分析方法	33
4.1 连锁分析简史	33
4.2 θ 的先验与后验分布	35
4.3 似然分数法	37
4.4 检验连锁	39
4.5 等量观察值	41
4.6 简单家系类型的严格检验	42
4.7 多重比较	44
4.8 家系资料的似然率	46
4.9 非参数方法	46
4.10 若干特殊方法	48

第五章	家系资料的信息量	49
5.1	信息量的度量	49
5.2	期望似然分数	50
5.3	期望信息	52
5.4	交配类型	53
5.5	两个等位基因的双杂交	54
5.6	多于两等位基因的双杂交	56
5.7	已知相双回交	58
5.8	两子女未知相双回交	58
5.9	多于两子女的未知相双回交	60
5.10	发现连锁所需要的观察值数	62
第六章	多位点连锁分析	64
6.1	符号与术语	64
6.2	三点分析	66
6.3	两子女的未知相三回交	69
6.4	干扰	71
6.5	多位点图距函数	74
6.6	排列位点	76
6.7	完全干扰下的定位	80
6.8	其它方法方面的问题	83
第七章	表现率	86
7.1	表现率的定义	86
7.2	不完全表现率的代价	87
7.3	估计表现率与重组率	88
7.4	与年龄有关的表现率	91
7.5	影响表现率的因素	94
7.6	数量表现型	95
第八章	数值与计算机方法	97
8.1	似然分数的解析法计算	97
8.2	爱尔斯顿-斯图雅特法	99
8.3	连锁分析的计算机程序	101
8.4	连锁程序的特殊应用	103
8.5	连锁工具程序	107
8.6	父性计算	109
8.7	计算机模拟方法	109
第九章	重组率的变异	114
9.1	男性与女性重组率	114
9.2	重组率与年龄	116

9.3	不同家系类型的异质性	116
9.4	混合家系所产生的异质性	119
9.5	异质性的多层与混合模型	126
9.6	混合性检验中的协变量	126
第十章 不一致性		128
10.1	选择偏差	128
10.2	分类错误	130
10.3	表现率的定义错误	132
10.4	异质性的错误分类	132
10.5	系谱错误	133
10.6	其它模型定义错误	134
第十一章 疾病位点的连锁分析		136
11.1	遗传因素	136
11.2	疾病表现型	137
11.3	遗传模式	138
11.4	表现型与粒子关联	141
11.5	候选位点	145
11.6	排除定位	146
11.7	遗传风险	147
11.8	非规则分离	149
11.9	计划连锁分析	150
习题解答		153
英汉词汇对照		158
参考文献		163

第一章 普通遗传学与细胞遗传学基础

1.1 孟德尔遗传学

遗传学与遗传连锁经常针对于染色体加以介绍。但是,从概念上,遗传与细胞现象是很不相同的。在本节与下一节,基因与遗传连锁将用纯遗传术语来定义,但随后交替使用普通遗传学、细胞遗传学及分子遗传学的术语将是方便的。

在用豌豆的杂交试验中,格里高尔·孟德尔(Gregor Mendel)提出了现今被称为孟德尔定律的假设(Mendel 1886; Bateson 1913)。在1900年由考兰斯(Correns)、滋切麦克(Tschermak)和德弗里斯(de Vries)重新发现了孟德尔定律后,苏顿(Sutton)和波威利(Boveri)在1902至1904年公式化了遗传的染色体理论。对于染色体的正确理解赋予了孟德尔定律所缺乏的说服力。以下是用现行术语对常染色体孟德尔遗传的一个简要描述。对普通遗传学与细胞遗传学基本概念的更详尽介绍可在许多教科书中找到。鲍勒(Bowler 1989)对孟德尔主义给予了有趣的历史及社会方面的分析。

可遗传性状是由基因控制的。不同的基因负责不同性状的表达。用现代术语,基因是一个特定的DNA译码序列(Elant-Johnson 1971, 1973; Hartl 1988),是传递、重组和功能的单位(Vogel和Motulsky 1986)。据估计,人类大约有50,000个基因而果蝇大约有5,000个基因(见McKusick和Ruddle 1977文中的参考文献)。在果蝇中主要基因被认为有约1,000个(Maynard Smith 1989, 1959)。每个个体在每个基因座上携带两个基因,一个来自母亲,一个来自父亲。一个基因可有不同的形态和状态,称为等位基因,而每个等位基因可能有不同的表达形式。例如,ABO血型的三个等位基因的相互作用决定各种ABO血型。区别不同性状是由同一基因的不同等位基因决定的还是由不同基因决定的有时是困难的,而对这个问题的解决方法将推迟到下一节。

一个基因的不同等位基因的相对频率叫做基因频率。每个个体对群体基因库贡献两个等位基因(如果假设常染色体的话)。例如ABO基因的三个等位基因A、B和O在白种人中的相对频率是 $P_A=0.28$, $P_B=0.06$, $P_O=0.66$ 。一个基因被称为多态的如果它最常出现的等位基因的频率不高于95%(有时用较不严格的99%这个标准)。

个体的两个等位基因构成该个体的基因型。 n 个等位基因可以构成 $n(n+1)/2$ 个基因型。例如,ABO血型基因($n=3$)可以构成六个基因型,A/A、A/B、A/O、B/B、B/O、O/O。个体的两个等位基因可能相同也可能不相同。由两个相同等位基因构成的基因型称为纯合体(如A/A、B/B、O/O),由不同等位基因构成的基因型称为杂合体(A/B、A/O、B/O)。在性连锁中,同样的术语适用于女性,但男性的X连锁基因的基因型称为半合子(如A/Y,这里Y代表Y染色体)。最后,当纯合体的两个等位基因是来自一个共同祖先时(基因同源),该基因型被称为自合体。

一个特定基因型的表达称为表现型。在ABO基因位点,六个基因型决定四个表现型(血型),即,A型,B型,AB型,及O型。基因型与质量性状表现型的关系可用一个表格来表

示，用行代表基因型，列代表表现型。表中的每个格是表现率。表现率是给定基因型后观察相应表现型的条件概率。在简单的情形中，表现率是 0 或 1。但是许多疾病有介于 0 与 1 之间的表现率。表 1.1 是 ABO 基因表现率的一个例子。两个基因型 A/A 和 A/O 有同一表现型，A 血型。或者，用遗传学家的话来说，A 对 O 是显性 ($A > O$)，因为 A 等位基因不管 O 等位基因的存在与否都得到表达。反过来说，O 对 A 是隐性基因，因为 A 存在时 O 对表现型无影响。O 等位基因只在纯合 O/O 基因型中得到表达，即导致 O 血型。B 对 O 也是显性，但 A 和 B 是共显性，因为 A 和 B 在 A/B 基因型中都得到表达。

表 1.1 在 ABO 位点上基因型与表现型（表现率）之间的关系

基因型	表 现 型			
	A 型	B 型	AB 型	O 型
A/A	1	0	0	0
A/B	0	0	1	0
A/O	1	0	0	0
B/B	0	1	0	0
B/O	0	1	0	0
O/O	0	0	0	1

在配子形成过程中，每个个体以 1/2 的概率传给孩子该个体两个等位基因的一个。从父母得到的等位基因构成孩子的基因型。例如，如果父母的血型是 AB 和 O（基因型 A/B 和 O/O）。他们孩子半数的基因型预计是 A/O，另一半的基因型预计是 B/O。如果所假设的遗传模型是正确的，对这个 1:1 比例的偏差将是由于机率摆动造成的。当一个新基因被发现并被假设服从孟德尔遗传模式，实践中经常要做的是通过已知父母基因型来预测孩子不同表现型的比例并检验观察与期望比例的符合程度。例如，朱尼加等 (Juneja 等 1988) 研究了原生质 $\alpha_1\beta$ -糖蛋白的四个等位基因。他们对 57 个家系根据父母基因型进行分类并验证了所观察的子女基因型与孟德尔遗传模式所预测的频率相符。两个或更多的基因型被称为遗传不一致或遗传不相等如果它们与孟德尔定律不相合。例如，母亲的 A/B 基因型与子女 O/O 基因型是遗传不一致的。

在随机交配的条件下，即没有迁移、突变和选择，一个群体被称为处于哈代-温伯格平衡，也就是基因型频率只取决于基因频率 (Hartl 1988)。例如， $P(A/A) = P_A^2$ ， $P(A/B) = 2P_AP_B$ 。要注意基因型 B/A 与 A/B 相同。哈代-温伯格定律的突出特点是，对于常染色体基因，无论父母代基因型的分布如何，哈代-温伯格平衡在一代随机交配后即达到。用另一句话来说，子女代基因型频率只取决于父母代的基因频率而不是父母代的基因型频率。但是，在有限群体中，基因型频率不是保持不变的。由于随机变动，第二代中的等位基因频率可能与第一代中的有所偏差，然后第二代中新的等位基因频率将决定第三代中期望的基因型频率。这种离散性的随机动力被称为遗传漂移 (Maynard Smith 1989)。

哈代-温伯格平衡的存在经常被用来表示随机交配的存在。虽然随机交配意味着哈代-温伯格平衡，但反过来不一定成立。某些类型对随机交配的偏差（称为拟随机交配）已被证明也导致子女代中的哈代-温伯格平衡 (Li 1988)。

现在我们同时考虑两个基因，基因 1 有两个等位基因，A 和 a，基因 2 也有两个等位基因，

B 和 b。例如，图 1.1 显示一个人为构造的家系，其中祖父母都是双纯合体；祖母在基因 1 是纯合的等位基因 A，在基因 2 是纯合 B，祖父是纯合 a 和 b。在实验动物中，这类纯合性可以通过对理想基因型的连续近交和选择而获得。这个家系的父亲（祖父母的儿子）肯定从祖母得到了 A 和 B，从祖父得到了 a 和 b。一个个体从一个亲体（父或母）得到的所有等位基因（在不同基因位点）叫做单倍体。所以，例子中的父亲从他母亲得到了 AB 单倍体，从他父亲得到了 ab 单倍体。

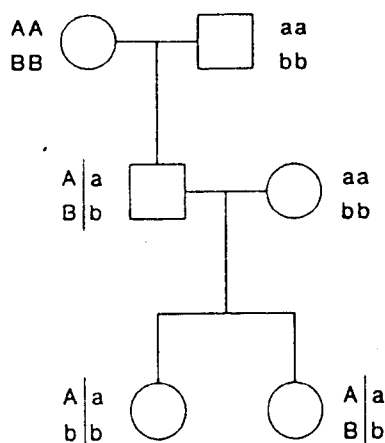


图 1.1 在两位点上表现一个重组及一个非重组后代的假定家系

原则上，双杂合体个体 ($A/a, B/b$) 从一个亲体得到 A 等位基因。以及 B 或 b 等位基因。这两种可能性被区分为杂合体的两个相（在两个以上基因位点是杂合的个体中存在着两个以上的相）。根据人们对等位基因的兴趣，两个相的一个叫顺相，另一个叫反相。通常，双杂合个体的相是未知的，也就是，哪两个等位基因构成从一个亲体得到的单倍体是不清楚的。图 1.1 中的父亲（中间的一代）的相是已知的。图中的相由竖线将基因型中从母亲得到的等位基因与从父亲得到的分开。对于最多只有一个基因是杂合的基因型，相总是已知的，因为等位基因来自哪个亲体是清楚的，尽管人们可能无法区别一个单倍体来自哪一个亲体。

与单个基因的基因频率相类似，可以对群体的单倍体频率加以定义。如果 n_i 是第 i 个基因的等位基因数， $i=1 \dots N$ ，所有 N 个基因的单倍体总数 H 是 n_i 的乘积，即， $H = \prod n_i$ 。不同的基因型总数是 $H(H+1)/2$ ，这里不同相但同一基因型的多重杂合体被计算为不同的基因型。例如，当三个基因各有 4, 2, 3 个等位基因时，单倍体总数是 24，可能的基因型数 300。当不同基因的等位基因在单倍体中独立出现时，一个单倍体的群体频率是构成该单倍体的等位基因 i 频率的乘积。对等位基因在单倍体中随机出现的偏差被称为等位基因关联或非连锁平衡。这个题目将在 11.4 节中详细讨论。

1.2 重组与遗传连锁

图 1.1 中假设了双杂合体父亲的配偶是一个双纯合体的个体 ab/ab ，而这个双纯合体个体与前一代一个亲体（祖父母中的一个）有同样的基因型。这样的交配被称为回交。在以上例子中，这个回交是一个已知相回交。两个双杂合体 ($A/a, B/b$) 个体的交配被称为杂交

(intercross)。如果图 1.1 中假设的两个基因相互独立, 父亲预计将以 $1:1:1:1$ 的比例传递给后代四个单倍体, AB, ab, Ab 和 aB。AB 和 ab 单倍体与他的父亲(系谱中的祖父)所得到的单倍体相同。在经典遗传学中, 得到这些单倍体的子女称为亲体型(如图 1.1 中右侧的子女)。另两个单倍体, Ab 和 aB, 与其父从祖父母得到的任何一个单倍体都不同, 并含有从祖父或祖母得到的一个等位基因(从祖父母得到的等位基因在父亲中一定发生了重新组合)。在经典遗传学中, 有这类单倍体的后代被称为非亲体型。在人类遗传学中, 非亲体型, Ab 和 aB, 被称为重组体, 而其它两个单倍体(AB 和 ab)被称为非重组体。两个基因间的重组表示两个不同祖父母对一个个体的单倍体提供了每个基因的一个等位基因, 而非重组体带有从同一祖父或祖母起源的两个等位基因。一个后代根据其是否显示父和母出现了重组被称为重组型和非重组型。所以, 在图 1.1 中, 女儿 1 是重组型, 女儿 2 是非重组型。

当两个基因独立遗传时, 后代中的重组与非重组体应以相等数量出现。对于某些基因对, 可能观察到在以下意义上对 $1:1$ 比例的一致偏差: 从祖父或祖母传给父或母的单倍体中的等位基因倾向于再由父或母传给子女。用另一句话来说, 不同基因的等位基因以遗传配对出现。这个现象叫做遗传连锁。如果双杂合体亲体只产生非重组配子, 两个基因被称为完全连锁。如果一个亲体产生同样数量的重组与非重组型, 两个基因无连锁。

遗传连锁程度是用重组率度量的。重组率是亲体产生一个重组型配子的概率。传统上, 在人类遗传学中, 重组率用希腊字母 θ (读 Cita) 来表示。独立分离的基因之间无连锁, 其特征是重组率 $\theta=1/2$ 。连锁基因的特征是 $\theta<1/2$ 。某些基因对有着紧密连锁, 以致于 θ 趋近于 0, 也就是, 在那两个基因间的重组是非常稀有的。估计重组与检验无重组 ($\theta=1/2$) 及相反的有连锁假设是连锁分析的目的。

两个或更多基因间遗传连锁有以下特点:

1. 由于重组事件只能从父母传给子女的单倍体来识别, 连锁分析不能用不相关的亲属而要有对亲属的观察值。所以, 为了分析连锁, 研究者要收集关于家系成员的表型信息。
2. 由一个亲体产生的重组和非重组单倍体不总是可以区别。例如由基因型为 Ab/ab 亲体产生的 Ab 单倍体。由于第二个基因是纯合型, Ab 可能是重组型也可能是非重组型。要使单倍体能得到区别, 亲体必须是双杂合体。这时遗传学家称那个亲体为可能有连锁信息。一个交配有可能是有连锁信息的, 如果至少一个亲体是双杂合体。
3. 对于许多基因的连锁分析表明, 重组率依产生配子的亲体性别不同而不同。所以, 人们对男性的重组率 (θ_m) 与女性的重组率 (θ_f) 加以区别。这里的性别是指发生重组事件的父母代个体的性别, 而不是用来计数重组或非重组型的子女的性别。重组率性别因素将在 9.1 节详述。

1.3 连锁群与同体

如果基因 1 在遗传上与基因 2 连锁, 而基因 2 又与基因 3 连锁, 基图 1 与基因 3 看上去不一定有连锁。连锁群被定义为一系列基因, 其中一个基因与至少同一系列的另一个基因相连锁。经验表明, 将基因分为连锁群在以下意义上是专一的: 属于不同连锁群的两个基因没有连锁。对于早期遗传学家, 连锁群遗传现象的解释通过以下观察而得以澄清: 在对于每个被研究的双倍体物种, 连锁群的数目与染色体的数目巧合。现在连锁群可被解释为基因在其

上线性排列的染色体的对等物。基因以成对出现的现象是由染色体也以成对存在这一事实来解释的,即所谓的同源染色体。相连锁的基因位于同一染色体上。最后,由于减数分裂过程,配子只包含一套单倍体染色体,这解释了为什么传给后代是每个基因的一个等位基因。

位于同一染色体上的基因被称为同体。一对基因的同体总表示这两个基因属于同一连锁群,尽管遗传上它们不一定有连锁。相反,非同体基因有时也表现连锁性的关联,有时被称为拟连锁或亲合性 (Bailey 1961)。有人用老鼠中的一个例子讨论过各种可能机制 (Sturt 等, 1967)。

不同基因的两个等位基因出现在同一单倍体中这一事实并不意味着这些基因是同体。图 1.1 父亲基因型的竖线只不过表示相已知,但这与这两个基因的连锁或同体状态无关。这是许多学连锁分析的学生理解混乱的一个来源。对这些关系有个清楚的理解是非常重要的。

在人类中除了 22 对常染色体外,还有两个性染色体, X 和 Y。女性有两个 X 染色体,男性有一个 X 和一个 Y 染色体。在这之前所涉及的染色体都是指常染色体。X 和 Y 染色体也携带基因,这些基因的特征是性连锁遗传模式,因为男孩 (X/Y) 从母亲得到 X,从父亲得到 Y,而女孩 (X/X) 从父母各得到一个 X 染色体。对 X 连锁基因,每个女性对群体等位基因的数目贡献两个等位基因,而男性对群体贡献一个等位基因。所以,假设群体中有同样数目的女性与男性, X 连锁基因的一个等位基因的频率是 $(2P_f + P_m) / 3$, 其中 P_f 是女性中的基因频率, P_m 是男性中的基因频率。在平衡条件下, $P_f = P_m$ (X 连锁基因不是在一代随机交配后就达到哈代-温伯格平衡,而是逐渐达到的)。X 连锁基因的著名例子是色盲与血友病。许多常染色体基因的遗传类似于 X 连锁基因,即,基因型在不同性别中表现不同。这个现象叫伴性遗传,但不应与性连锁遗传相混淆。

近些年来,分子遗传学的进展极为迅速,对基因的结构与功能已了解甚多。所有这些都有对应的 DNA 分子的核苷酸序列。它们包括 DNA 的功能序列,即所谓的外显子 (Exon)。这些外显子分散于各种片段中的非表达序列,即所谓的内含子 (Intron),或干扰序列。有些基因非常长。例如,导致杜氏肌萎缩病的基因至少有 65 个外显子,而这些外显子分散于大约 2000 公对 (kb) 基因组上 (Koenig 等, 1987) [一公对=1000 核苷酸对,一译者]。有些 DNA 序列能改变它们在基因组上的位置。这种变化可由感染造成 (Shapiro 1983, Maynard Smith 1989)。这种序列称为转位子 (Transposon)。对基因的群体与分子遗传方面及基因表达的简述可在哈特尔 (Hartl, 1988) 文中找到。对重组的生化过程只有有限的理解。对噬菌体和植物的研究表明重组发生在染色体段中不重迭的特定地点上 (White 和 Lalouel, 1987)。细胞酶可能参与重组过程,通过在减数分裂过程中沿染色体对活跃地折断染色体来诱发重组 (Watson 等, 1987a)。结果是,重组率不一定是连锁基因实体距离的好指标。

除了细胞核内染色体上的核基因外,线粒体也含有 DNA (mtDNA) (“另一类人类基因组”, Science 249: 1104-5, 1990)。线粒体基因控制下的表现型服从母性遗传的模式 (Wallace 1989)。母性遗传病的一个著名例子是所谓的利波 (Leber's disease) 病。这个病已被证明是线粒体 DNA 中一个核苷酸的变化,但在某些家系中似乎是由不同的基因引起的 (Vilkkil, Savontaus, 和 Nikoskelainen, 1989)。由于高突变率和缺乏重组, mtDNA 非常适用于对遗传多样性的研究 (如 Excoffier, 1990)。有人推测,人类现今所有成员的线粒体 DNA 都起源于 20 万年前非洲的一个女性 (Cann 等, 1987),但也有对线粒体进化的其它解释 (Krüger 和 Vogel 1989; Excoffier 和 Langaney 1989)。大多数现代非洲人比现在白种人更接近于古代线粒