遗传学数据分析

——群体遗传学离散型数据分析方法

[美] Bruce S. Weir 著
徐云碧 王志宁 俞志华 译
朱 军 校

中国农业出版社

遗传学数据分析

——群体遗传学离散型数据分析方法

〔美〕 Bruce S. Weir 著
徐云碧 王志宁 俞志华 译
朱 军 校

中国农业出版社

Genetic Data Analysis

Methods for Discrete Population Genetic Data

Bruce S. Weir

North Carolina State University

本书根据 Bruce S. Weir 所著的 Genetic Data Analysis 一书1990年英文版翻译。

遗传学数据分析

——群体遗传学离散型数据分析方法

〔美〕Bruce S. Weir 著

徐云碧 王志宁 俞志华 译

朱 军 校

* * *

责任编辑 徐建华

中国农业出版社出版（北京市朝阳区农展馆北路2号）

新华书店北京发行所发行 北京市密云县印刷厂印刷

850×1168mm32开本 13.75印张 348千字

1996年2月第1版 1996年2月北京第1次印刷

印数 1—900 册 定价 25.50 元

ISBN 7-109-03764-9/Q·228

译 者 的 话

美国北卡罗来纳州立大学著名统计遗传学家 B. S. Weir 博士所著的《遗传学数据分析——群体遗传学离散型数据分析方法》，是一本在国外颇受欢迎的、遗传学与统计学相结合的科技著作和教学参考书。

浙江农业大学在本书出版之始，就以英文版本作教材为遗传育种专业的博士研究生开设了“高级群体遗传学”的课程。参加首期教学的师生深感其内容之新颖，方法之独到，无论对从事遗传学和统计学研究的同行，还是有志于这一专业的学生都是一本难得的参考书，有必要将其译成中文，使更多的同行能够从中得到启发和教益，从而推动我国遗传育种研究及其相关学科的发展。这就是我们要将本书的中译本奉献给读者的本意，也是我们的期望之所在。

本书共8章。作者在论述了哈代—温伯格非平衡和连锁非平衡估计、群体遗传结构分析和遗传距离估计等方法之后，还向读者介绍了进化树的构建以及 DNA 和蛋白质序列等分子数据的分析方法。其中第1章介绍了各种不同类型的离散型遗传数据、术语和将要采用的基本的统计学工具，为后续章节的学习打下了基础。第2—6章讨论了应用新型的统计分析手段来分析一些经典的统计遗传学数据，第7—8章涉及分子数据的分析方法。书末附有统计图表和计算机程序。

本书的前言和第1至2章由王志宁翻译，第3至5章由俞志华翻译，第6至8章由徐云碧翻译，附录部分由俞志华和徐云碧负责翻译。全部译稿由徐云碧初校。最后由朱军教授作统一校阅。同时，根

据作者提供的修正稿以及我们的教学实践对原著作了多处修正。为了方便读者学习，我们将原作以Fortran/77语言编写的程序全部列出。此外，在教学中，我们还用Quick BASIC语言为几乎全部的习题编写了计算程序，可以提供给有兴趣的读者和教师。

在本书出版之际，我们对浙江农业大学申宗坦和季道藩两位教授的关心和帮助，表示感谢。

由于译者水平所限，本书不免有翻译不当之处。我们热忱欢迎读者批评指正。

译校者

一九九四年六月于浙江农业大学

原　序

本书阐述了离散型群体遗传数据的统计分析方法。这些数据通常是一些计数，可以是基因型也可以是碱基类型。而由度量诸如株高或产量之类遗传性状所得的数量遗传数据，则不属本书所述范围。

群体遗传学家正面临着分析新型遗传数据的令人兴奋的任务，现在似乎是评述一些适合于分析限制性片段长度多态性和DNA序列的统计技术的时候了。本书可望表明，即使是新型的数据，也能用R. A. Fisher，J. B. S. Haldane和Sewall Wright在早期研究中发展起来的，并在此后为许多工作者所扩展的统计技术来处理。我希望以此填补20年前Regina Elandt—Johnson所著《遗传学中的概率模型和统计方法》一书中精彩的处理方法与当前科技文献之间存在的空缺。我力图拓宽涉及面，并在适当处参阅了一些更加专门的书籍。例如，读者也许要阅读Jurg Ott的《人类遗传连锁之分析》一书，以便详细了解连锁、尤其是人类连锁的估计，以及Masatoshi Nei的《分子进化遗传学》，以便全面合理地分析分子遗传数据。期望本书能成为一本用传统方法分析哈迪—温伯格及连锁不平衡，以及描述群体结构和估计遗传距离的指导书。本书还向读者介绍了一些构建进化树的方法，以及DNA或蛋白质序列排列和比较中所遇到的一些难题。

撰写有关分子遗传数据分析方法的书籍，好比是瞄准一个移动着的靶子。随着人类基因组计划中数据的积累，日趋复杂的分析方法正迅速涌现，并且这些方法也确实需要。如果能够收到有关未来版本中应包括的新论题的建议，以及告知本版中的错误或不明确之处，我将会很高兴由于许多人认真仔细的审阅，才使本书得以较为完善。我衷心感谢Bill Hill (Edinburgh)，Wyatt Anderson，

Jonathan Arnold, Jamie Cuticchia 和 Yunxin Fu(Athens, Georgia), Joe Felsenstein (Seattle), Mike Clegg (Riverside), 和 Chris Basten, Ken Dodds, Rebecca Doerge, Spencer Muse, 和朱军 (Raleigh)。在一些专业问题上我还得到了 Walfer Fitch (Irvine), Anthony Edwards (Cambridge) 和 John Monahan (Raleigh) 的指点。本书出自我在 Raleigh 讲课时所用的一些讲稿。几位北卡罗来纳大学和 Duke 大学的学生帮助证明了我所用的一些方法。

感谢许多作者和出版社允许我引用一些未发表过的数据和享有版权的资料。

在成书过程中，我与 Sinauer Associates 一起工作始终感到很愉快。要特别致谢 Andy Sinauer，正是在他的鼓励下，我才去啃 LATEX。我学得迟缓，他对此赋予极大的耐心，因而使得这本书价格适中，并避免了导致印刷错误的一个环节。在使用这一打字软件过程中，我得到了 Chris Basten 和 Hal Caswell 的帮助，尤其是在绘图上 Chris Basten 给予了帮助。

我战战兢兢地在书末列上了一些计算机程序的源编码清单，并冒着使应用者可能得出错误结论的危险。但是，由于不断有人向我索要有关不平衡和 F—统计量程序的拷贝，这鼓励我将这些程序附上。我很高兴能够得到有关在本书中所列程序的应用价值的评论。尽管这些程序并非特别有效，但它们至少是正确的。使用者应对照我所列出的样本输入文件对其进行检验。这些程序都是用 FORTRAN/77编写的，并在微机上运行过。不过对于据此所作的任何分析，我不承担责任。

很明显，只要你打开这本书，就会感觉到 Clark Cockerman 的见解对我的影响有多大。Clark 指导了我25年，我深感荣幸。

最后，我还要感谢 Beth, Claudia 和 Henry 所给予的支持与耐心。

布鲁斯·威尔 (Bruce S. Weir)

1990年3月于北卡罗来纳

目　录

1. 引　言

1.1　遗传数据的属性

本书对离散型性状数据作了论述。这些性状包括从花朵颜色之类的形态性状到DNA序列中某个碱基的类型。不论数据的属性怎样，本书所提供的分析方法适用于分析离散型遗传单位（它们具有孟德尔单位（Mendelizing units）的特征）的观察结果。换言之，所观察的单位可由亲代向子代传递，并且该传递过程是随机的。这样的单位曾被称作基因。但是，现在已不再需要将注意力局限于那些编码蛋白质的DNA区段。可以发展一些统计理论，用以分析那些确实是基因的遗传单位，或者那些取决于基因产物的物理性质的遗传单位以及那些与编码序列毫无关系的遗传单位。

在开始讨论这一内容以前，先来看看群体遗传学中离散型数据的类型。本章列举了几种不同类型的数据。

1.2　遗传数据实例

1.2.1　表现型数据

最为人们熟知的表现型数据是孟德尔的豌豆（*Pisum sativum*）试验数据。孟德尔在一系列杂交组合中观察了7对性状，并报道了杂种自交后代的试验结果（表1.1）。根据这些结果孟德尔提出了显性与隐性表现型之比为3∶1，尽管有人对其结果与这一比例

的过分符合提出了讨论。关于这一点，在本章稍后一节中将会提到。

表 1.1 孟德尔对豌豆（*Pisum sativum*）7 个显性性状的观察结果

（孟德尔，1866）

性 状		显性表现	隐性表现
		籽粒性状	
A	种子外形	5474 圆	1850 皱
B	子叶颜色	6022 黄	2001 绿
		植株性状	
C	种皮颜色	705 灰棕色	224 白
D	豆荚形状	882 饱满	299 皱缩
E	未熟豆荚颜色	428 绿色	152 黄色
F	花着生位置	651 腋生	207 顶生
G	植株高度	787 高秆	277 矮秆

1.2.2 共显性的表现型数据

正如 Fisher（1936）在讨论孟德尔的研究工作时所指出的，显性性状有个问题，即无法区分杂合体与显性纯合体，故不能肯定亲本的基因型。而象血型这样的共显性性状，就不会引起这类问题。表 1.2 列举了 Race 等（1949）所报道的一组相当完整的共显性表现型数据。

表 1.2 Race 等（1949）报道的几个家系的 MNS 血型

序号	双亲		子女					
	父亲	母亲	1	2	3	4	5	6
1	MSMs	MsMs	MsMs	MSMs	MSMs			
2	MM.S	MsMs	MSMs	MSMs				
3	MM.S	MM.S	MM.S	MM.S	MM.S			
4	MM.S	MM.S	MM.S					
5	MM.S	MM.S	MM.S					
6	MM.S	MM.S	MM.S	MM.S	MM.S			
7	MsMs	MsNs	MsNs	MsNs				
8	MsNs	MsMs	MsMs					
9	MsNs	MsMs	MsNs					
10	MsMs	MsNs	MsMs=	=MsMs				

（续）

序号	双亲		子女					
	父亲	母亲	1	2	3	4	5	6
11	MsMs	MsNs	MsNs	MsMs	MsNs	MsMs		
12	MsNs	MSMs	MSNs—	—MsMs				
13	MSMs	MsNs	MsNs	MSNs	MsMs			
14	MsNs	MSMs	MsMs	MSNs	MSMs			
15	MM. S	MsNs	MSMs	MSMs	MSNs	MSNs	MSNs	MSNs
16	MsNs	MM. S	MSMs	MSMs=	=MSMs			
17	MM. S	MsNs	MSMs	MSMs				
18	MsNs	MM. S	MSMs=	=MSMs				
19	MsNs	MM. S	MSNs					
20	MsNs	MM. S	MSMs	MSMs				
21	MsMs	MSNs	MSMs	MsNs	MSMs	MsNs		
22	MsMs	MSNs	MSMs	MSMs	MSMs—	—MSMs		
23	MSNs	MSMs	MM. S	MsNs				
24	MSNs	MSMs	MM. S	MsNs				
25	MsNS	MSMs	MsMs					
26	MM. S	MN. S	MM. S	MN. S	MN. S			
27	MM. S	MN. S	MM. S	MN. S				
28	MM. S	MN. S	MM. S	MM. S				
29	MN. S	MM. S	MN. S	MM. S				
30	MM. S	MN. S	MN. S					
31	MM. S	MN. S	MM. S	MN. S				
32	MM. S	MN. S	MN. S	MM. S				
33	MN. S	MM. S	MN. S—	—MM. S				
34	MM. S	MN. S	MM. S	MM. S				
35	MM. S	MN. S	MM. S	MM. S	MN. S			
36	NsNs	MsMs	MsNs	MsNs				
37	NsNs	MsMs	MsNs	MsNs	MsNs	MsNs	MsNs	MsNs
38	NsNs	MsMs	MsNs	MsNs	MsNs			
39	MsMs	NsNs	MsNs					
40	MSMs	NsNs	MsNs	MsNs				
41	MSMs	NsNs	MSNs	MsNs				
42	NsNs	MSMs	MSNs	MsNs	MSNs			
43	NsNs	MM. S	MSNs					
44	NN. S	MsMs	MsNS					

（续）

序号	双亲		子女					
	父亲	母亲	1	2	3	4	5	6
45	NsNs	MSMs	MsNs	MN. S				
46	MM. S	NN. S	MN. S					
47	NN. S	MM. S	MN. S=	=MN. S				
48	MsNs	MsNs	MsNs	MsNs				
49	MsNs	MsNs	MsNs	MsMs	NsNs			
50	MsNs	MsNs	NsNs					
51	MsNs	MsNs	MsNs	NsNs				
52	MSNs	MsNs	MSNs	NsNs	MsNs			
53	MsNs	MSNs	NsNs	MSNs				
54	MsNs	MSNs	MSMs	MSNs	NsNs			
55	MSNs	MsNs	NsNs	MSNs	MSMs			
56	MsNs	MSNs	MSMs	MsNs	MSNs			
57	MSNs	MsNs	NsNs	MSNs	MSNs			
58	MSNs	MsNs	NsNs					
59	MsNs	MSNs	MSNs	MsNs	MsNs	MSMs	NsNs	
60	MsNs	MsNS	MsMs					
61	MsNs	MN. S	MN. S—	—MsNs	MN. S			
62	MsNs	MN. S	NSNs					
63	MsNs	MN. S	MSMs	MN. S				
64	MN. S	MsNs	MN. S					
65	MN. S	MsNs	MN. S					
66	MSNs	MSNs	NsNs	MSNs				
67	MSNs	MSNs	MSMS	MSNs	NsNs	MSMS	MsNs	
68	MSNs	MSNs	MSMS	NsNs				
69	MN. S	MN. S	MN. S	NSNs	MsNs	NSNs		
70	MN. S	MN. S	MN. S	MM. S				
71	MN. S	MN. S	MN. S	MN. S				
72	MN. S	MN. S	MN. S	MN. S				
73	NsNs	MsNs	NsNs					
74	MSNs	NsNs	MSNs	NsNs				
75	NsNs	MSNs	MSNs	NsNs				
76	NsNs	MSNs	NsNs	NsNs				
77	MSNs	NsNs	NsNs					
78	NsNs	MSNs	MSNs	NsNs	NsNs			