



HANDBOOK of COMPARATIVE GENOMICS

PRINCIPLES AND METHODOLOGY

比较基因组学手册

— 原理与方法

[意] C. 萨科内 (Cecilia Saccone) 著
G. 佩索莱 (Graziano Pesole)

王进 严明 等译



化学工业出版社

HANDBOOK of COMPARATIVE GENOMICS PRINCIPLES AND METHODOLOGY

比较基因组学手册 ——原理与方法

[意] C. 萨科内 (Cecilia Saccone) 著
G. 佩索莱 (Graziano Pesole)

王进 严明 等译



化学工业出版社

· 北京 ·

图书在版编目 (CIP) 数据

比较基因组学手册——原理与方法 / [意] 萨科内 (Saccone, C.), [意] 佩索莱 (Pesole, G.) 著; 王进等译. —北京: 化学工业出版社, 2008. 2

书名原文: Handbook of Comparative Genomics: Principles and Methodology
ISBN 978-7-122-02085-7

I. 比… II. ①萨…②佩…③王… III. 基因组-对比研究 IV. Q343. 1

中国版本图书馆 CIP 数据核字 (2008) 第 019452 号

Handbook of Comparative Genomics: Principles and Methodology, by Cecilia Saccone and Graziano Pesole

ISBN 0-471-39128-X

Copyright © 2003 by John Wiley & Sons, Inc. All rights reserved.

Authorized translation from the English language edition published by John Wiley & Sons, Inc.

本书中文简体字版由 John Wiley & Sons, Inc. 授权化学工业出版社独家出版发行。

未经许可，不得以任何方式复制或抄袭本书的任何部分，违者必究。

北京市版权局著作权合同登记号: 01-2005-2927

责任编辑: 傅四周 孟嘉
责任校对: 战河红

文字编辑: 张春娥
装帧设计: 关飞

出版发行: 化学工业出版社 (北京市东城区青年湖南街 13 号 邮政编码 100011)

印 刷: 北京市振南印刷有限责任公司

装 订: 三河市宇新装订厂

720mm×1000mm 1/16 印张 25 1/4 字数 569 千字 2008 年 7 月北京第 1 版第 1 次印刷

购书咨询: 010-64518888 (传真: 010-64519686) 售后服务: 010-64518899

网 址: <http://www.cip.com.cn>

凡购买本书, 如有缺损质量问题, 本社销售中心负责调换。

定 价: 65.00 元

版权所有 违者必究

谨以此书献给 Ernesto Quagliariello!

前 言

基因组时代的序幕刚刚拉开，后基因组时代即迅速来临。大规模测序实施以来，人们一直在探寻新的方法和新的工具以用于从分子水平上研究生命体，特别是对推动生物学发展的新假说进行研究。全基因组测序所带来的前景被寄予了厚望，但目前仍然存在许多问题有待解决；基因组信息的开发利用还处于萌芽阶段，需要发展新的方法学来对基因组信息进行充分利用。

一些基本的问题包括：大规模测序在多大程度上增加了我们关于物种性质和功能（或广义地看，有关基因组结构与功能）的知识？测序技术广泛应用以后出现了什么更有价值的观点或想法？人们对生物过程分子基础的认识是如何得到推进的，如果说有的话，是否是根本性的？我们在进行比较基因组研究方面有什么样的能力？可以得出哪些一般性的规律？

众所周知，比较生物学研究大多建立在类比和同源概念的基础上，由于这个原因，必须进行全基因组水平的比较，而不是仅仅对单个基因、基因家族或特定的基因组区域进行比较。与之前只能专注于局部不同，原核生物、真核生物和细胞器的全基因组测序已经有效地促进了将基因组作为整体进行结构与进化的工作。因而产生了进化基因组学这个新的领域，并将在揭示核酸、蛋白质线性结构、三维折叠、细胞遗传学、基因表达和调控途径方面产生革命性的影响。

我们相信比较进化基因组学是揭示隐藏于生命体中的信息的关键，我们使用的概念“基因组学”包括其表达和调控机制，这是所有生物过程的基础，也就是说，基因组学也包括转录组学、蛋白质组学和其他“组学”，这些概念相互衔接、相互交叉。“组学”研究需要广泛的数据库和支持包括生物学、化学、信息学、数学和物理学多学科间合作研究的动态技术。

本书旨在为具有生物学背景并试图接触现代生物学的新方向——基因组学的读者提供一本工具书，对技术开发人员和管理人员、工业界和基金机构也很有用。原则上，任何有兴趣了解这一新领域的激动人心的新应用的人都可一读。必须说明的

是，这本书只陈述个人的观点，由于涉及范围广，问题复杂，所以无法囊括所有的文献，尽管这样，我们已经尽力突出里程碑性的工作、新产生的概念、关键的方法技术和这一新领域最急需的知识。本书以描述完整测序的生物全基因组为起点，提炼从已获得的数据中涌现出来的新概念，同时也介绍用于研究完整基因组及其进化的最新方法。

总体来说，本书是作为学生和研究人员的一般性的指导书，读者不一定是专门研究基因组学的，任何人都可一读，不管是理论研究还是带有实际目的，只要有兴趣涉猎这一迷人的生物学新篇章——实际上已经成为一个新的学科领域。我们将本书分为三个部分。

第一部分描述了利用现代生物技术所获得的主要生物学过程的最新分子知识。介绍了由基因组测序带来的新认识。特别地，我们总结了完整测序的原核生物、真核生物和细胞器基因组的主要特征。

第二部分是基因组学中最新的方法学。介绍了现有实验技术和生物信息学工具，特别强调的是分子生物学技术、生物学数据库和用于序列分析的计算方法。

第三部分介绍了从比较研究中获得的结果。我们在这一部分讨论了一些基础性的前沿问题，如基因组大小的进化、碱基组成的约束以及分子层次上的生物体结构与起源。最后，我们论述了分子系统发生学的新进展。

致谢

感谢我们怀念的亲爱的朋友和同事 Giuliano Preparata，一位著名的理论学家，他将我们引入了分子进化研究的多学科探索；也热忱地感谢为本书提供过指导和建议的好几位同事和朋友，他们是 Marcella Attimonelli、Giorgio Bernardi、Rita Casiadio、Nicla Cataldo、Victor De Lorenzo、Annamaria D’ Erchia、Ilenia D’ Errico、Carmela Gissi、Alessandro Minelli、Aurelio Reyes、Teresa M. R. Regina、Elisabetta Sbisa 和 Apollonia Tullo。最后，特别感谢 Alessandra Larizza 整理文献和组织资料工作，以及 Marilina Lonigro 在语言文字方面的帮助。

随着测序技术的飞速发展，人类基因组计划于1990年启动，基因组学应运而生。基因组学是研究生物体所有遗传信息的学科，通过综合运用分子生物学、生物化学、生物物理学、数学、计算机科学等多学科的知识，去探索生命活动中基因的作用机理。

基因组学的研究对象是生物体的全部遗传信息，即基因组。基因组要从整体上研究，不能只局限于个别基因或少数几个基因，因为基因之间存在着密切的相互作用，一个基因的表达可能影响另一个基因的表达，一个基因的突变可能影响另一个基因的表达。

目 录

上篇 基因组特性	
第1章 原核生物	3
1.1 引言	3
1.2 形态和分类	9
1.3 基因组的形状和大小	12
1.4 基因数目和组成方式	15
1.5 碱基组成	19
1.6 密码子的使用	26
1.7 复制和表达	28
第2章 真核生物	34
2.1 引言	34
2.2 真核生物的分类与时间尺度	34
2.3 基因组的形状和大小	37
2.4 碱基组分	42
2.5 复制、修复和重组	44
2.6 基因表达	48
2.6.1 转录与转录后调控	48
2.6.2 遗传密码和密码子的使用	54
2.6.3 翻译与翻译后修饰	55
2.7 完整测序的真核生物基因组	58
2.7.1 酿酒酵母基因组	60
2.7.2 栗酒裂殖酵母基因组	63
2.7.3 线虫基因组	63

2.7.4 黑腹果蝇基因组	65
2.7.5 拟南芥基因组	67
2.7.6 水稻基因组	69
2.7.7 人基因组	70
第3章 细胞器	77
3.1 线粒体	77
3.1.1 一般结构和功能	77
3.1.2 DNA与遗传系统	79
3.1.3 基因组的特征	88
3.2 叶绿体和其他质体	114
中篇 方法学	115
第4章 基因组学中的分子生物学技术	121
4.1 基因组DNA测序	121
4.1.1 DNA测序技术	121
4.1.2 人类基因组计划	124
4.2 转录组分析	125
4.2.1 基因表达分析	125
4.2.2 表达序列标签	126
4.2.3 基因表达序列分析	127
4.2.4 差异显示	129
4.2.5 表现度示差分析	130
4.2.6 DNA微阵列	131
4.3 蛋白质组分析	137
4.3.1 二维凝胶电泳	137
4.3.2 蛋白质鉴定	139
4.3.3 蛋白质-DNA和蛋白质-蛋白质相互作用的研究	141
4.3.4 生物芯片法分析蛋白质组	143
第5章 基因组时代的生物学数据库	145
5.1 引言	145
5.2 基础数据库和专业化的数据库	145
5.3 数据库结构	148
5.4 数据库链接与互通性	150
5.5 数据库注释	151
5.6 检索系统	154
5.6.1 SRS	154
5.6.2 Entrez	155

5.6.3 其他检索系统	155
5.7 核酸数据库	155
5.8 蛋白质数据库	156
5.9 其他蛋白质数据库	156
5.10 基因组数据库和资源	159
5.11 基因数据库与资源	163
5.12 转录组数据库	164
5.13 代谢组数据库	165
5.14 突变数据库	167
5.15 线粒体数据库和资源	168
第6章 基因组序列分析的计算方法	171
6.1 引言	171
6.2 点阵图	172
6.3 两序列比对	175
6.3.1 Needleman-Wunsch 全局比对算法	175
6.3.2 Smith-Waterman 算法	176
6.3.3 cDNA 以及基因组 DNA 序列的比对	179
6.3.4 基因组比对	179
6.3.5 清除序列库中的冗余序列	181
6.3.6 同源序列相似度的计算	185
6.4 数据库搜索	189
6.4.1 FASTA	190
6.4.2 BLAST	192
6.4.3 BLAST 和 FASTA 程序包	198
6.4.4 过滤不需要的序列匹配块	200
6.4.5 重复序列匹配的过滤	201
6.4.6 比对分值的统计学意义	201
6.5 多序列比对	205
6.6 可识别远源相关蛋白或蛋白质模块的比对谱	209
6.7 序列的组装方法	213
6.7.1 序列的净化	213
6.7.2 序列的聚类	214
6.7.3 构建一致序列	215
6.7.4 用电子 PCR 绘制序列图谱	217
6.7.5 基因组和 EST 项目的序列组装	217
6.7.6 构建用于基因索引的组装序列	219
6.8 生物序列的语言学分析	222
6.8.1 马尔可夫链式生物序列	224

6.8.2	生物序列语言的复杂度	225
6.8.3	基因组重复序列的识别	228
6.8.4	生物序列中的模式搜索	229
6.8.5	识别染色体序列中的启动子区域	234
6.8.6	发掘识别基因调控元件和蛋白质模块组件的模式	237
6.8.7	基因预测	239
6.8.8	基因组序列中 CpG 岛的识别	243
6.8.9	密码子使用策略分析	244
6.9	RNA 二级结构预测	245
6.10	蛋白质序列分析	252
6.10.1	蛋白质一级序列的分析	253
6.10.2	预测跨膜蛋白螺旋	258
6.10.3	蛋白质信号肽的识别及亚细胞定位的预测	259
6.10.4	蛋白质二级结构的预测	265
6.10.5	预测卷曲螺旋和螺旋-转角-螺旋结构	270
6.10.6	蛋白质三级结构预测	270
6.10.7	蛋白质折叠的识别与分类	273
6.10.8	蛋白质功能预测的基因组比较工具	273
6.11	进化与系统发育分析	274
6.11.1	同源序列间遗传距离的评价	276
6.11.2	分子系统发育学	278

下篇 比较基因组学

第 7 章	分子进化	293
7.1	引言	293
7.2	基因组尺寸的进化	294
7.3	碱基组分在进化中的作用	297
7.4	原核基因组的进化	301
7.5	从原核生物到真核生物	302
7.5.1	真核细胞的起源	302
7.5.2	线粒体基因组的进化	304
7.5.3	质体的起源与进化	309
7.6	从单细胞到多细胞	312
7.7	核基因组的进化	312
7.7.1	内含子	313
7.7.2	基因数目和蛋白质数目	313
7.7.3	非编码元件	314

7.7.4	基因家族的扩展	315
7.7.5	基因组倍增	320
7.7.6	结论	321
第8章 分子系统发育学		322
8.1	引言	322
8.2	分子钟	322
8.3	相似性测量：直系同源与旁系同源	324
8.4	基因组学时代的分子系统发育学	326
8.5	远源物种间的相互关系：进化树	328
8.6	后生动物的系统发育	329
8.6.1	细胞器分类学与核分类学	329
8.6.2	哺乳动物系统发育	330
附录 本书引用的 URL		333
参考文献		335
索引		384
译后记		390

生物学基因基础

8.8	人类基因组计划	337
8.9	线虫基因组	338
8.10	酵母基因组	339
8.11	果蝇基因组	340
8.12	拟南芥基因组	341
8.13	水稻基因组	342
8.14	小麦基因组	343
8.15	玉米基因组	344
8.16	高粱基因组	345
8.17	大豆基因组	346
8.18	甘蓝基因组	347
8.19	拟南芥基因组	348
8.20	拟黑果蝇基因组	349
8.21	斑马鱼基因组	350
8.22	小鼠基因组	351
8.23	大鼠基因组	352
8.24	狗基因组	353
8.25	牛基因组	354
8.26	山羊基因组	355
8.27	绵羊基因组	356
8.28	猪基因组	357
8.29	马基因组	358
8.30	驴基因组	359
8.31	兔基因组	360
8.32	狗基因组	361
8.33	狼基因组	362
8.34	牛基因组	363
8.35	山羊基因组	364
8.36	绵羊基因组	365
8.37	猪基因组	366
8.38	马基因组	367
8.39	驴基因组	368
8.40	兔基因组	369
8.41	狗基因组	370
8.42	狼基因组	371
8.43	牛基因组	372
8.44	山羊基因组	373
8.45	绵羊基因组	374
8.46	猪基因组	375
8.47	马基因组	376
8.48	驴基因组	377
8.49	兔基因组	378
8.50	狗基因组	379
8.51	狼基因组	380
8.52	牛基因组	381
8.53	山羊基因组	382
8.54	绵羊基因组	383
8.55	猪基因组	384
8.56	马基因组	385
8.57	驴基因组	386
8.58	兔基因组	387
8.59	狗基因组	388
8.60	狼基因组	389
8.61	牛基因组	390
8.62	山羊基因组	391
8.63	绵羊基因组	392
8.64	猪基因组	393
8.65	马基因组	394
8.66	驴基因组	395
8.67	兔基因组	396
8.68	狗基因组	397
8.69	狼基因组	398
8.70	牛基因组	399
8.71	山羊基因组	400
8.72	绵羊基因组	401
8.73	猪基因组	402
8.74	马基因组	403
8.75	驴基因组	404
8.76	兔基因组	405
8.77	狗基因组	406
8.78	狼基因组	407
8.79	牛基因组	408
8.80	山羊基因组	409
8.81	绵羊基因组	410
8.82	猪基因组	411
8.83	马基因组	412
8.84	驴基因组	413
8.85	兔基因组	414
8.86	狗基因组	415
8.87	狼基因组	416
8.88	牛基因组	417
8.89	山羊基因组	418
8.90	绵羊基因组	419
8.91	猪基因组	420
8.92	马基因组	421
8.93	驴基因组	422
8.94	兔基因组	423
8.95	狗基因组	424
8.96	狼基因组	425
8.97	牛基因组	426
8.98	山羊基因组	427
8.99	绵羊基因组	428
8.100	猪基因组	429
8.101	马基因组	430
8.102	驴基因组	431
8.103	兔基因组	432
8.104	狗基因组	433
8.105	狼基因组	434
8.106	牛基因组	435
8.107	山羊基因组	436
8.108	绵羊基因组	437
8.109	猪基因组	438
8.110	马基因组	439
8.111	驴基因组	440
8.112	兔基因组	441
8.113	狗基因组	442
8.114	狼基因组	443
8.115	牛基因组	444
8.116	山羊基因组	445
8.117	绵羊基因组	446
8.118	猪基因组	447
8.119	马基因组	448
8.120	驴基因组	449
8.121	兔基因组	450
8.122	狗基因组	451
8.123	狼基因组	452
8.124	牛基因组	453
8.125	山羊基因组	454
8.126	绵羊基因组	455
8.127	猪基因组	456
8.128	马基因组	457
8.129	驴基因组	458
8.130	兔基因组	459
8.131	狗基因组	460
8.132	狼基因组	461
8.133	牛基因组	462
8.134	山羊基因组	463
8.135	绵羊基因组	464
8.136	猪基因组	465
8.137	马基因组	466
8.138	驴基因组	467
8.139	兔基因组	468
8.140	狗基因组	469
8.141	狼基因组	470
8.142	牛基因组	471
8.143	山羊基因组	472
8.144	绵羊基因组	473
8.145	猪基因组	474
8.146	马基因组	475
8.147	驴基因组	476
8.148	兔基因组	477
8.149	狗基因组	478
8.150	狼基因组	479
8.151	牛基因组	480
8.152	山羊基因组	481
8.153	绵羊基因组	482
8.154	猪基因组	483
8.155	马基因组	484
8.156	驴基因组	485
8.157	兔基因组	486
8.158	狗基因组	487
8.159	狼基因组	488
8.160	牛基因组	489
8.161	山羊基因组	490
8.162	绵羊基因组	491
8.163	猪基因组	492
8.164	马基因组	493
8.165	驴基因组	494
8.166	兔基因组	495
8.167	狗基因组	496
8.168	狼基因组	497
8.169	牛基因组	498
8.170	山羊基因组	499
8.171	绵羊基因组	500
8.172	猪基因组	501
8.173	马基因组	502
8.174	驴基因组	503
8.175	兔基因组	504
8.176	狗基因组	505
8.177	狼基因组	506
8.178	牛基因组	507
8.179	山羊基因组	508
8.180	绵羊基因组	509
8.181	猪基因组	510
8.182	马基因组	511
8.183	驴基因组	512
8.184	兔基因组	513
8.185	狗基因组	514
8.186	狼基因组	515
8.187	牛基因组	516
8.188	山羊基因组	517
8.189	绵羊基因组	518
8.190	猪基因组	519
8.191	马基因组	520
8.192	驴基因组	521
8.193	兔基因组	522
8.194	狗基因组	523
8.195	狼基因组	524
8.196	牛基因组	525
8.197	山羊基因组	526
8.198	绵羊基因组	527
8.199	猪基因组	528
8.200	马基因组	529
8.201	驴基因组	530
8.202	兔基因组	531
8.203	狗基因组	532
8.204	狼基因组	533
8.205	牛基因组	534
8.206	山羊基因组	535
8.207	绵羊基因组	536
8.208	猪基因组	537
8.209	马基因组	538
8.210	驴基因组	539
8.211	兔基因组	540
8.212	狗基因组	541
8.213	狼基因组	542
8.214	牛基因组	543
8.215	山羊基因组	544
8.216	绵羊基因组	545
8.217	猪基因组	546
8.218	马基因组	547
8.219	驴基因组	548
8.220	兔基因组	549
8.221	狗基因组	550
8.222	狼基因组	551
8.223	牛基因组	552
8.224	山羊基因组	553
8.225	绵羊基因组	554
8.226	猪基因组	555
8.227	马基因组	556
8.228	驴基因组	557
8.229	兔基因组	558
8.230	狗基因组	559
8.231	狼基因组	560
8.232	牛基因组	561
8.233	山羊基因组	562
8.234	绵羊基因组	563
8.235	猪基因组	564
8.236	马基因组	565
8.237	驴基因组	566
8.238	兔基因组	567
8.239	狗基因组	568
8.240	狼基因组	569
8.241	牛基因组	570
8.242	山羊基因组	571
8.243	绵羊基因组	572
8.244	猪基因组	573
8.245	马基因组	574
8.246	驴基因组	575
8.247	兔基因组	576
8.248	狗基因组	577
8.249	狼基因组	578
8.250	牛基因组	579
8.251	山羊基因组	580
8.252	绵羊基因组	581
8.253	猪基因组	582
8.254	马基因组	583
8.255	驴基因组	584
8.256	兔基因组	585
8.257	狗基因组	586
8.258	狼基因组	587
8.259	牛基因组	588
8.260	山羊基因组	589
8.261	绵羊基因组	590
8.262	猪基因组	591
8.263	马基因组	592
8.264	驴基因组	593
8.265	兔基因组	594
8.266	狗基因组	595
8.267	狼基因组	596
8.268	牛基因组	597
8.269	山羊基因组	598
8.270	绵羊基因组	599
8.271	猪基因组	600
8.272	马基因组	601
8.273	驴基因组	602
8.274	兔基因组	603
8.275	狗基因组	604
8.276	狼基因组	605
8.277	牛基因组	606
8.278	山羊基因组	607
8.279	绵羊基因组	608
8.280	猪基因组	609
8.281	马基因组	610
8.282	驴基因组	611
8.283	兔基因组	612
8.284	狗基因组	613
8.285	狼基因组	614
8.286	牛基因组	615
8.287	山羊基因组	616
8.288	绵羊基因组	617
8.289	猪基因组	618
8.290	马基因组	619
8.291	驴基因组	620
8.292	兔基因组	621
8.293	狗基因组	622
8.294	狼基因组	623
8.295	牛基因组	624
8.296	山羊基因组	625
8.297	绵羊基因组	626
8.298	猪基因组	627
8.299	马基因组	628
8.300	驴基因组	629
8.301	兔基因组	630
8.302	狗基因组	631
8.303	狼基因组	632
8.304	牛基因组	633
8.305	山羊基因组	634
8.306	绵羊基因组	635
8.307	猪基因组	636
8.308	马基因组	637
8.309	驴基因组	638
8.310	兔基因组	639
8.311	狗基因组	640
8.312	狼基因组	641
8.313	牛基因组	642
8.314	山羊基因组	643
8.315	绵羊基因组	644
8.316	猪基因组	645
8.317	马基因组	646
8.318	驴基因组	647
8.319	兔基因组	648
8.320	狗基因组	649
8.321	狼基因组	650
8.322	牛基因组	651
8.323	山羊基因组	652
8.324	绵羊基因组	653
8.325	猪基因组	654
8.326	马基因组	655
8.327	驴基因组	656
8.328	兔基因组	657
8.329	狗基因组	658
8.330	狼基因组	659

上篇 基因组特性

第1章

原核生物

1.1 引言

就在我们编写本书的同时，人们正以每月两个细菌基因组的速度完成测序工作，所产生的原核生物基因组序列可以从美国国家生物技术信息中心（NCBI）的网站上获得。本章涉及的数据都是已完成测序的全基因组序列，见表 1.1。很显然，到本书面世的时候，肯定会有更多的基因组完成了测序，并且某些观点也可能随之而发生改变。可以想象，如果技术取得巨大进步的话，我们目前所拥有的知识体系将会发生革命性的变化。

表 1.1 列出了迄今为止已完成测序的原核生物基因组，包括它们的名称和属性，如 EMBL 数据库索引号、大小、形状、是否存在染色体外遗传因子以及参考文献。从表中可以看出选择测序物种的原因是多种多样的，主要取决于它在基础科学或应用科学中的地位，如：对系统发育、代谢机制（主要是古细菌）研究的重要性，作为人类或动物病原体以及用于工业化制酶的重要性。换句话说，优先对那些目前熟知或者应用前景广泛的物种进行测序，而从系统发育的角度看，这种选择是非常随意的。

我们当前正处于基因组时代的起步阶段，尽管完成测序的物种数目非常少，但它们已经显示出惊人的迹象。本章将介绍测序方面的重要成就，这些成就为我们提供了原核生物基因组的新知识，也促进了相关领域研究方法的发展。本章还包括原核生物的一些基本知识，如形态、分类，以及遗传物质的结构、复制和表达等主要性质。当然，这些描述较为简短，因为我们侧重的是全基因组测序方面的重要问题。希望读者进一步参考该领域更深入的研究工作以及各种综述和论文。

表 1.1 已完成测序的原核生物基因组

物种	染色体	索引号	大小/bp	索引号	染色体外遗传因子 大小/bp	文献	
						古细菌 (Archaea)	
敏捷好氧热球菌 (<i>Aeropyrum pernix</i>)	BA000002	1669695				Kawarabayasi, Hino 等 (1999)	
闪烁古球菌 (<i>Archaeoglobus fulgidus</i>)	AE000782	2178400				Klenk, Clayton 等 (1997)	
盐杆菌 NRC-1 (3 个染色体) (<i>Halobacterium sp. NRC-1</i>)	AE004437	2014239		AF016485	191346	Ng, Kennedy 等 (2000)	
	AE004438	365425		AE004438	365425		
	AF016485	191346					
热自养甲烷杆菌 (<i>Methanobacterium thermoautotrophicum</i>)	AE006666	1751377				Amith, Doucette-Stamm 等 (1997)	
詹氏甲烷球菌 (<i>Methanococcus jannaschii</i>)	L77117	1664970	177118		58407	Bult, White 等 (1996)	
			177119		16550		
坎氏甲烷球菌 AV19 (<i>Methanococcus kandleri</i> AV19)	AE009439	1694969				Slesarev, Mezhevaya 等 (2002)	
嗜乙酸甲烷八叠球菌 C2A (<i>Methanosaerina acetivorans</i> str. C2A)	AE010299	5751492				Galagan, Nusbaum 等 (2002)	
马氏甲烷八叠球菌 Goel (<i>Methanosaerina mazei</i> Goel)	AE008384	4096345				Deppenmeier, Johann 等 (2002)	
嗜气热棒菌 (<i>Pyrobactrum aerophilum</i>)	AE009441	2222430				Fitz-Gibbon, Ladner 等 (1998)	
深海热球菌 (<i>Pyrococcus abyssi</i>)	AL096836	1765118				Lecompte, Ripp 等 (2001)	
强烈热球菌 DSM36338 (<i>Pyrococcus furiosus</i> DSM36338)	AE009950	1908256				Robb, Maeder 等 (2001)	
掘越热球菌 (<i>Pyrococcus horikoshii</i>)	AP000001-AP000007	1738595				Kawarabayasi, Sawada 等 (1998)	
硫黄矿硫化叶菌 (<i>Sulfolobus solfataricus</i>)	AE006641	2992245				She, Singh 等 (2001)	
常田硫化叶菌 (<i>Sulfolobus tokodaii</i>)	BA000023	2694765		AJ010405	41229	Kawarabayasi, Hino 等 (2001)	
嗜酸热原体 (<i>Thermoplasma acidophilum</i>)	AL445063-	1564905				Ruepp, Graml 等 (2000)	
火山热原体 (<i>Thermoplasma volcanium</i>)	AL445067						Kawashima, Amano 等 (2000)
	AP009911-AP00996	1584799					

表
续

物 种	种加词	拉丁学名	染 色 体				文 献	
			索引号	大 小/bp	索引号	大 小/bp	染色体外遗传因子	大小/bp
细菌(Bacteria)								
根瘤土壤杆菌 C58 (Cereon) [Agrobacterium tumefaciens str. C58 (Cereon)]		AE007869	2841581					Goodner, Hinkle 等(2001)
根瘤土壤杆菌 C58 (U. Washington) [Agrobacterium tumefaciens str. C58 (U. Washington)]		AE008688	2841490	AE008687	542780			Wood, Setubal 等(2001)
				AE008690	214234			
				AE006667	39456			Deckert, Warren 等(1998)
风产液菌 (<i>Aquifex aeolicus</i>)		AE000657	1551335					Takami, Nakasone 等(2000)
耐盐芽孢杆菌 (<i>Bacillus halodurans</i>)		BA000004	4202353					Kunst, Ogasawara 等(1997)
枯草芽孢杆菌 (<i>Bacillus subtilis</i>)		AL009126	4214814					Fraser, Casjens 等(1997);
伯氏疏螺旋体 (<i>Borrelia burgdorferi</i> ①)		AE000783	910725	AE000791	9386			Casjens, Palmer 等(2000)
乙酸杆菌 F-13 (<i>Vibrio Fischeri</i> F-13)				AE000792	26498			
				AE001575	30750			
				AE001576	30223			
				AE001577	30299			
				AE001578	29838			
				AE001579	30800			
				AE001580	30885			
				AE001581	30651			
				AE001583①	5228			
				AE000793①	16823			
				AE001582①	18753			
				AE000785①	24177			
				AE000794①	26921			
				AE000786①	29766			
				AE000784①	28601			
				AE000789①	27323			
				AE000788①	36849			
				AE000787①	38829			
				AE000790①	53561			
				AE001584①	52971			

续表

物种	索引号	染色体		染色体外遗传因子		文献
		大小/bp	索引号	大小/bp	索引号	
细菌(Bacteria)						
马尔他布鲁菌(<i>Brucella melitensis</i>)	AEO08917	2117144				DelVecchio, Kapatral 等(2002)
巴克纳菌 APS (<i>Buchnera</i> sp. APS)	AP000398	640681	AP001070	7258		Shigenobu, Watanabe 等(2000)
空肠弯曲杆菌(<i>Campylobacter jejuni</i>)	AL111168	1641481	AP001071	7786		Parkhill, Wren 等(2000)
新月柄杆菌(<i>Caulobacter crescentus</i>)	AE005673	4016947				Nierman, Feldblyum 等(2001)
肺炎衣原体 AR39(<i>Chlamydia pneumoniae</i> AR39)	AE002161	1229853	JB1039	4524		Read, Brunham 等(2000)
肺炎衣原体 CWL029(<i>Chlamydia pneumoniae</i> CWL029)	AE001363	1230230	JB1040			Kalman, Mitchell 等(1999)
沙眼衣原体 MoPn(<i>Chlamydia trachomatis</i> MoPn)	AE002160	1069412	AE002162	7501		Read, Brunham 等(2000)
沙眼衣原体血清型 D(<i>Chlamydia trachomatis</i> serovar D)	AE001273	1042519				Stephens, Kalman 等(1998)
肺炎嗜性衣原体 J138(<i>Chlamydophila pneumoniae</i> J138)	BA000008	1225655				Shirai, Hirakawa 等(2000)
丙酮丁醇梭菌(<i>Clostridium acetobutylicum</i>)	AE001437	3940880	NC_001988	192000		Nolling, Breton 等(2001)
产气荚膜梭菌(<i>Clostridium perfringens</i>)	BA000016	3031430	JB1041			Shimizu, Ohtani 等(2002)
谷氨酸棒杆菌(<i>Corynebacterium glutamicum</i>)	AX114121	3309400	AE001826	177466		Tauch, Hombann 等(2002)
耐辐射异常球菌 RI(2个染色体)[<i>Deinococcus radiodurans</i> RI (2 chromosomes)]	AE000513	2648638	AE001827	45704		White, Eisen 等(1999)
大肠杆菌 K-12 (<i>Escherichia coli</i> K-12)	U00096	4639221	JB1042			Blattner, Plunkett 等(1997)
大肠杆菌 O157 : H7 EDL933 (<i>Escherichia coli</i> O157 : H7 EDL933)	AE005174	5528970	AB011549	92721		Perna, Plunkett 等(2001)
大肠杆菌 O157 : H7 Sakai (<i>Escherichia coli</i> O157 : H7 Sakai)	BA000007	5498450	AB011548	3306		Hayashi, Makino 等(2001)
具核梭杆菌且核亚种 ATCC25586(<i>Fusobacterium nucleatum</i> subsp. <i>nucleatum</i> ATCC25586)	AE009951	2174500	JB1043			Kapatral, Anderson 等(2002)
流感嗜血杆菌 Rd (<i>Haemophilus influenzae</i> Rd)	L42023	1830138	JB1044			Fleischmann, Adams 等(1995)
幽门螺杆菌 26695 (<i>Helicobacter pylori</i> 26695)	AE000511	1667867	JB1045			Tomb, White 等(1997)
幽门螺杆菌 J99 (<i>Helicobacter pylori</i> J99)	AE001439	1643831	JB1046			Alm, Ling 等(1999)
乳酸球菌乳亚种(<i>Lactococcus lactis</i> subsp. <i>lactis</i>)	AE005176	2365589				Bolotin, Wincker 等(2001)
无害李斯特菌 (<i>Listeria innocua</i>)	AL592022	3011208	AL592102	81900		Glaser, Frangeul 等(2001)
单核细胞增生李斯特菌 EGDe- <i>Listeria monocytogenes</i> EGDe	NC_003210	2944528				Glaser, Frangeul 等(2001)
百脉根中慢生根瘤菌 (<i>Mesorhizobium loti</i>)	BA000012	7036074	AP003015-16	351911		Kaneko, Nakamura 等(2000)
			AP003017	208315		