

上海科普创作出版专项资金资助

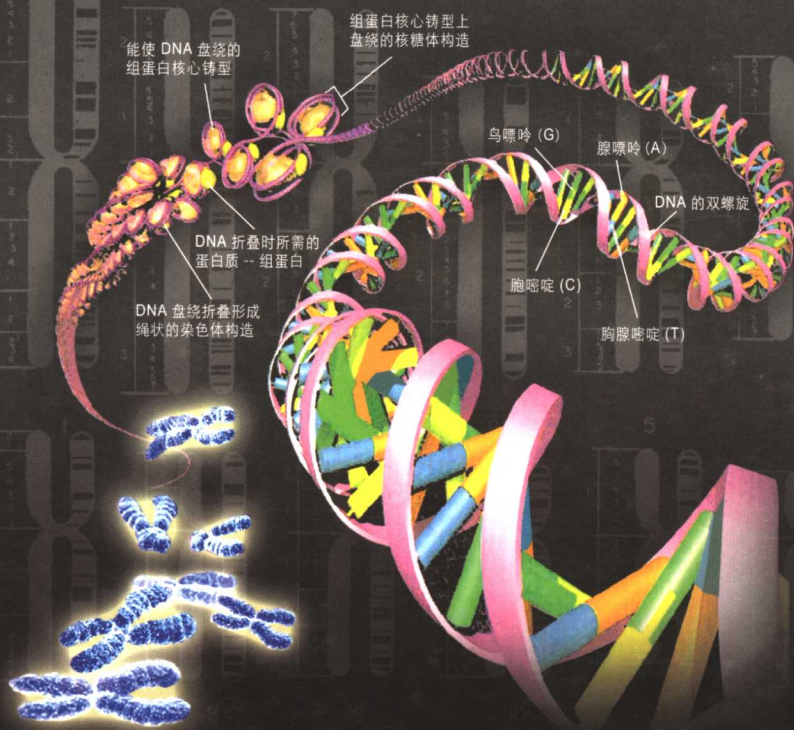
基因宝库丛书

谈家桢 主编

上海市农业生物基因中心 编

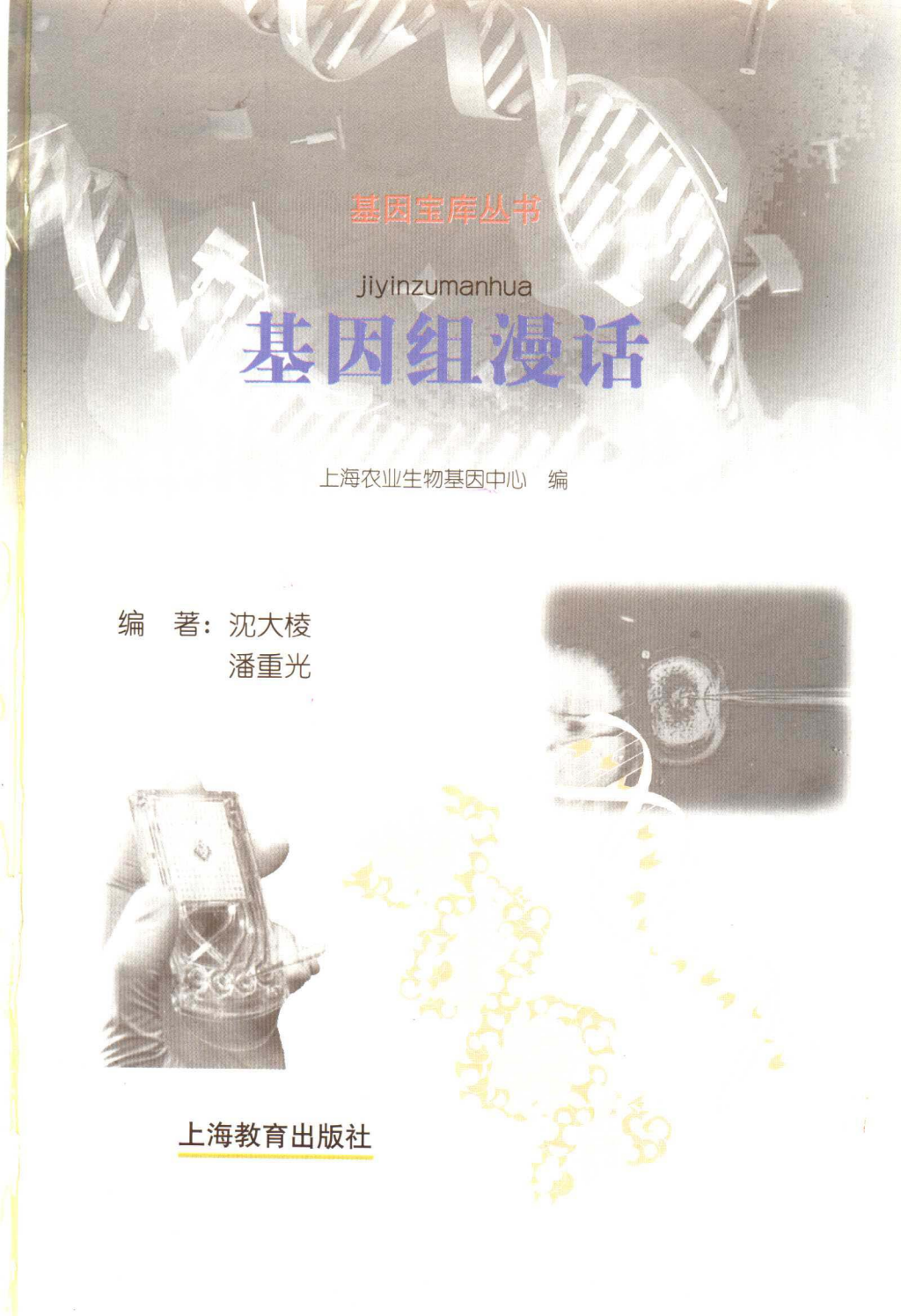
基因组漫话

沈大棱 潘重光 >>> 编著



上海教育出版社

SHANGHAI EDUCATION PUBLISHING HOUSE



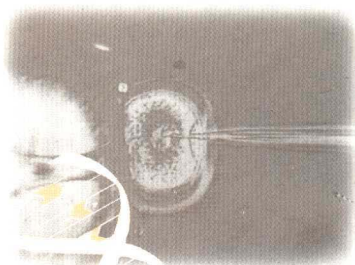
基因宝库丛书

Jiyinzumanhua

基因组漫话

上海农业生物基因中心 编

编 著：沈大棱
潘重光



上海教育出版社

图书在版编目 (C I P) 数据

基因组漫话 / 沈大棧, 潘重光编著. —上海: 上海教育出版社, 2005. 12

(基因宝库丛书 / 谈家桢主编)

ISBN 7-5444-0529-X

I. 基... II. ①沈...②潘... III. 基因组—青少年读物 IV. Q343.1-49

中国版本图书馆CIP数据核字 (2005) 第154935号

基因宝库丛书

基因组漫话

谈家桢 主编

上海世纪出版集团 出版发行
上海教育出版社

易文网: www.ewen.cc

(上海永福路 123 号 邮政编码: 200031)

各地新华书店经销 上海新华印刷有限公司印刷

开本 889×1194 1/32 印张 4.25 字数 78,000

2005 年 12 月第 1 版 2005 年 12 月第 1 次印刷

印数 1-3,000 本

ISBN 7-5444-0529-X/Q·0007 定价: 16.00 元

(如发生质量问题, 读者可向工厂调换)

主 编：谈家桢

副主编：吴爱忠 罗利军

编 委：沈大棱 林榕辉

袁正守 潘重光

（按姓氏笔划）

编辑策划：肖征波 吴延恺

序



年初，上海农科院吴爱忠教授和上海市农业生物基因中心罗利军教授告诉我，上海市科委和科协将设专项基金资助科技工作者撰写科普书籍。他们打算组织长期从事教育和科技工作的专家编写基因科学丛书，定名为“基因宝库”。我认为科委和科协的决定及两位教授的打算很有意义。向公众传播科学知识，无疑能提高劳动者的科技素质，促进先进生产力的发展。

生命科学自上世纪50年代进入分子生物学时代以来，基因科学突飞猛进，新概念、新名词日新月异，与时俱增。基因也成为运用次数最多的字眼之一。但由于基因科学既包含遗传、变异，个体、群体，分子、细胞，基因、环境，核酸、蛋白质等诸多矛盾的统一，基因科学又与国计民生关系十分密切，丰衣足食、安居乐业、健康长寿、天下太平都离不开基因科学。因此要较全面地了解基因科学知识及基因科学在工业、农业、医学等诸多方面的应用价值，实非易事。组织专家编写普及基因科学的系列丛书，无疑又是先进文化发展的需要，我

是非常支持的。

自我国取得抗击SARS的初步胜利后，吴爱忠、罗利军两位教授委托上海交大潘重光教授转告我，市科委、科协已正式同意资助“基因宝库”的编写，我很高兴。我因年迈已不能亲自参加丛书的编写，但我很乐意做力所能及的事。我托潘重光同志转告吴、罗两位教授，编写“基因宝库”丛书是一件很有意义的事，希望在编写过程中，特别要重视科学性，在保证科学性的基础上，应该积极探索趣味性和可读性，努力把“基因宝库”编成公众喜欢阅读的丛书。

谈家桢

2003年10月9日



目 录

前言	1
基因	3
遗传因子与基因	3
染色体与基因	5
DNA 与基因	8
基因组	19
基因组的类别	20
C 值与 C 值悖理	22
基因组计划	26
遗传作图	34
连锁现象	35
连锁与染色体	36
连锁与交换	38
三点测交试验	42
系谱分析	48
人鼠细胞融合与辐射	50
DNA 标记作遗传图	52
物理作图	60
限制酶切位点作图	60
荧光原位分子杂交 (FISH 技术)	61
最有用的物理图标记——STS	64
DNA 克隆文库	67
人类基因组的物理图和综合图	74

基因组测序	76
DNA 测序方法	76
DNA 序列的连续组装	83
人类基因组计划的测序	97
搜寻 DNA 序列中的基因	100
寻找阅读框	100
筛查同源性序列	104
基因定位技术与基因功能研究	105
基因定位的实验技术	105
基因功能的研究	109
从基因组到细胞	121
转录组分析	121
蛋白质组分析	123
搜寻基因的成绩	125

前 言

人类基因组被一些科学家称为“生命之书”，因为基因组中包含着有关人类的喜怒哀乐、高矮胖瘦、生老病死、智力程度等全部生命活动的奥秘。

探索人类的生命奥秘，是人类最为雄心勃勃的研究，对人类基因组开展全面研究是人类在科学史上的一大壮举，它首次将全世界人民团结起来，共同来做一件对整个人类具有深远意义的事。

2000年6月26日，从北京到东京、从巴黎到柏林、从伦敦到华盛顿，参与人类基因组计划的六国政府和有关科学家分别以不同的方式，宣布了人类基因组工作草图绘制成功，国际社会对这一重大科学进展给予高度评价，一致呼吁共享这一全人类的共同财富。

时任美国总统的克林顿在白宫举行的庆祝仪式上表示，人类基因组草图是迄今“人类所绘制的最为奇妙的图谱”，英国首相布莱尔也说“这是21世纪第一项伟大的科技成就”，是“医学科学领域的一场革命，其意义远超过抗生素的发现”。

基因科学的专家们认为，根据基因组研究的成果，人类可了解自身体质的弱点和对某种疾病的易感性，从

而为不同人群研制特殊的药物。根据基因组研究的成果，人类可以掌握那些专门控制衰老的基因，从而延长人的平均寿命，实现长命百岁的愿望。

基因组研究掀起了世界性热潮，到2000年6月28日，时任中国国家主席的江泽民的讲话将这股热潮推到了阶段性的顶峰。他在中共中央思想政治工作会议上特别对人类基因组工作框架草图的绘制成功作了简要的总结，他指出：“基因组计划是人类科学史上伟大的科学工程，它对于人类认识自身，推动生命科学、医学，以及制药业的发展具有极其重要的意义。经全球科学家的共同努力，人类基因组序列的‘工作框架图’已经绘就，这是该计划实施过程中的一个里程碑。人类基因组序列是全人类的共同财富，应该用来为全人类造福。”

基 因

现今，“基因”这个词已走进千家万户。人类为了自身的丰衣足食，必须了解自然，揭示生命发生发展的规律；为了自身的健康长寿，渴望了解自我，了解自身的生命活动奥秘，破译人的遗传机密。

机密也好，规律也好，都离不开“基因”这个迷人的字眼。

遗传因子与基因

遗传因子是出生在奥地利布隆（现为捷克的布尔诺）的孟德尔提出来的，他在修道院中作为一个传经布道的神甫时，就选用豌豆为研究材料进行了大量的杂交试验，根据对试验结果的分析 and 归类，发现了杂种的显性现象和分离规律，以及自由组合定律。



孟德尔

他提出豌豆的每个性状都是由一对遗传因子决定的。由于每个性状都有两种不同的表现形式，如花色这个性状就有红与白的不同表现，株高这个性状也有高与矮的差别，当同一性状的两种不同类型的豌豆杂交时，杂种却只表现出一个亲本的性状，另一个亲本的性状仿佛在杂种一代中不见了。在后来的研究中发现，杂种一代所表现出的那个亲本的性状是显性性状，没有表现的那个亲本的性状就称为隐性性状。

由于同一个性状有显性与隐性之分，因此决定同一性状的遗传因子也有显性因子和隐性因子之别。所有遗传因子像一粒粒的骰子或一张张的扑克牌，而绝不像一滴滴的水，两滴水碰在一起就会混合起来，而放在一起的两张牌依然保持着各自的个性，红桃还是红桃，黑桃还是黑桃。

孟德尔是在1858~1865年间提出遗传因子这一概念的，在他提出遗传因子、分离定律和自由组合定律后的几十年间，正统科学家们对这位神甫的结论不屑一顾，许多人不相信一个投身宗教的人会发现什么科学规律。直到1900年，孟德尔的超前发现才受到科技界的重视，大概是受到“遗传因子是决定生物体性状的基本原因之一”的思想启发吧，丹麦学者约翰逊就提出了用“基因”代替孟德尔的“遗传因子”。约翰逊认为，“基因”是一个很好使用，且很容易限制的字眼组合。从此以后，基因就代替了孟德尔文章中的“遗传因子”。但除了“遗传因子”这一叫法被“基因”代替外，孟德尔关于“遗

传因子”的其他假定都没有变。例如孟德尔提出：一个性状由两个“遗传因子”所决定，在后人的文献中仅将“基因”替代了“遗传因子”，即一个性状由两个“基因”决定；又如决定同一性状的“遗传因子”有显性与隐性的区别，同样，这一假定也仅被替换为决定同一性状的两个“基因”有显性和隐性的区别等等。

根据孟德尔的假定：决定同一性状的两个“遗传因子”称为等位的“遗传因子”，等位的“遗传因子”在生物体产生生殖细胞时一定要分开并各自进入不同的生殖细胞中。用“基因”代替“遗传因子”后，那就成为：决定同一性状的两个“基因”（一对“基因”）称等位“基因”，等位“基因”在生物体形成生殖细胞时一定要分开并各自进入不同的生殖细胞中。

可见，孟德尔关于“遗传因子”的研究和结论对以后生命科学的发展有着非常重要和深远的意义。

染色体与基因

染色体是细胞核内能被一种碱性的红色染料染上颜色的物质。刚开始时，这种位于细胞核内的物质被称为染色质。后来科学家发现这种能染上颜色的物质，在细胞一分为二的过程中会聚集起来，成为有一定结构和一定长度的物体。在细胞停止分裂时，细胞核中的这种物质被称为染色质，而当细胞进入分裂期时，这种物质被称为染色体了。实际上染色质和染色体是细胞在不同状

态下的同一种物质。

染色体一般存在于细胞核中，有些生物虽然还没有形成真正的细胞核，如细菌，但这种称为前核（原核）生物的细胞中也有染色体，不过原核细胞中的染色体比较简单，仅仅是一个头尾相连的脱氧核糖核苷酸（DNA）分子。而像我们人类等高等生物，在细胞中的染色体，不仅在数量上不止一个，而且在类型上也不是一种。如人的身体细胞中就有23种长短、形态不同的染色体，而且每种染色体各有两条，由此可见，人的每个细胞中有23对染色体，那一对对染色体中两个形态结构看起来完全相同的染色体，互为同源染色体。这里必须指出，只有在女性的细胞中，23对染色体才都是对对成双的，而男人的细胞中，成双又成对的染色体只有22对，还有两条虽然成双却不成对，因为其中一条染色体长，一条染色体短。即使是由男变成女的变性人，他们染色体的组成也绝对不会随着性别的改变而改变，用显微镜看他们的细胞，能清楚地看到一长一短这成双不成对的染色体。这两条染色体因能判别男与女，所以特称为性染色体，用X和Y代表，女性的性染色体用XX表示，男性则用X与Y表示。其余22对染色体男与女都一样，就称为常染色体。

随着研究技术的改进，染色体的结构和组成也被研究清楚了，所有染色体都离不开DNA和蛋白质这两大类物质，每条染色体都有一个DNA分子贯穿始终，因为人的细胞中有23对染色体，由此可知人的细胞中含有23对DNA分子。染色体的外形看起来也不复杂，大多数染



细胞中的染色体

染色体在显微镜下看起来像腰部特别细的法国女郎，染色体上的细腰部称为着丝粒，着丝粒的两端称为染色体臂，如果两端的臂长短不同，那么长的就是长臂，短的就是短臂。实际上像法国女郎那样一个腰连着两个手臂两条腿的染色体在细胞分裂过程中存在的时间很短，在细胞分裂过程中较长时间内看到的染色体是线条状的，即一个着丝粒的两端各有一个臂。

在孟德尔的分离定律和自由组合定律重新被科学界发现后，美国哥伦比亚大学的青年学生萨顿把染色体当作了基因，他的根据是一个生物体的身体细胞在形成生殖细胞时，细胞中的同源染色体的两个成员必然分开并进入不同的生殖细胞，这与孟德尔说的等位基因在形成生殖细胞时的分离非常对应，如果把染色体看成是基因，完全可以解释孟德尔所得到的研究结果。

然而，染色体就是基因的观点仅仅是一场闹剧，很快就因染色体的数目无论如何不能与生物体多种多样的性状相匹配而草草收场了。但萨顿的假说给那些善于捕



摩尔根

捉战机的优秀科学家点燃了智慧的火花。与萨顿在同一学校的摩尔根用果蝇作研究材料，经过红眼配白眼、灰体配黄体等杂交试验，终于证明了基因与染色体的关系。根据果蝇杂交试验的结果，摩尔根不仅从孟德尔定律的怀疑论者摇身一变成为孟德尔的虔诚信徒，而且找到了基因所在的位置——

染色体。与萨顿不同的是，摩尔根并没有把基因与染色体等同，他认为基因是染色体上的一个点。染色体相当于一条长长的线，基因就是线上的一个一个点。那些直线排列在染色体上的基因就称为连锁基因，存在许多基因的一条染色体就是一个连锁群。

DNA 与基因

既然每个染色体由蛋白质和 DNA 共同组成，那么在染色体上的基因究竟是蛋白质还是 DNA，或者是蛋白质和 DNA 的复合物呢？这个问题到上世纪 40 年代得到了答案。美国细菌学家埃弗里进行了肺炎球菌的有毒类型转变为无毒类型，及无毒类型转变为有毒类型的转化试验，根据确凿无疑的研究结果，肯定了只有 DNA 才有资

格充当基因的物质基础，而蛋白质没有资格组成基因。

DNA 作为基因物质基础的地位确定以后，许多科学家，包括生物学家、物理学家等对 DNA 的结构非常感兴趣，许多人把揭示 DNA 结构作为终身奋斗的目标。1953 年，沃森和克里克在总结别人的经验和教训的基础上，经过苦战，终于揭开了 DNA 结构的秘密。

1. DNA 的分子结构

每个 DNA 分子是由两条多核苷酸长链靠互补的碱基连接起来的双链大分子，这个双链大分子犹如向右旋的一架两边有扶手的梯子，扶手的骨架是由磷酸和脱氧核糖构成的，扶梯的踏脚板就是互补的碱基构成的碱基对。

DNA 组每条多核苷酸链中的每一个核苷酸都是由一个磷酸分子、一个脱氧核糖分子和一个碱基组成的复合物，根据核苷酸中的碱基差别可把核苷酸分成不同类别，如核苷酸中的碱基是嘌呤，那就叫嘌呤核苷酸，如果碱基是嘧啶，那就叫嘧啶核苷酸。由于嘌呤类碱基有腺嘌呤（A）与鸟嘌呤（G）的区别，因此嘌呤类核苷酸也有腺嘌呤核苷酸（仍用 A 代表）和鸟嘌呤核苷酸（仍用 G 代表）两类。同理嘧啶类碱基有胞嘧啶（C）和胸腺嘧啶（T）这两类，所以嘧啶类核苷酸也包括胞嘧啶核苷酸（仍用 C 代表）和胸腺嘧啶核苷酸（仍用 T 代表）两类。两条链中核苷酸与核苷酸的对接靠的是互补性的碱基，A 与 T、G 与 C 就是互补性的碱基配对。

现在常把 DNA 分子简化成由两条 ATGC 连成的长链。这里的 ATGC 实际上是代表了四种不同的核苷酸。