



解读基因

——来自DNA的信息

—THE LANGUAGE AND MEANINGS OF DNA

罗伯特·波拉克 / 著
Robert Pollack

杨玉龄 / 译
中国青年出版社

SIGNS OF LIFE

◎ 生命传奇丛书

大视野文库

解读基因

——来自 DNA 的信息

THE LANGUAGE AND MEANINGS OF DNA

罗伯特·波拉克 / 著
Robert Pollack

杨玉龄 / 译

中国青年出版社



大视野
文库

◎ 生命传奇丛书

(京)新登字083号

图书在版编目(CIP)数据

解读基因:来自DNA的信息/(美)波拉克/著;
杨玉龄/译.-北京:中国青年出版社,2000
(生命传奇丛书)
ISBN 7-5006-3986-4

I.解... II.①波... ②杨... III.脱氧核糖核酸
-普及读物 IV.Q523-49

中国版本图书馆CIP数据核字(2000)第47858号

北京市版权局著作权合同登记

图字:01-1999-3442

Copyright © 1994 by Robert Pollack

All rights reserved

For information about permission to reproduce selections from
this book, write to Permissions, Houghton Mifflin Company,
215 Park Avenue South, New York, New York 10003

Library of Congress Cataloging-in-Publication Data

Pollack, Robert, date.

Signs of life: the language and meanings
of DNA / Robert Pollack.
p. cm.

Includes bibliographical references and index.

ISBN 0-395-64498-4

*

《解读基因——来自DNA的信息》

罗伯特·波拉克/著

杨玉龄/译

责任编辑:徐泳

装帧设计:敬人设计工作室 吕敬人+杨林青

中国青年出版社 出版发行

社址:北京东四12条21号 邮政编码:100708

中国青年出版社印刷厂印刷 新华书店经销

*

889×1194 1/32 8.5印张 2插页 200千字

2000年10月北京第1版 2000年10月北京第1次印刷

印数10000册

定价:19.00元

作者简介

波拉克(Robert Pollack)

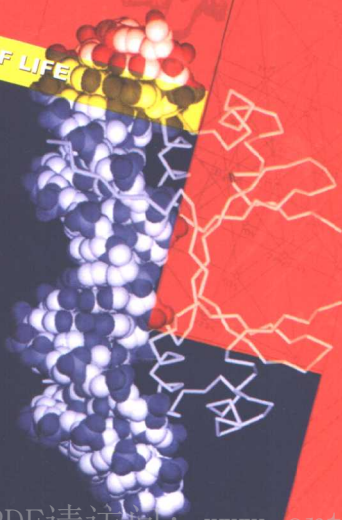
1938 年出生于美国东岸(祖父母皆来自东欧), 哥伦比亚学院物理系毕业, 布兰戴斯大学 (Brandeis University)细菌遗传学博士。

取得博士学位后, 波拉克先是在纽约大学医学院病理系研究癌症的分子机制。因为表现杰出, DNA双螺旋结构发现人之二的沃森博士, 力邀他加入冷泉港研究所, 继续从事病毒及癌症转化机制、分子细胞学的研究。

波拉克从事研究及教学工作数十年, 80年代曾经返回母校哥伦比亚学院, 担任院长。曾经获颁古根汉姆基金奖(Guggenheim fellow)。目前在纽约市及佛蒙特州, 均有教职。

责任编辑=徐 泳
装帧设计=敬人设计工作室
吕敬人+杨林肯

SIGNS OF LIFE



目 录

导 读	人类基因组解读计划——一个无可逃避的挑战	001
楔 子	DNA 时代已来临	010
第一章	看不见的城市，晶体状的书堆	027
第二章	染色体和圣典	061
第三章	说文、解字，转录、转译	099
第四章	分子的文字处理机	135
第五章	释义、造句，基因工程阶段	175
第六章	介于物理与历史之间：生物学的新典范	215
结 语	给分子生物学家的诤言	252
附 录	名词注解	259

导读

人类基因组解读计划 ——一个无可逃避的挑战

提到“基因”，大家也许都听说过，知道它是决定遗传特征的基本单位。也许有人更知道基因是由DNA构成，而在基因里蕴含的是，构建生物个体的遗传信息。就像电脑软件用“0”与“1”作为机器语言的密码，生物世界，小至细菌、病毒，大至人类，都是以A、T、G、C这四个字母的密码形式，将遗传信息储存在长链的DNA分子里。然后，细胞得想法子把A、T、G、C的遗传密码透过“转译机器”，制造出特定结构的蛋白质，来执行复杂的生物功能。到了细胞要开始进行分裂繁殖的时候，还得把这套密码忠实地复制一份，分别藏到二个子女细胞，以便未来可以继续执行任务。

决定DNA里遗传信息的密码，就这样一代一代地保存下来。同时带有这些遗传信息的细胞，也就在密码的指引下，忠实地显示出DNA内涵的表征了！

1953年，沃森和克里克两人发现DNA的双螺旋结构，不仅解开了基因的化学结构之谜，同时也揭示了基因所携带的遗传信息，如何精确地从上一代传递到下一代。过去40年里，分子生物学家已经成功地确定了遗传密码的“翻译原则”，另外，遗传信息“切割/连结技术”的成熟，让我们更可以“任意”改变不同生命形式的遗传内涵。步入20世纪80年代，对个别基因的认识，不论是结构、功能、以至于调控的机制，都已日趋完备。分子生物学家开始野心勃勃地把目光投射到另一个更大的目标：我们是否可以了解构成一个生物个



体“所有基因”的功能，和它们之间彼此相互作用的法则和最终的结果？于是“基因组”(genome)的想法逐渐浮出台面。

“基因组”是什么？简单地说，就是一个生物体所有基因的总合。从70年代开始，科学家就汲汲于追求，对某些简单生物“基因组”的完全定序。最早基因被完全定序的“生物”是一种叫作“ $\phi \times 174$ ”的噬菌体。这个噬菌体的基因全部只有5383个遗传密码，1977年由英国剑桥大学的圣格(Frederick Sanger, 1918—，1958年及1980年两届诺贝尔化学奖得主)教授首先完成了它的密码定序。随后陆续有一些更大的基因组，像天花病毒(186000个密码)、线粒体(187000个密码)和叶绿体(121000个密码)等等的基因被完全定序^①。

也许大家会注意到，上面所提到的，都是病毒或是细胞器。换言之，它们都是必须寄生到宿主细胞内，利用宿主细胞的资源及新陈代谢，才能进行复制繁衍。病毒并不能独立营生，因此知道了病

^①截止2000年10月初共有32种微生物的全基因组被完全定序，水稻基因组之完全定序接近完成，人类基因组全序列也将在2003年完成。

毒全部的基因序列后,最多是告诉我们病毒的组成,和它与宿主细胞间的互动关系。这些蓝图还不足以让我们知道,一个最基本、能独立营生的生命个体,需要多少不同基因的相互作用才能形成?

相反,细菌就不同了!每一个细菌都能独立呈现所有的生命现象——有组织的结构、新陈代谢、复制繁殖、侦测与适应环境等等。而细菌的基因组所提供的遗传信息,构成了运作上述所有生命现象的基础。1995年7月,位于美国马里兰州一家私人研究所首次决定了一个叫做流感嗜血杆菌的细菌全部基因的密码序列——这个细菌的基因组,共有180万个遗传密码。

从这180万个密码中,大约可以读出1734个决定蛋白质的区域。换句话说,这1734种不同的蛋白质,可以透过在不同时空、不同数量的交互作用,产生一个独立自主的生命个体,呈现一个单一细菌所能表达出所有的生命现象!在这1734种蛋白质当中,可以分辨出功能的大约有1000个,负责氨基酸与脂肪的代谢、细胞外套的合成、能量的产生、蛋白质的输送、DNA的复制与基因的转录等等。180万个密码(还不及“视窗九五”这套软件的四分之一)竟然可以谱出一曲生命的乐章,这毫无疑问使我们对生命的解读,向前迈进了一步。

“生命”就好像电脑音乐

但是从这180万个密码中,我们究竟能参透多少生命中的奥秘呢?坦白地说,我们像是个才上幼稚园的孩童,刚刚学会了26个英文字母,而面对需要解读的是《莎士比亚全集》。当我们问道:一个单一的细菌,如何展现它的生命活力时,我们不只是要知道细菌的一生一共动用了多少个种类不同的蛋白质,共同协力来维系生命;同时还要了解,在它的一生(大肠杆菌的一生只有20分钟)每一个不同的片刻,是哪些蛋白质登场在生命舞台的哪一个角落献艺。

换言之,生命是一个涵括了“时空”的整体,而遗传密码代表的只是一个“线性”的信息。从“线性”的信息转换成“四维空间”的

呈现,需要的是另一套“硬件”的运作。因此“生命”就好像电脑音乐,带有“乐谱”的软件,必须透过“硬件”的解读才能奏出一首气势磅礴的交响乐,而软件/硬件些许的差异,整合后都可以呈现出形似、但品味完全不同的风貌!

从单一细胞到多细胞生物,不仅基因组的复杂程度大幅提高,而掌握“生命现象”的困难度,更是呈几何级数增加。像我们身体大约是由一亿兆(10^{14})个细胞所组成,除了极少数像成熟的红血球外,每个细胞都有一个细胞核。细胞核里储存了构建人体所需要的遗传信息。这些遗传信息也就是人类的“基因组”,包含了大约30亿个遗传密码。这些密码大约可以组成10万个基因。每个基因(大约含了3万个密码)决定一种蛋白质的结构与功能;也就是说,这个基因,决定了我们身体的基本构造与特征。

我们身体每一个体细胞里都有两套共计46条染色体;其中一套染色体来自爸爸,另一套来自妈妈。而所有的遗传信息就是储存在这些染色体里的DNA分子上,生殖器官里某些体细胞会经过一个减数分裂的过程,产生只带有单套染色体的精子或卵子。精子与卵子结合以后,就再度形成一个带有双套染色体的受精卵,而开始另一个胚胎发育的周期。

一个受精卵呈现出的生命活力是一个奇迹,它不仅具有无穷生长分裂的潜力——从单一细胞发育成一个由一亿多兆个细胞组成的个体;在细胞生长分裂的过程中,还有另一套精致的分化程序也同时在展开——从开始看来完全相同的胚胎细胞,逐渐分化成外观、功能完全不同的(像神经、肌肉、肝、肾等)细胞,同时组成特定的器官。是谁在发号施令,指挥这一系列生长、分化程序的进行?我们相信,这些指令加上制造10万种蛋白质的蓝图,全都是存在人类的基因组里。

因此要了解人类基因组,首先,就必须决定这30亿个密码的排列顺序。然后再决定:哪些密码是用来决定蛋白质结构;哪些密码是用来作为调节指令,让某些特定的蛋白质,只在个体发育特定的期

间，于特定的细胞内制造出来。

生物学家的“登月计划”

最早提出“人类基因组解读计划”(Human Genome Project)的是美国能源部。能源部对日本广岛与长崎的居民有一个长期追踪的研究计划，想要了解这些曾暴露于原子弹辐射的人，是否会产生一些基因突变而遗传给他们的后代。但是许多基因突变从外表并不容易看得出来，有些可能要经历好几个世代才会显现其特征。1984年12月，能源部在犹他州的盐湖城召开了一个研讨会，会中大家都同意，最快、最“简单”的解决之道，就是直接比较这些人与其子女全部的遗传密码顺序，其间的差异可能就代表是由辐射暴露所带来的后果。

当然，要作这样的比较，前题就是要有能力，很快地把人类30亿个密码的顺序排列出来。如果能把人类30亿个密码的顺序排出来，我们不就马上知道构成人体的10万个基因是什么了吗？就像潘朵拉的盒子，一旦发现大规模基因定序的技术可行后，我们就有可能完全了解构建人体蓝图的每一个细节。这是多么大的一个诱惑啊！“人类基因组解读计划”马上就成为每一个分子生物学家怦然心动的目标。

1986年5月，有关“人类基因组解读计划”的讨论正式公开露面。在冷泉港研究所每年举办一次的研讨会上，大家第一次听到这样的计划。科学家立即的反应当然是分歧的。遗传密码的定序是件呆板、枯燥、但又十分花钱的工作。大家最担心的就是对有限资源是否会产生排挤效应，另外也有人担心让年轻科学家参与这样的工作，会抹杀他们的创造力。总而言之，多数人的心态是乐观其成，大家都对这个计划产出的结果有兴趣，但谁来做这件事？当然不是我！

但这好像是生物学家的“登月计划”，目标是如此鲜明，而且容易打动人心，于是经过三年反复讨论，美国国家卫生研究院终于在1989年，正式成立了“人类基因组研究中心”，同时请冷泉港研究所的所长，也就是发现DNA双螺旋结构的沃森教授，担任第一任中心

主任。而计划预期完成的日程也确定为2005年。换言之，到了2005年，我们每个人手上都可以有一本含有30亿字母排列的天书。下一个问题就是我们该怎么去读懂它！

未来生命科学研究的主流

人类基因组解读计划展开以后，随着技术的熟练与进步，慢慢地大家发现，其实DNA定序的工作并不是那么枯燥无聊；相反地，当信息累积的速度愈来愈快，面对巨大的遗传密码资料库，新的分析方法不断推出，过去从未想像过的信息，便幽然浮现。“生物信息学”这个五年前还没有听说过的学问，现在已成了就业市场上的抢手货，药厂与生物技术公司突然好像掘到金矿，像淘金客般地涌入基因组的研究中。成百上千个新的基因从不同的生物、不同的组织、不同病态的细胞中涌现出来，每一个新基因功能的鉴定，都牵引出无限商机的梦想。基因组研究毫无疑问地，将成为未来第二波工业革命——生物技术的基石。

人类基因组解读计划所建立的各种分析技术，使我们不单单只满足于知道人类的10万个基因是什么；我们还想知道，每一个基因在身体发育过程中扮演什么样的功能；最后剩下来的问题是，这些基因究竟从何而来。换句话说，人类基因组解读计划的对象从人延伸到酵母菌、线虫、果蝇、斑马鱼和老鼠。比较不同生物基因组的信息，使我们对基因演化的历程逐渐得到一个概括的印象。

许多基因都有一些结构类似的区域。我们可以想像，在简单的原始生物中，基因的种类可能不多，每一个基因就好像一个简单的积木，凑合着拼出一个图像。随着演化的进程，简单积木间的排列组合可以逐渐形成更复杂的积木，复杂积木的组合产生更复杂的图像。从由简而难的图像演化中，我们依稀可以追溯构筑积木的变迁。

人类基因组的研究从“结构基因学”，到现在风起云涌的“功能基因学”，以至未来的“演化基因学”，都让我们对“生命”产生不同

方面的认识，这是20世纪生物科学上最重要的一个里程碑。而基因组的研究不仅带动了当代生物医学研究的风潮，同时对下个世纪的科学教育也会产生重大的影响。

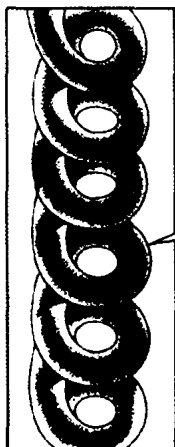
前面提过的信息处理、复杂系统、多重非线性的交互作用等等，都会成为未来生命科学研究的主流，缺少这方面的认识，我们就无法培养出能够适应未来这种研究趋势的科学家。

周成功

染色体与DNA双螺旋构造、遗传信息

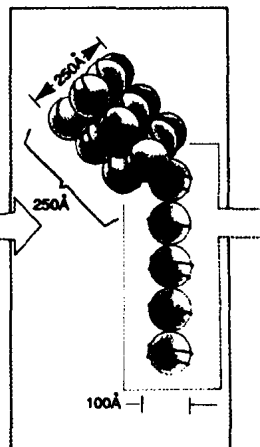


1. 人类的染色体
(人类有 23 对染色体)

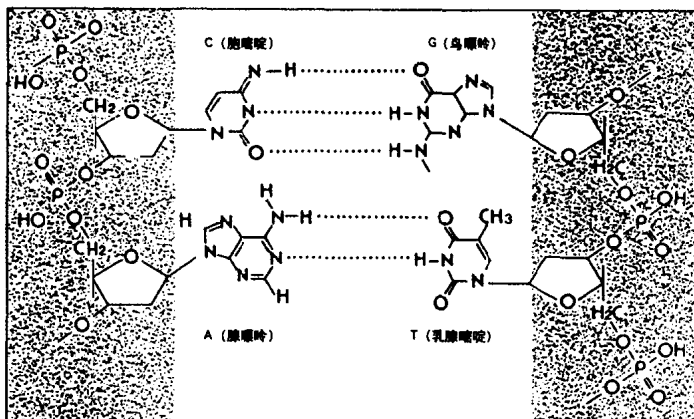


2. 染色体细部
(染色体由长丝状的染色
线纤维层层缠绕而成)

(1Å 等于 10^{-8} 厘米)



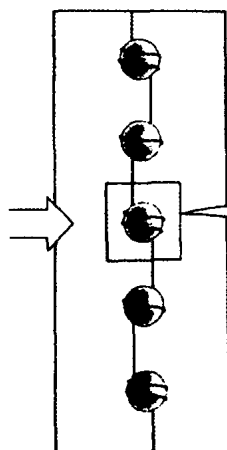
3. 染色线纤维
(染色线纤维由 DNA 盘卷在
组蛋白上, 再折叠而成)



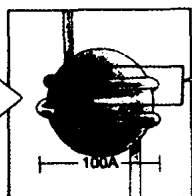
9. 遗传信息

(A 一定与 T 配对, C 一定与 G 配对。A、T、G、C 四种碱基对的排列组合, 决定了遗传信息)

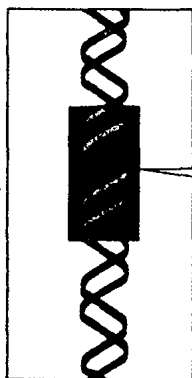
◎绘图: 陈立君



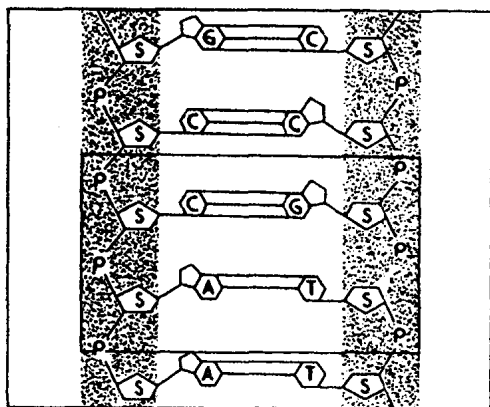
4. 染色线纤维再放大
(图中的球状物即为组蛋白, 细线为 DNA)



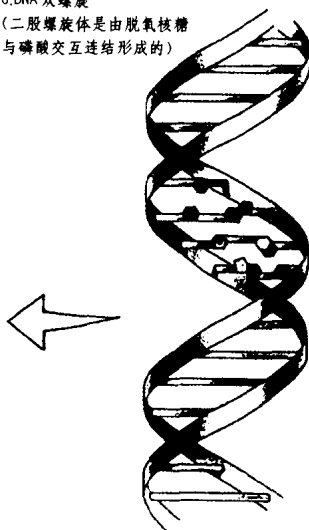
5. 染色线
(每一染色线约含 140 个碱基对)



6. DNA 双螺旋
(二股螺旋体是由脱氧核糖与磷酸交互连接形成的)



8. 碱基对
(灰点区域代表螺旋体, 其中的 P 为磷酸, S 为脱氧核糖。中央白底区的 G 与 C、A 与 T 两两的配对, 即为碱基对)



7. 双螺旋再放大
(碱基对如竹梯般, 将双股螺旋体栓架住)

楔子

DNA 时代已来临

我敢说，
要是我们活得够久，
久到有朝一日能够回顾的话，
20 世纪后半叶将会被冠上
“DNA 时代”而非“原子时代”的封号。

这本书的写作始于 20 世纪 70 年代晚期，当时我和家人首次得知父亲罹患了阿耳茨海默氏症(Alzheimer's disease)。经过一段漫长岁月后，父亲终于过世了；病中这段期间，他甚至没办法认出我们。每当我们一想到他那神秘、缓慢又悲哀的结局，就不禁更加难受。

在他刚被诊断出罹患这种疾病时，我的反应就像任何头脑清醒的分子生物学家一样。我花了许多时间埋首在哥伦比亚大学医学院图书馆里，苦读各种科学及医学文献，希望能找出对付这种病症的线索。那时我并没有意会到这件事有多么令我苦恼，我只是一味认定：只要让我找着那必然藏匿在研究文献堆里的固定模式，我就有办法帮助父亲。

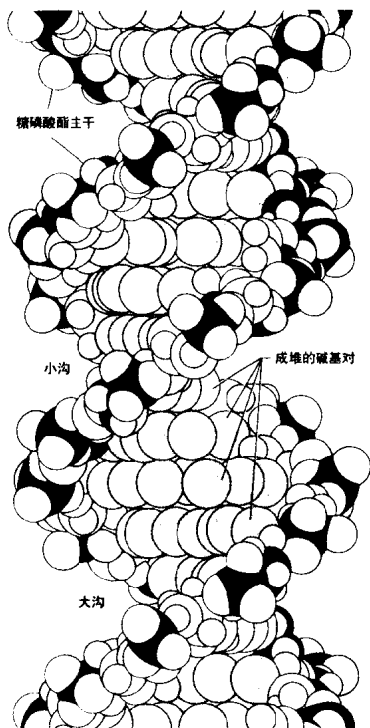
但我错了。十年前，完全没有任何可行或可预见的方法，能扭转阿耳茨海默氏症的病程；即使到了今天，虽说迷雾中似乎出现了些许诱人的闪光，我还是看不到任何妙方。当时我很难接受这一点，而且到现在还是一样，但是，不论我用什么样的方式活下去，我必须相信、接受这种事实。

自从父亲的检验报告出来后，我的生活大致和从前一样，只除了一点明显的差别：我仍然相信，我们能通过科学来了解大自然的运作方式；然而当我明白，科学发现并不能免除我丧失亲人的痛苦后，我也开始对科学的极限产生兴趣。

编织科学家之梦

我从小就爱看书，上了高中之后就想成为科学家。但在那之前，

◎ 双螺旋模型



我什么书都看，而且一心想 当个作家。

20 世纪初，我的祖父母自东欧来到美国。当年他们登上爱丽丝岛时，身无长物，有的只是希望：如果自己不行，希望子女将来能在美国过好日子。我父母和许多其他移民者的子女一样，没能实现双亲的梦想，便早早离开学校去工作。但他们却紧抓着一种 19 世纪对科学的看法，认为科学是进步的动力，是社会阶级里的平等主义者。他们拜读过威尔斯 (H.G.Wells, 1866—1946, 英国科幻小说家)、德克鲁夫 (Paul de Kruif)、霍登 (J.B.S.Haldane, 1892—1964, 英国遗传学家、生化学家) 这些作家的作品，而且也传承给我。这令我一度相信，科学定律可以控制整个历史及所有作为，而且这些定律就和那些支配能量和物质的定理一样，都是可以预期、具有普遍性的。

然而，在我青春叛逆期间有一个奇特的转折。当同伴们故意打破父母的饮食戒律，跑到中国餐馆去大啖猪肉时，我则以“坚称科学定律没办法延伸到人类事务，而且历史也是不可预知的”，来表现我的独立精神。在这样的叛逆行为里，我开始想