

21世纪科技新视野丛书

New Horizons in The 21st Century's Science & Technology

(英汉对照读物)

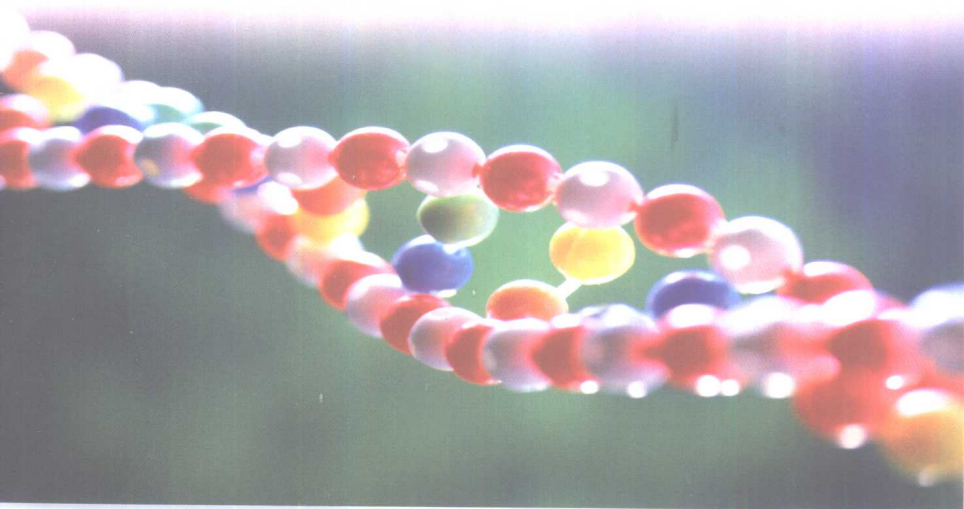
◆丛书主编 吴文智 徐 新

THE FUTURE Biotechnology

未来生物技术

◆ 刘世红 倪正琴 编译

煤炭工业出版社



New Horizons in The 21st Century's Science & Technology

21 世纪科技新视野丛书(英汉对照读物)

未来生物技术

The Future Biotechnology

刘世红 倪正琴 编译

煤炭工业出版社

图书在版编目 (CIP) 数据

未来生物技术/刘世红,倪正琴编译. —北京:煤炭工业出版社, 2002

(21 世纪科技新视野丛书/吴文智, 徐新主编)

ISBN 7-5020-2134-5

I. 未… II. ①刘…②倪… III. 英语—对照读物, 生物技术—英、汉 IV. H319.4 : Q81

中国版本图书馆 CIP 数据核字 (2002) 第 001050 号

未来生物技术

丛书主编 吴文智 徐新

编译 刘世红 倪正琴

策划 向云霞

责任编辑: 向云霞 姚有超 杨成俊

*

煤炭工业出版社 出版

(北京市朝阳区芍药居 35 号 100029)

北京密云春雷印刷厂 印刷

新华书店北京发行所 发行

*

开本 787×1092mm¹/₃₂ 印张 4¹/₂

字数 95 千字

2002 年 2 月第 1 版 2002 年 2 月第 1 次印刷

社内编号 4905 定价 7.50 元

版权所有 违者必究

本书如有缺页、倒页、脱页等质量问题, 本社负责调换

主 编：吴文智 徐 新
副 主 编：金光辉 王立非 张 权
编 委：王 坤 陈东东 李 瑛
 梅智宇 李恩宁 王慧娟
电子编辑：靳红华 何 磊

序

人类社会进入 21 世纪的今天,科学技术日新月异的发展速度真正地到了匪夷所思的程度。那些在过去常常被人们认为不可能的梦想,今天大多成了事实。如果将来有一天你突然发现汽车可以像飞机一样在大街小巷穿梭飞行,或当你在某个餐厅就餐时竟然发现你对面就坐着一个与你百分之百相象的你,请不要吃惊,因为这正是现代科学技术创造的结果。

科学探索是一项伟大的冒险活动,充满了刺激与振奋。它使人类的求知欲和好奇心得到满足,并且益发地激起人们愈来愈大的想像力,去欣赏和理解科学技术所带来的种种美妙与神奇。e 时代的到来更使人们对知识的力量不再有丝毫的怀疑,唯有对科学知识的需求更多地增添了紧迫感。“让科学知识为我们插上腾飞的双翅”成了我们绝大多数人潜意识的追求,正是在这样一种背景下,我们构想了这套《21 世纪科技新视野》丛书。意欲从浩瀚的科学海洋中撷取那些对我们明天的开拓进取富有启迪意义的新知识,奉献给一切热爱学习,热爱科学的人们。

《21 世纪科技新视野》是一套以英汉对照方式编排的“语言学习+科技知识”的“链接”式丛书。在编写过程中,所有参编者遵照“应用价值、文化价值、精神价值”相结合的原则精心选择每篇文章,努力把最能体现人类创造力与想像力的科学成果介绍给广大读者,所有原文均摘自英语国家的现版期刊或网络杂志。英文地道,原汁原味。内容讲求知识性、趣味性、通俗性、新颖性,

使得广大英语爱好者在学习英语的同时可以接受新科学知识的熏陶，也使那些钟爱新科学知识的人们在掌握新知识的同时得以强化和提高自己的英语水准，特别是与这个时代特点相融合的那些“与时俱进”的科技英语水准。这在加入 WTO 后的今天尤为重要，因为 WTO 已不容置疑地把每一个中国人深深地卷入到了全球一体化发展的新浪潮中。作为链接未来科学技术的知识纽带——《21 世纪科技新视野》丛书，将把我们与新科学和新知识紧紧地联接在一起，从而为广大读者打造出一个再次提升自己的知识平台，以便可以从容应对 WTO 时代扑面而来的任何挑战。

如果本丛书的出版发行确能使读者对我们的上述编写意图认同十之一二，那就是对我们所有编写人员的莫大奖赏。此外，本书得以顺利出版，除了我们所有编写人员的努力外，还折射了煤炭工业出版社决策者的创新意识和与时俱进的奋发精神，渗透了本丛书责任编辑的辛勤汗水。在此一并表示感谢。

对于书中可能存在的不足之处，我们将在下次再版时改进，敬请广大读者批评指正。

《21 世纪科技新视野》丛书编委会

2002 年元旦于南京

CONTENTS

目 录

The Meaning of Life	2
生命的真谛	
Code Breakers	22
基因密码破译者	
Potential Benefits of Human Genome Project Research	42
人类基因组研究的潜在价值	
A Clone in Sheep's Clothing	56
克隆羊	
Pass the Genes, Please	64
请赐基因	
Ready-to-Wear Flesh	82
人造皮肤	
Placental Puzzle	92
胎盘之谜	
Know Your Enemy	110
认识你的敌人	
New Gene-therapy Techniques Show Potential	128
新基因治疗技术显露潜力	



Rapid progress in genome science and a glimpse into its potential applications have spurred observers to predict that biology will be the foremost science of the 21st century.

基因组科学的迅猛发展及其潜在的应用价值已经促使人们预见到：21世纪，生物学将是最前沿的科学。



The Meaning of Life

By Tina Hesman

In a football field-size room that genetics researchers have dubbed “the factory”, row upon row of sequencing machines churn through strands of *DNA¹ and record long strings of As, Ts, Gs, and Cs. Each symbol represents the stuff of life—the chemical *bases² adenine, thymine, guanine, and cytosine that make up the genetic code of every living organism on Earth.

The 300 sequencing machines at the factory in *Celera Genomics³ headquarters in Rockville, Md., run day and night in the race to decipher the genetic blueprints of dozens of organisms including people, mice, flies, and flowers. Celera isn’t alone in its efforts. Other scientists from around the world, many affiliated with a competing, massive public effort to map genomes, dump more than 100 million bases each week into a public data repository.

Raw genetic sequences, however, tell little. It’s the messages among all those letters that the scientists are after. Only by finding patterns in the long strings of DNA will scientists understand “how the genome is wired” and ultimately how life is structured, says mathematician Pavel A. Pevzner of the University of Southern California in Los Angeles.

The task of making sense of the raw information is formidable.



生命的真谛

在一个足球场般大小,被基因学家们称为“工厂”的房子里,一排排测序仪正对 DNA(脱氧核糖核酸)链进行搅动并记录下长串的 A、T、G 和 C 符号。每一个符号都代表着生命的一个组成要素——A 代表腺嘌呤, T 是胸腺嘧啶, G 是鸟嘌呤, 而 C 则代表胞嘧啶, 所有这些符号组成了地球上一切生物的基因密码。

马里兰州罗克维尔市塞勒尔基因组研究公司总部工厂里的 300 台测序仪正夜以继日地工作, 力争尽快揭开包括人类、老鼠、苍蝇和花朵在内的几十种生物的基因组图谱奥秘。从事这项工作的并非只有塞勒尔公司一家单位。世界各地的科学家们每周都要向一个公众数据库补充一亿多个碱基, 其中许多人参与了绘制基因组图谱这项极富竞争性又需要广泛合作的工作。

然而, 人们从原始的基因序列中收益甚微。科学家们感兴趣的是那些字母中所蕴涵的信息。来自洛杉矶南加州大学的数学家帕威尔·佩夫斯纳说, 只有在了解 DNA 分子链的结构之后, 科学家们才能弄清楚“基因组是怎样构成的”这个问题, 并最终揭开生命之源的奥秘。

搞清原始信息的意义是很艰巨的任务。去年 2 月, 在弗吉尼亚

-
1. DNA: 脱氧核糖核酸, 一种在细胞中带有基因信息的核酸, 能够自行复制并合成核糖核酸, 由两个核甙酸长链组成, 这两个核甙酸链交结成一个双螺旋体
 2. base: 碱基嘌呤(腺嘌呤和鸟嘌呤)或嘧啶(胸嘧啶、胸腺嘧啶和尿嘧啶)的一种, 附着在脱氧核糖核酸或核糖核酸的糖分子上
 3. Celera Genomics: 美国 PE 生物系统公司做人类基因组测序的子公司



“The speed of acquiring data is now exceeding our ability to comprehend it and put it into the proper biological context,” said Michigan State University biologist George M. Garrity at a conference on microbial genomes in *Chantilly⁴, Va., last February.

Gone are the days when biologists could analyze most of their data with a pencil and a sheet of paper, says Steven L. Salzberg of the Institute for Genomic Research, also in Rockville. Today’s biologists need computing power to find even the most obvious needles in molecular haystacks of information, he says.

That’s where the field of *bioinformatics⁵ comes in, says Sean Eddy, a computational biologist at the Washington University School of Medicine in St. Louis. The burgeoning field, also called biological computing, straddles the lines dividing biology, computer science, and mathematics.

The process of making sense out of a DNA sequence by finding genes and other interesting patterns in the strings of letters is called annotation, and it’s often the most difficult aspect of a sequencing project, says computer scientist Peter D. Karp of *SRI⁶ Internationale in Menlo Park, Calif.

“Genome annotation is a lot like passing a piano through a nine-inch hole,” he told attendees of the Chantilly Conference. It’s very difficult, and it isn’t immediately obvious how such a task can be accomplished. It’s also essential to understanding biology, he says.

Processing all the data is going to take a lot of time and resources, says Celera’s president, J. Craig Venter. His company has already identified all the bases of one person’s DNA sequence and is planning to



的尚蒂伊举行了一次微生物基因组大会,密歇根州立大学的生物学家乔治·加里第在会上说,“对于数据的需求速度超过了我们的领悟能力,我们不知道该把这些信息放到什么样的生物演化关系中去。”

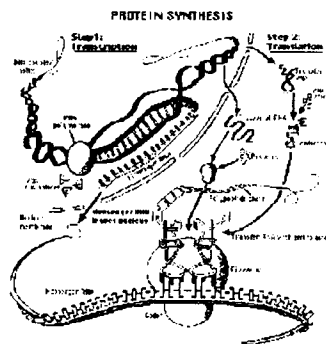
同样是位于罗克维尔的基因组研究学会的史蒂文·萨尔茨伯格称,生物学家们仅靠一枝笔、一张纸就可以对大部分数据进行分析的时代已一去不复返了。他说,对于今天的生物学家们来说,哪怕是分子海洋中最明显的信息,也需要依靠计算机的力量来获得。

正因为如此,才有了生物信息学的兴起,来自圣路易斯华盛顿大学医学院的计算机生物学家肖恩·埃迪如是说。这个新兴的领域,也被称为生物计算机学,跨越了生物、计算机和数学这三门学科。

彼得·D·卡普,这位来自位于加利福尼亚州门洛帕克市斯坦福研究所的计算机科学家说,通过发现基因和字母链中的其它有趣的组合来弄清 DNA 序列的含义的过程被称为注解。它往往是整个序列研究中最难的一个环节。

“基因组注解工作就好像让一架钢琴穿过一个直径仅九英寸的洞口。”他这样告诉出席尚蒂伊大会的人。这项工作很难做,无法立刻知道如何着手。他说,要想做好基因组注解工作,还必须对生物学有深刻的了解。

塞勒尔公司总裁 J·克雷格·文特称,处理全部的数据需要大量的时间和资源。他的公司已经确认了一个人的 DNA 序列中全部



4. Chantilly: 尚蒂伊,弗吉尼亚州一城市

5. bioinformatics: 生物信息学

6. SRI: 斯坦福研究所



decode the sequences of four or five more people. Celera's announcement on April 6 came a week after the Human Genome Project, a publicly funded consortium of researchers, reported that it had finished determining 2 billion of the 3 billion bases of the human genome.

Venter predicts, however, that it will take most of this century to analyze the data.

"It's only through having phenomenal computers and computer tools that we will be able to try and understand how biology works," Venter says. "It doesn't matter what analysis [of the human genome] we do this year, it will be only the most cursory analysis." Scientists will have to invent ever-more-powerful computer algorithms to deal with and understand the data, he says. "Scientists will be making major discoveries from the human genetic code a hundred years from now," he says.

Scientists generally start by searching for obvious patterns in the DNA with a computer program. The trouble is, Pevzner says, that scientists don't know how a cell processes all the information contained in its DNA. But one thing is certain: "The way we do annotation today is very different from the way nature does it," Pevzner says.

Ultimately, researchers want to use computers to take raw DNA-sequence information and construct an entire biochemical model of an organism, says Karp. That's still a long way off, but some patterns are beginning to take shape on computer screens.

Without annotation, the billions of bases of DNA sequence are essentially useless, says bioinformatician Sylvia J. Spengler of *Lawrence Berkeley (Calf.) National Laboratory⁷. All those As, Cs,



碱基的类别,并且计划再破解四、五个人的基因序列。塞勒尔公司在4月6号发表了这项声明。一个星期之前,由公众资助成立的一家科学家联营公司人类基因组工程报告称,它已经确定了人类基因组全部30亿个碱基中的20亿个。

不过,文特预测说,这个世纪的大部分时间将花在对这些数据的分析上。

“我们只有在先进计算机及其它计算机辅助工具的帮助下才能尝试搞清生物的机能。”文特说,“我们今年[对人类基因组]进行怎样的分析并不重要,那将只是最粗略的分析。”科学家们必须设计出更强有力的电脑运算程序来处理并理解这些数据。他说,“在今后的一百年里,科学家们将通过人类基因密码获得许多重大发现。”

科学家们一般首先通过某个计算机程序来寻找DNA中比较明显的模式。佩夫斯纳说,头疼的是科学家们不知道细胞是怎样处理它的DNA中所包含的全部信息的。但有一点可以肯定:“我们现在采用的注解方式和生物本身的自然方式是很不相同的。”佩夫斯纳说道。

卡普说,研究人员的最终目的是想利用计算机获取原始DNA序列信息并构建一个完整的生物生化样本。这条路还很长很长,但有些结构图已经在电脑屏幕上逐渐成型了。

来自加州劳伦斯伯克利国家实验室的生物信息学家西尔维亚·J·施本格勒说,如果没有注解过程,DNA序列中几十亿个碱基就



Gs. and Ts might as well be alphabetized, she quips. “If we can’t make sense of it, we don’t have any information,” she says. “All we have is data.”

Right now, biologists are most interested in finding the genes. “That’s where all the action is,” says Salzberg.

Most genes lay the plan for strings of *amino acids⁸, which make up proteins. Some genes, however, encode various forms of *RNA⁹ that interact with proteins and other molecules to run the machinery of cells. And long segments of DNA between genes—and even within them—appear to code for nothing. These strings of bases, which are nevertheless being sequenced in the factory and other genome laboratories, are called junk DNA.

Genes that code for proteins have some easily recognized patterns. A string of three-letter words, called codons, spells out the code for the 20 amino acids used to build proteins. For example, GCC spells alanine in the cell’s language, while ACC spells threonine. Each protein gene also has a starting codon—the letters ATG—and one of three different three-letter stop signs—TGA, TAG, or TAA.

Even though protein-coding genes obligingly follow these rules, it’s not easy to recognize a gene, says Salzberg. The genes are big, some ATG sequences don’t indicate the beginning of a gene, and it’s difficult to decipher exactly how to group the letters to form the codons, he says. For instance, the letters CCCCCAAGAC could be read as GCC (alanine) CGA (arginine) AGA (arginine) C, but the pattern might also read GCCC (proline) GAA (glutamic acid) GAC (aspartic acid).

A complex statistical analysis can tell scientists the likelihood that



毫无用处。最好还得将所有那些 A、C、G 和 T 排一下顺序，她打趣地说：“如果对这些弄不清楚，我们就什么信息也没有了，”她说，“我们所拥有的只是数据。”

目前，生物学家们最感兴趣的是寻找基因。“那是一切活动的根源。”萨尔茨伯格说道。

大部分基因确定合成蛋白质的氨基酸链的序列。不过有些基因将各种形态的 RNA(核糖核酸) 进行编码，让它们和蛋白质及其它分子发生反应以促进细胞的运作。而基因之间(甚至基因内部)长长的 DNA 基因片段似乎没有提供任何遗传信息。这些碱基链被称为废旧 DNA，尽管这家“工厂”和其它基因组实验室也把它们进行了排序。

确定蛋白质遗传密码的基因具有某些易于辨认的结构特征。一些被称为密码子的三字母单词详细说明了用来合成蛋白质的 20 种氨基酸的遗传密码信息。例如，GCC 这种细胞语言说明了丙氨酸的信息，而 ACC 说明了苏氨酸的信息。每个蛋白质基因还有一个起始密码子 ATG，以及三个不同的三字母终止密码子中的一个：TGA、TAG 或 TAA。

萨尔茨伯格说，即使确定蛋白质遗传密码的基因都遵循这些规则，界定一个基因也非易事。他认为，基因信息很长，有些 ATG 序列没有指明基因的起始信息，很难确定如何组合字母以构成密码子才算恰当。例如，字母 GCCCGAAGAC 可以读作 GCC(丙氨酸)、CGA(精氨酸)、AGA(精氨酸) C，但这个字母组合也可以读作 GCCC(脯氨酸)、GAA(谷氨酸)、GAC(天冬氨酸)。

通过一项复杂的统计分析，科学家们能推测一个碱基与它之前

8. amino acid: 氨基酸，同时含有氨基(NH₂)和羧酸(COOH)的有机化合物，特别是含分子式 NH₂CH(R)COOH 的 20 种化合物，并且由肽键联结起来形成蛋白质

9. RNA: 核糖核酸，是所有活细胞和许多病毒的一种聚合组成要素，由一条长的、通常是单线的相互交替的磷酸盐和核糖单位组成，核糖上又附着有腺嘌呤、鸟嘌呤、胞嘧啶及尿嘧啶等基本成分。核糖核酸的结构及基本顺序是蛋白质的合成及遗传信息的决定因素



a base fits with the bases that come before or after it to form a codon. That's something people can't do very quickly and efficiently.

But computers can. Bioinformaticians have developed mathematical and statistical formulas, or algorithms, for sorting through large chunks of raw data to locate genes. Most gene-finding programs use a statistical method to test sequences by determining their “coding potential,” the likelihood that a string of bases codes for a protein.

For bacterial genes, the process is relatively straightforward because each gene is a continuous unit. In plants, animals, and some other organisms, however, the genes are often interrupted by chunks of junk DNA called *introns¹⁰.

Cells make RNA copies of genes, then slice out the introns and splice the protein-coding stretches—called *exons¹¹—back into a single molecule that's the template for making a protein. Although cells identify protein-coding regions and junk DNA with aplomb, computer programs can have difficulty searching over long stretches of junk—sometimes several thousand bases—to find the next exon, says Salzberg.

Luckily, the boundaries between introns and exons are marked. The borders aren't as pronounced as the start and stop codons, Salzberg says, but there is a pattern to them that cells and computer programs can pick up.

Gene-finding computer programs mark the stretches of DNA that are likely to contain a gene. Some programs perform the task better than others do, and programmers train their algorithms to recognize subtle differences in the way genes are flagged in different organisms. A