

*Genome*  
The Autobiography of  
A Species in 23 Chapters



# 基因组

人类自传

*Genome*  
The Autobiography of  
A Species in 23 Chapters



# 基因组

## 人类自传

[英] 马特·里德利 著 李南哲 译  
(Matt Ridley)



机械工业出版社  
China Machine Press

## 图书在版编目 (CIP) 数据

基因组：人类自传 / (英) 里德利 (Ridley, M.) 著；李南哲译. —北京：机械工业出版社，2015.6

书名原文：Genome: The Autobiography of a Species in 23 Chapters

ISBN 978-7-111-50424-5

I. 基… II. ①里… ②李… III. 人类基因—基因组—研究 IV. Q987

中国版本图书馆 CIP 数据核字 (2015) 第 121164 号

本书版权登记号：图字：01-2013-6892

Matt Ridley. Genome: The Autobiography of a Species in 23 Chapters.

Copyright © 1999 by Matt Ridley.

Simplified Chinese Translation Copyright © 2015 by China Machine Press.

Simplified Chinese translation rights arranged with Matt Ridley through Andrew Nurnberg Associates International Ltd. This edition is authorized for sale in the People's Republic of China only, excluding Hong Kong, Macao SAR and Taiwan.

No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system, without permission, in writing, from the publisher.

All rights reserved.

本书中文简体字版由 Matt Ridley 通过 Andrew Nurnberg Associates International Ltd. 授权机械工业出版社在中华人民共和国境内 (不包括中国香港、澳门特别行政区及中国台湾地区) 独家出版发行。未经出版者书面许可，不得以任何方式抄袭、复制或节录本书中的任何部分。

## 基因组：人类自传

出版发行：机械工业出版社 (北京市西城区百万庄大街 22 号 邮政编码：100037)

责任编辑：赵艳君 方琳

责任校对：殷虹

印刷：北京盛兰兄弟印刷装订有限公司

版次：2015 年 7 月第 1 版第 1 次印刷

开本：170mm × 242mm 1/16

印张：21

书号：ISBN 978-7-111-50424-5

定价：79.00 元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换

客服热线：(010) 68995261 88361066

投稿热线：(010) 88379007

购书热线：(010) 68326294 88379649 68995259

读者信箱：hzjg@hzbook.com

版权所有·侵权必究

封底无防伪标均为盗版

本书法律顾问：北京大成律师事务所 韩光 / 邹晓东



人类基因组是一套完整的人类基因，位于 23 对独立的染色体里。将其中的 22 对依照尺寸大小进行排序，并分别命名为 1 ~ 22 号染色体，1 号最大，22 号最小。剩下的那对是性染色体，在女性体内是两条大的 X 染色体，在男性体内则是一条大的 X 染色体和一条小的 Y 染色体。就尺寸大小而言，X 染色体处于第 7 号和第 8 号染色体之间，Y 染色体则是最小的。

就基因数量而言，“23”这个数目没有任何特殊的意义。许多物种拥有更多的染色体，与人类关系最近的近亲猿类就是这样的，也有很多物种的染色体要少一些。而相似的基因或有着类似功能的基因也不一定聚集在同一条染色体上。几年前我通过笔记本电脑和演化生物学家戴维·黑格 (David Haig) 聊天时，他说最喜欢第 19 号染色体。我很惊讶。他解释道，那条染色体上有各种调皮捣蛋的基因。在那之前我从来没有想过染色体竟然还有不同的性格，归根结底，它们不过是基因的组合。但是黑格的话却在我心中植入了一个想

法，一直以来都挥之不去：人类第一次探索人类基因组，发现了各种细节，并将其一一展现出来。为什么不试着从每条染色体上都选择一个最具代表性的基因，讲述一下基因组的故事呢？普里莫·莱维 (Primo Levi) 在他的自传体短篇小说集里就用类似的方法讲述了化学元素周期表，每个化学元素是一个章节，分别讲述了他接触到该元素的那段时期里发生的事情。

于是，我想到：人类基因组本身就是一部人类的自传，它从生命诞生之时起，使用“基因语言”记录了人类和人类祖先所经历的事更迭与变迁。有些基因从最早的单细胞生物诞生起，就没有发生变化；有些基因是当我们的祖先还是线虫的时候出现的；有些基因是当我们的祖先还是鱼的时候出现的；有些基因因为近期的一场流行病，变成了目前的形态；还有些基因可以用来研究过去几千年里人类迁移的历史。从 40 亿年以前到最近几个世纪，人类基因组谱写了人类的自传，记录了人类历史中的每个重要时刻。

我写下这 23 对染色体的名单，并在每条旁边列出对应的人类本性中的一些重要内容。这是一个缓慢而艰难的过程，我开始寻找那些具有代表性的基因，有时无法找到合适的基因，有时找到了合适的基因却发现它位于其他染色体上，真是令人沮丧。如何排列 X 和 Y 染色体是我遇到的另一个难题，在这本书中，我根据尺寸大小把它们放在了第 7 对染色体之后——对于 X 染色体而言，这再合适不过了。

乍一看来，本书似乎颇具误导性，似乎我在暗示 1 号染色体最早出现，其实不然。抑或是我在暗示 11 号染色体只与人类的性格有关，亦非如此。人类基因组里有 6 万~8 万个基因，我无法一一道来，部分是因为截至本书成稿，只发现了不到 8 000 个（尽管这个数字还在以每月几百个的速度增加）。还有一部分原因是它们中的大部分主要负责人体生化反应，描述起来既枯燥又无味。

我要向大家展示的是人类完整基因组中的惊鸿一瞥，在基因组里最有趣的地方稍作逗留，看看它们告诉我们的关于人类自己的一些事情。我们这一代是幸运的，因为我们是阅读《基因组》这本书的第一批人。能够阅读基因组，意味着我们能够更好地了解人类的起源、人类的进化、人类的本性以及人类的思维，这将超过迄今为止科学带给我们的关于人类的所有知识。它将给人类学、心理学、医学、古生物学和几乎所有其他学科带来一场革命。这并非宣扬基因就是一切，或者基因比其他任何因素都重要。但是，基因的重要性是毋庸置疑的。

本书的主题不是关于人类基因组工程——基因定位与测序技术的，而是讲述了这一工程有何发现。2000年6月26日，第一份人类完整的基因组草图的绘制完成，仅仅用了短短几年时间，我们就从对于人类基因几乎一无所知变得无所不知。我坚信，我们正在经历着人类智力活动史上无与伦比的伟大时刻。同时，也有人发出了不同的声音。他们认为，仅用基因无法涵盖人类的全部。我并不否认这一点，每个人所拥有的远远超过一份遗传密码。但在此之前，人类基因几乎是一个谜，我们将是第一批破解这个谜团的人——我们即将揭晓重要的新答案，但也将面对更多的新问题，这也是我希望在本书中呈现给读者的。

## 导论

前言的第二部分作为本书的导论。在这个部分中，我把与基因和基因作用机制有关的主题以词汇表的形式展现出来，并加以解释。希望读者快速浏览这部分内容，以便在阅读本书的过程中遇到不清楚的术语时，能够返回此部分查询。现代遗传学术语无数，错综复杂。本书力求少用技术术语，但有些

还是无法避免的。

人体有大约 100 万亿个细胞，大多数细胞直径不到 1/10 毫米。每个细胞内部都有黑色的一团，叫作细胞核。细胞核里有两套完整的人类基因组（卵细胞、精子细胞和血红细胞除外。精子细胞、卵细胞只有一套基因组，血红细胞内没有基因），其中一套来自母亲，另一套来自父亲。从理论上讲，每套基因组都有相同的 23 对染色体，上面有相同的 6 万~8 万个基因。但实际上，来自父亲和母亲的基因对上常常有着细小的差别，正是这些差别决定了人的眼睛是蓝色还是棕色。人类在生育时，来自父亲和母亲的染色体经过交换和重组，便将一套完整的基因组传给了下一代。

想象一下，“基因组”是这样的一本书：

- 全书共 23 章，每章都是一对**染色体**。
- 每章都包含几千个故事，每个故事都是一个**基因**。
- 每个故事由不同的段落组成，称为**外显子**。段落之间是广告，名为**内含子**。
- 每个段落由词语组成，叫作**密码子**。
- 每个词语由字母构成，叫作**碱基**。

如果基因组是一本书，那么这本书有 10 亿个单词，长度是本书的 5 000 多倍，或者是《圣经》的 800 倍。如果按照每秒一个词、每天 8 小时的速度把基因组读给你听，需要 100 年才能读完。如果把人类基因组写下来，每个字母一毫米，则总长度堪比多瑙河。基因组可以看作巨型的文档、巨大的图书、超长的菜谱，其所有内容都收录在比针尖还小的细胞的细胞核内。

严格来讲，将基因组看作一本书不能算作一个比喻，因为它确实是一本书。一本书即一组数码信息，其内容是线性、一维和单向的。一个个小小的字母符号，按照特定的组合顺序，根据一定的规则，拼合成有意义的词汇，汇集成一

本书。基因组也是这样的。不同之处在于，所有的英文书都是从左向右读，而基因组的某些部分从左向右读，某些部分又是从右向左读的，好在这两种情况不会同时出现。

（顺便提一句，这段之后，你将不会在本书中看到“蓝图”这个词了。原因有三。首先，我们读的是书。只有建筑师和工程师才使用蓝图，即便是他们，在计算机时代也开始放弃蓝图。其次，对于基因而言，“蓝图”是个很糟糕的类比。蓝图是二维的图示，而非一维的数字编码。最后，对于基因而言，“蓝图”的定义过于字面化，无法表达基因的内在对应关系。蓝图中的每一部分都对应着机器或建筑上的一个部分，但菜谱里的一句话并不对应蛋糕上要吃的那一口）。

英文书是用英文单词写成的，英文单词长短不一，由 26 个字母组成。而整个基因组是由三个字母的词写成的，且仅用了 4 个字母：A、C、G 和 T（分别代表腺嘌呤——Adenine，胞嘧啶——Cytosine，鸟嘌呤——Guanine 和胸腺嘧啶——Thymine）。而且，它们不是写在平面的纸张上，而是写在由糖与磷酸组成的长链上。人们将这条长链称为 **DNA 分子**，碱基就附着在长链的侧面，像螺旋的阶梯一样。每条染色体是一对非常长的 DNA 分子。

基因组是一本非常聪明的书，它懂得在合适的条件下复印自己、读出自己。复印的过程叫作**复制**，读出自己的过程称为**翻译**。基因组之所以能够复制，是因为它的 4 个碱基拥有一种独特的属性：A 总与 T 配对，G 总与 C 配对。所以，一条单链 DNA 将所有的 T 对应所有的 A，A 对应 T，C 对应 G，G 对应 C，依此类推，通过 A-T、C-G 互补配对的方式进行自我复制。实际上，DNA 的通常状态即为著名的**双螺旋**，由原来那条 DNA 链和与它互补配对的一条缠绕在一起形成。

这样，互补链再次进行复制，就把原来的内容带了回来。在这次复制中，



序列 ACGT 变成了 TGCA，在下次复制中又被转录回原来的 ACGT。这使得 DNA 能够无限制地复制下去，却永远携带同一套信息。

翻译的过程要更加复杂一些。首先，一个基因的内容经过相同的碱基配对程序被**转录**成一份副本，但这份副本并非由 DNA 构成，而是由 RNA 构成。RNA 是一种与 DNA 区别非常细微的化学物质。RNA 也携带一个线性密码，与 DNA 使用同样的字母，不同的是使用 U 代替 DNA 中的 T，表示尿嘧啶。这个 RNA 副本被称作“信使 RNA”，通过切除所有内含子并连接所有外显子的方式进行编辑。

之后，这个信使 RNA 与**核糖体**结合，核糖体是一种微小的结构，其自身的一部分也是由 RNA 构成的。核糖体沿着 RNA 移动，依次将由 3 个字母组成的密码子翻译成另外一个字母表，这个字母表代表着 20 种不同的**氨基酸**中的一种，每一种氨基酸由不同的分子带来，这类分子被称为**转运 RNA**。这些氨基酸首尾相连，形成一条与密码子顺序一致的链条。当全部信息都被翻译之后，氨基酸链将自己折叠成一个特殊的形状（具体形状由其序列决定），便形成了**蛋白质**。

几乎身体里的所有东西，从头发到激素，不是由蛋白质构成的，就是由蛋白质制造出来的。每个蛋白质都是被翻译出来的基因。需要特别指出的是，一种名为**酶**的蛋白质对身体里的化学反应起到催化作用。甚至 DNA 和 RNA 分子本身的加工、复印、纠错和组装（复制和翻译）也是在蛋白质的帮助下完成的。蛋白质还负责基因功能的开启和关闭，它们将自己附着于基因内容起始处附近的**启动子**和**增强子**，从而实现对基因开关的控制。不同的基因在身体不同的部位被开启。

基因在复制时，有时会产生错误。偶尔会丢掉或错误地加上一个字母（碱基）。有时整个句子或段落会重复出现、丢失或次序颠倒。这些被称为**突变**。许多突变既无害处也无益处，例如，一个密码子被改成另一个拥有相同氨基酸含

义的密码子。这是因为，总共有 64 个不同的密码子，但对应的只有 20 种氨基酸，这就意味着 DNA 中许多“词语”（此处指密码子）有着相同的含义。人类的每一代里会积累大约 100 个突变，因为人类基因组里有 100 多万密码子，这看上去并不算多。但是如果突变发生在错误的地方，即使只有一个，也可能是致命的。

凡事皆有例外（也包括人类基因）。并非人类所有的基因都在这 23 对主要的染色体上——有很少一部分存在于名为线粒体的小球里，并且，很有可能线粒体还是自由存在的细菌时就是如此。并非所有的基因都是由 DNA 组成——有些病毒由 RNA 构成。并非所有的基因都能够形成蛋白质，有些基因被转录为 RNA，但并不被翻译成蛋白质。这些 RNA 或者成为核糖体的一部分，或者成为转运 RNA，从而直接发挥作用。并非所有的化学反应都由蛋白质来催化，有少量是靠 RNA 催化的。并非所有的蛋白质都来自某个单独的基因，有些是由不同的基因组合而成的。并非所有的 64 个由三个字母组成的密码子都能够用来确定一个氨基酸——其中三个用来下达**终止**指令。最后，并非所有的 DNA 都能够形成基因，大多数 DNA 的序列是重复或者随机的，很少或从来不被转录，即所谓的无用 DNA。

了解完这些，让我们一起开启人类基因组之旅吧！



## 前 言

1号染色体  
生命 / 1

2号染色体  
物种 / 14

3号染色体  
历史 / 30

4号染色体  
命运 / 47

5号染色体  
环境 / 58

6号染色体  
智力 / 69

7号染色体  
本能 / 85

XY染色体  
冲突 / 101

8号染色体  
自身利益 / 115

9号染色体  
疾病 / 129

10号染色体  
压力 / 140

11号染色体  
个性 / 153

12号染色体  
自装配 / 166

13号染色体  
史前 / 178

14号染色体  
永生 / 188

15号染色体  
性别 / 199

16号染色体  
记忆 / 212

17号染色体  
死亡 / 225

18号染色体  
疗法 / 236

19号染色体  
预防疾病 / 250

20号染色体  
政治 / 263

21号染色体  
人种优化 / 278

22号染色体  
自由意志 / 293

参考文献 / 307



# 生 命

一死一生，川流不息；  
灭亡之后，振兴继之；  
一祸一福，起伏相寻；  
有如水中，忽生泡影；  
自起自灭，幻化无穷。

《人论》(亚历山大·蒲柏)

生命起源之初，地球汪洋一片，这时便有这样一个“词”，它携带着自己的信息，自我复制，永不停息，改变了整个海洋的组成。这个“词”能够重新排列化学物质的结构，以使它们从无序的环境中吸收负熵<sup>⊖</sup>，从而成为有序的生命体，地球从风沙肆虐的地狱变成了郁郁葱葱的天堂。最终，这个“词”发展到极致，创造出一种奇妙的机器——人的大脑，而大脑则意识到了这个“词”的存在。

每当思考这个问题时，我的大脑就像煮沸的粥一样，翻腾不停。地球距今已有 40 亿年的历史，我能够活在当今这个时代，是一种幸运——地球上 500 万个物种，我有幸成为一个有意识的人；地球上 60 亿人口，我有幸出生在发现这个“词”的国家；更加幸运的是，我与 DNA 这一宇宙中最伟大、最本质而又最惊人的秘密有着密切的关系：就时间而言，这一秘密揭示之后的第 5 年，我便出生了；就地理位置而言，我的出生地与揭示这个秘密的地点，不过 320 多公里；就生物关系而言，揭示这个秘密的，是我的两个同类——人。当然，你可以嘲笑我对此的狂热，认为我如此热衷于 DNA 这样一个英文缩写词，简直不可理喻。但是，请跟随我去探索生命的源头，你将相信这个“词”是那么迷人。

早在 1794 年<sup>1</sup>，博学的诗人兼内科医生伊拉斯谟斯·达尔文（Erasmus Darwin）曾提出这样一个问题：“在动物诞生之前，陆地和海洋就充满了各种植物；在某种动物诞生之前，其他动物就已经存在，据此，我们是否能够得出：所有的有机生命都源自同一种有生命的‘丝状物’？”在那个时代提出这样的假设，是令人震惊的。不仅因为他提出了“所有有机生命都有共同祖先”这一大胆的假说（65 年后，他的孙子查尔斯才出版了相同主题的书），也因为他使用了“丝状物”这个古怪的词语。而事实上，生命的秘密就藏在一条细丝里。

---

⊖ 一个无序的世界是不可能产生生命的，有生命的世界必然是有序的。生物进化是由单细胞向多细胞、从简单到复杂、从低级向高级进化，也就是说向着更为有序、更为精确的方向进化，这是一个熵减小的过程，可以说生物进化是熵变为负的过程。——译者注

问题是，一条“丝状物”怎么就能创造出有生命的东西呢？生命是很难定义的，但生命有两种能力：自我复制和建立秩序。有生命的东西能够产生与自身相似的副本。正所谓“龙生龙，凤生凤，老鼠的儿子会打洞”——兔子生兔子，蒲公英的下一代还是蒲公英。当然，兔子还会做些别的。这个世界本是随机而混沌的，但兔子吃草，却将其转化为骨与肉，从而形成有序而复杂的身体。这一转化过程并没有违反热力学第二定律——在一个封闭系统内，所有的事物都倾向于从有序变成无序。兔子不是一个封闭的系统，因此它通过消耗大量的能量，建立起一个复杂而有序的身体结构。用埃尔温·薛定谔（Erwin Schrodinger）的话来讲：生物从周围环境汲取秩序（从外界引入“负熵”）。

信息是生命这两种能力的关键所在。生命之所以能够自我复制，是因为存在着创造新的身体所需的各种信息。兔子的受精卵携带的信息可以“组装”一只新的兔子。生命通过新陈代谢建立起秩序，同样要依靠信息来创建和维护。如同在烘焙蛋糕时，要在准备蛋糕配方时就预先设计停当——一只只有繁殖能力和新陈代谢能力的成年兔子，也是由它的生命“丝状物”预先规划和决定好的。这一理念最早源于亚里士多德。他曾说过，形成一只鸡的原理都蕴含在鸡蛋里，一颗橡实包含着整棵橡树的信息。亚里士多德这种原始的信息理论观点，曾被化学和物理学埋没多年，随着现代遗传学的发现又被重新挖掘出来。马克斯·德尔布吕克（Max Delbruck）开玩笑道：DNA 是由这位古希腊哲学家发现的，为此应追授他诺贝尔奖。<sup>2</sup>

DNA 的“丝状物”就是信息，由化学物质的密码写成，每种化学物质即为一个字母。令人不可思议的是，DNA 的密码是以一种我们能够理解的方式写成的。遗传密码和书面英语一样，是一种线性的语言，沿一条直线书写。遗传密码就像数字一样，每个字母都有着重要的意义。和英语相比，DNA 的语言要简单得多，它的字母表里只有 4 个字母，即通常所说的 A、C、G 和 T。

当我们了解到基因其实是一种加密的信息后，便很难理解为何提出这种可能性的人少之又少。20 世纪上半叶，生物学界有个问题一直被反复提及——什么是基因？那时，DNA 简直就是无解之谜。DNA 的对称结构于 1953 年被发现，但让我们首先回到 10 年前的 1943 年。1943 年，那些将在 10 年后为破解 DNA 秘密做出卓越贡献的人，都正从事着其他工作。那一年，弗朗西斯·克里克（Francis Crick）在朴茨茅斯（Portsmouth）设计水雷；15 岁的“神童”詹姆斯·沃森（James Watson）刚刚被芝加哥大学录取，立志要倾其一生去研究鸟类学；莫里斯·威尔金斯（Maurice Wilkins）在美国协助研发原子弹；罗莎琳德·富兰克林（Rosalind Franklin）在英国政府工作，研究煤炭的结构。

1943 年，在奥斯维辛（Auschwitz）集中营里，约瑟夫·门格勒（Josef Mengele）为完成其“科学探究”，将一对对双胞胎折磨致死，可谓惨绝人寰。门格勒尝试去理解遗传的原理，但他的“人种优化论”被证明是伪科学，对于改善人类基因是没有用处的，他的实验结果对之后的科学更是毫无意义可言。

1943 年，在都柏林，有一个从门格勒那种人的手下逃出来的难民——大物理学家埃尔温·薛定谔，正在三一学院开展“什么是生命”的专题讲座。他知道染色体内蕴含了生命的秘密，但生命的秘密是以何种形式存储的，却不得而知。于是，他尝试着去解决这个问题：“就是这些染色体……使用某种编码，存储了一个人未来发育的全部信息，以及发育成熟后的各种机制和功能。”他说，基因太小了，应该归为一种大分子。他的这一见解似乎为这个问题提供了解决思路，包括克里克、沃森、威尔金斯和富兰克林在内的一代科学家受到启发，开始攻克难题。然而，尽管距离答案触手可及，薛定谔却偏离了方向。他钟爱量子论，认为可以从中找到这种分子承载遗传信息的原因。他固执地坚持自己的观点，却最终被证明走进了死胡同。生命的秘密与量子状态无关，无法从物理学中找出答案。<sup>3</sup>

1943年，在纽约，66岁高龄的加拿大科学家奥斯瓦尔德·艾弗里（Oswald Avery）的一个实验进入收尾阶段。这个实验即将决定性地证明DNA就是遗传信息的化学物表现形式。此前，他经过一系列精心的实验，证明一种与肺炎有关的细菌，通过吸收一种化学溶剂，就能从无害的变为有毒菌株。到了1943年，艾弗里已经得出了结论：发生转变的就是DNA。但在发表自己的成果时，他的表达过于保守，致使很长一段时间内无人关注。1943年5月，艾弗里写信给他的兄弟罗伊，信中的表述也仅仅放开了一点点：<sup>4</sup>

尽管尚待证明，但如果我们是对的，那就意味着DNA不仅在结构上重要，而且是一种功能活跃的物质，它对于细胞的生化活动和某些特性起着决定性的作用。这也意味着，有可能利用已知的化学物质，根据需要去改变细胞，并使这种改变遗传下去。这正是遗传学家们长期以来的梦想。

艾弗里几乎已经实现了这个梦想，但他的思考仍然仅局限在化学的层面。扬·巴普蒂斯塔·范·海尔蒙特（Jan Baptista van Helmont）在1648年做出一个猜想：“一切生命都是化学。”1828年，弗里德里希·维勒（Friedrich Wohler）说：“至少有些生命是化学。”那时，他刚利用氯化物和氰化银合成了尿素，从而打破了化学与生物学界之间不可逾越的界限，在此之前，尿素都是生物体产生的。说生命是化学，理论上是对的。但这个比喻让人兴致索然，就像有人说足球就是物理一样。粗略来讲，氢、碳、氧是构成生命的主要化学元素。生物体的98%都是由这三种原子构成的。真正令人兴致盎然的是生命中那些突出的特性，比如遗传性，而非组成生命的那些元素。然而，艾弗里是无法回答DNA是如何承载遗传特性的，因为这个问题无法从化学中得出答案。

1943年，在英国的布莱奇利（Bletchley），天才数学家阿兰·图灵在极度保密的环境下，亲眼见证了他最伟大的想法（首台可编程计算机）变成现实。图灵曾



论证过数字能够自己进行运算。为了破译德国军方的洛伦兹编码器，英国根据图灵理论，制造了一台计算机，命名为“巨人号”（COLOSSUS）。这是一台可编程存储程序的通用计算机。当时，没有一个人意识到图灵也许比其他任何人都更接近生命的秘密。图灵更是没有意识到，遗传物质本质上就是一种可编程的存储程序，新陈代谢就是一台通用计算机。将两者连接起来的是一种编码，是一种化学的、物理的，甚至无形的抽象信息。其奥秘就在能够进行自我复制。任何能够利用世界上的各种资源进行自身复制的事物，都是有生命的。这种信息最有可能以数码信息的方式呈现：可能是数字、程序脚本，或单词。<sup>5</sup>

1943年，在美国的新泽西州（New Jersey），克劳德·香农（Claude Shannon），这位默默无闻的学者，正在反复思考几年前在普林斯顿大学时的一个想法。他认为，信息和熵都与能量有着密切的关系，但属性却又是相反的。一个系统的熵越小，它所含的信息就越多。蒸汽机之所以能够吸收煤燃烧产生的热能，并将其转化为动能驱动机器运转，就是因为蒸汽机设计者为蒸汽机注入的信息量很大。人体也是如此。香农结合了亚里士多德的信息论和牛顿的物理学原理。和图灵一样，香农也没有考虑生物学的因素，但他的想法很深刻，远比堆积成山的物理化学理论更加接近“什么是生命”这一问题的答案。生命也是一种用DNA写成的数码信息。<sup>6</sup>

本章在开头曾提到，生命起源之初，就有那样一个“词”，但那个“词”并不是DNA。DNA是在生命诞生之后才出现的，此时生命已经有了两种分工，一种是化学反应与信息存储，另一种是新陈代谢与复制，这两种活动是独立进行的。但是，在DNA中记录了这个“词”的信息，并将其原封不动地传递下来，历经岁月变迁，直到今天。

想象一下在显微镜下人类受精卵细胞核的模样，如有可能，将23对染色体