

导 言

人类基因组——一套完全的人类基因——包装在23对独立的染色体^①里。在这些染色体中，有22对是大致按照大小来编号的，从最大的（第1号）到最小的（第22号）剩下那一对是性染色体：在女性体内是两条大的 X 染色体，在男性体内是一条大的 X 染色体和一条小的 Y 染色体。从大小上说，X 染色体在第七和第八号之间，Y 则是最小的。

“23”这个数字没有什么重要意义。包括我们的近亲猿类在内的许多物种都有更多的染色体，也有的物种染色体少得多。相似的基因或有着类似功能的基因也不一定都在同一条染色体上扎堆。所以，几年以前我从一台笔记本电脑上俯过身去与一位进化生物学家戴维·黑格（David Haig）说话的时候，很惊讶地听他说第十九号染色体是他最喜欢的染色体。他解释说，那上面有各种各样调皮捣蛋的基因。在那之前我还从来没有把染色体想成是 5 有性格的东西。归根结底，它们只不过是基因的随机组合。但是黑格这偶然的一句话在我头脑里植下了一个驱之不去的想法。为什么不试着在每一条染色体上都选一个与正在进行的故事相配合的基因，把正在我们面前展开的人类基因组——现在我们正在第一次发现它的细节——的故事按照染色体逐条地讲出来呢？普瑞

细胞核里的长条或环状的脱氧核糖核酸长链，绝大多数基因就在这些染色体上。因用碱性染料染色后在显微镜下观察时着色较重，易于分辨而得名。——译者注

默·利维(Primo Levi)^①在他的自传体短篇小说集中就用类似的方法讲述了元素周期表。他把自己生命中的每一章与一个元素联系起来，这个元素是他描述的那段时期里他接触过的。

我开始想象，人类基因组本身就是一种自传——用基因语言写成的一个记录，记录了从生命最初开始属于我们这个物种以及我们的祖先的偶然事件和发明。有一些基因从第一种单细胞生物居住在原始的“汤”里开始就没有怎么变化过；有一些基因是当我们的祖先还是线虫的时候发展出来的；有一些基因肯定是当我们的祖先是鱼的时候出现的；有一些基因之所以以它们目前的形式存在，是因为近期的一场流行病；还有一些基因可以被用来书写过去几千年中人类迁移的历史。从40亿年以前到仅仅几百年以前，基因组是我们这个物种的一种自传，当重要事件发生的时候把它们记录在案。

我写下了一个23条染色体的名单，然后在每一条旁边，我开始罗列人类本性中的一些主题。慢慢地，费了很大工夫，我开始寻找可以为我的故事作象征的基因。当我无法找到合适的基因或是发现了一个合适的基因它却在错误的染色体上的时候，我常感到恼火。还有一个难题是怎样处理X和Y染色体，现在我把它们放在第七条染色体之后了，这对于X染色体的大小来说很合适。现在你就知道了，为什么这本书在副标题里吹嘘自己有23章，最后一章却是第22章。

第一眼看上去，我干的这件事是最误导人的。也许我看上去是在暗示第一条染色体出现得最早，其实并不是；也许我看上去是在暗示第十一条染色体只与人的性格有关，其实并不是。在人类基因组里有大约3~8万个基因，我没法把它们都讲给你，这部分地是因为只有不到8000个被找到了（尽管这个数字以每月几百个的速度增加），还因为它们的绝大多数是枯燥无趣的生物化学

的中层管理者。

但是，我能给你的是一个表述清楚的对于全局的一瞥：在基因组里一些最有趣的地方短暂停留观光，看看它们告诉我们关于我们自己的什么事情。因为我们这幸运的一代是第一批阅读基因组这部书的人。能够阅读基因组比到今天为止科学的所有努力都能够得知更多关于我们的起源、我们的进化、我们的本性，以及我们的思维；它将要给人类学、心理学、医学、古生物学，以及几乎所有其他学科都带来一场革命。这不是要宣布基因里有着一切，或者基因比其他任何因素都重要。很显然并不是如此。但是它们重要，这是可以确定的。

这不是一本关于人类基因组工程——定位与测序技术——的书，而是关于这个工程发现了什么的书。2000年6月26日，科学家宣布他们完成了整个人类基因组的一份草稿。在短短几年里，我们将从对于我们的基因几乎一无所知变得无所不知。我确信我们正在经历人类智慧历史上无与伦比的最伟大的时刻。有些人会抗议说一个人的内涵绝非其基因所能囊括。我不否认：我们每一个人都有比一份遗传密码多得多的东西。但是此前，人类基因还几乎是完全神秘的。我们将是穿透那层神秘的第一代人。我们站在重要的新答案的边缘上，更重要的，是我们站在重要的新问题的边缘上。这是我在这本书里试图传达的。

入 门

导言的第二部分是一个类似叙述性的词汇表的入门简介，讲述基因及其工作方式。我希望读者在开始时先简略看一下，然后，如果遇到没有解释的术语，再回来看看。现代遗传学是个吓人的术语丛林。在这本书里，我力求少用技术术语，但是有一些是无法避免的。

人类的身体有着大约 100 万亿个细胞，大多数细胞的直径还

不到十分之一毫米。在每一个细胞里有一个黑团儿，叫做细胞核。在细胞核里有两套完全的人类基因组（只除了卵细胞和精子细胞，它们每个只带有一套，还除了血细胞——里面没有基因）。一套基因来自母亲，一套来自父亲。从原理上讲，每一套都有着相同的 23 条染色体，上面有相同的 3~8 万个基因。实际上，在父方与母方的两套基因之间常常有着小而细微的差异，例如，决定眼珠是蓝色还是棕色的差异。当我们生育的时候，我们传下去一套完整的基因组，但这只是在把父方与母方的染色体在一个叫做“重组”的过程中这一点那一点地进行了交换之后。

把基因组想象成是一部书：

全书 23 章，叫做染色体。

每一章有几千个故事，叫做基因。

每个故事是由段落组成的，叫做外显子，它们被名叫内含子的广告打断。

每一段是由词组成的，叫做密码子。

每一个词是由字母组成的，叫做碱基。

在书里有 10 亿个词，把基因组的书搞得比我这本书长 5000 多倍，或者说，是《圣经》的 800 倍那么长。如果我把基因组用每秒钟一个词、每天念 8 小时的速度念给你，要用 100 年才能念完。如果我把人类基因组写下来，每一个字母 1 毫米长，我这份文件会与多瑙河一样长。这是一份巨型文件，巨大的书，有着过长的“菜谱”，所有这些都塞进了一个很容易就能够放在针尖上的微小细胞里的更小的细胞核里。

把基因组看成是一部书的想法，严格来说都不是一个比喻。

- 8 它确实如此。一部书是一组数码式的信息，用线性、一维和单向的形式写成，并用一套密码通过它们的组成顺序把一个小小的字母表中的符号拼成一个巨大的有意义的词汇的表。基因组也是这

样。仅有的麻烦是所有英文书都是从左往右读的，而基因组的某些部分是从左往右读，某些部分又是从右往左读，尽管这两者不会同时出现。

（顺便提一句，在这一段之后，你就再也不会在这本书里看到“蓝图”这个字了，原因有三。第一，我们都用书，只有建筑师和工程师才用蓝图，而且就是他们，也在电脑时代开始放弃蓝图。第二，蓝图对于基因来说是个很糟的比喻。蓝图是二维的图，不是一维的数码符号。第三，蓝图对于遗传来说太过逐字逐句了，因为蓝图上的每一部分都对应着机器或建筑上的一个部分，而一本“菜谱”里的每一句话并不对应着蛋糕里的一口。）

英文书是用26个字母组成长度不同的词写成的，而基因组整个是用三个字母的词写成的，只用四个字母：A，C，G和T（它们代表腺嘌呤、胞嘧啶、鸟嘌呤、胸腺嘧啶）。而且，它们不是写在平平的纸上，而是写在糖与磷酸组成的名叫DNA分子的长链上，碱基就附着在这长链的侧面形成侧环。每一条染色体是一对（非常）长的DNA分子。

基因组是非常巧妙的一本书，因为在合适的条件下它可以复印自己也可以读出自己。复印的过程叫做复制，读出自己叫做翻译。复制之所以能够成功，是因为这四个碱基有一个相当巧妙的性质：A喜欢和T配对，G喜欢和C配对。所以，一个单根链子的DNA可以通过组装一根与自己互补的链子的方式来复制自己，用T对应所有的A，A对应所有的T，C对应所有的G，G对应所有的C。事实上，通常情况下DNA的状态是著名的“双螺旋”，由原来的那一条链与和它互补的一条缠绕而成。

这样，再复制一条与互补链互补的链，就把原先的“文本”带回来了。ACGT这样一个序列在复制版里成为TGCA，在复制版的复制版里又被转回了ACGT。这使得DNA能够无限地复制下去，却仍然带有同一套信息。

翻译要更复杂一点。首先，一个基因的内容要用同样的碱基

配对方法被转录成一份复制品，但是这一次，复制品不是由 DNA 组成，而是由 RNA 组成，一种与 DNA 有着非常细微区别的化学物质。RNA 也可以携带一个线性的密码，它与 DNA 使用的是同样的字母，只除了它在 T 的位置上用了 U，代表尿嘧啶。这个 RNA 的复制版被叫做信使 RNA，它通过把所有内含子都切除和把所有外显子（见前面所述）都连接起来的方式进行编辑。

然后，这个信使交上了一个朋友——一个叫做核糖体的微观机器，它本身也部分地由 RNA 组成的。核糖体沿着 RNA 移动，按着顺序把每三个字母组成的密码子翻译成另外一个单词表中的一个词，这个单词表有 20 个不同的氨基酸，每一个都由一个不同的叫做转运 RNA 的分子带来。每一个氨基酸都接在上一个的后面，形成一条与密码子顺序一致的长链。当整个信息被翻译之后，氨基酸链就把自己折叠成一个特殊的形状，这形状是由它的序列决定的。现在，它被叫做蛋白质。

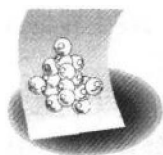
身体里几乎所有东西，从头发到激素，要么是由蛋白质组成的，要么是由蛋白质造出来的。每个蛋白质都是一个翻译出来的基因。具体点说，身体里的化学反应是由名叫酶的蛋白质来催化的。甚至 DNA 和 RNA 分子本身的加工、复印、改错和组装——复制和翻译——也是在蛋白质的帮助下完成的。蛋白质还负责把基因开启和关闭，这是通过把它们自己附着到基因内容的起始处附近的启动子和增强子来完成的。不同的基因在身体的不同部位被启动。

当基因被复制的时候，有时会出现错误。一个字母（碱基）会偶尔被丢掉，或者一个错误的字母被加上去了。整个的句子或段落有时会被重复一遍、被丢掉，或者顺序被颠倒。这些称为突变。很多突变是既无害也无益的，例如，如果它们把一个密码子改成了另一个，但是含义是同样的一个氨基酸：一共有 64 个不同的密码子，但只有 20 种氨基酸，所以很多 DNA 词汇有着相同的含义。人在每一代里会积累起大约 100 个突变，考虑到在人类基因

组里有100多万个密码子，这可能不算多，但是如果发生在错误的地方，即使是一个也可以致命。

所有规则都有例外（也包括这个规则）。不是所有人类基因都在23条主要的染色体上，有很少的一些住在名叫线粒体的小球里，而且，可能从线粒体还是自由生活的细菌时就是如此；不是所有的基因都由DNA组成，有些病毒用的是RNA；不是所有基因都是蛋白质的配方，有些基因被转录成RNA，却不被翻译成蛋白质——这些RNA直接进行活动，要么成为核糖体的一部分，要么成为转运RNA；不是所有的化学反应都由蛋白质来催化，有少量是由RNA催化的；不是所有的蛋白质都来自一个单个基因，有些是从几个配方拼凑起来的；不是所有的64个三个字母的密码子都确定一个氨基酸，相反地，有三个确定的是“停止”的指令。最后，不是所有DNA都拼写出基因，大多数DNA是混乱的重复序列或随机序列，很少或从来不被转录，即所谓的垃圾DNA。

这些就是你需要知道的所有东西。人类基因组之旅可以开始了。



第一号染色体 生命

一切归于腐朽之物皆源于他方
(一个接一个地,我们抓住生命的气息而后死亡)
如同产生于物质之海的泡沫
上升、破裂、重归海洋

——亚历山大·波普^①:《论人类》

太初有“词”。这个词以自己携带的信息充斥了整个海洋，永不停息地复制它自己。这个词发现了如何重组化学物质，以便抓住熵的潮流中微小的逆流并给它们以生命。这个词把我们这个星球上的陆地从布满灰尘的地狱变成了郁郁葱葱的天堂。最终，这个词到达了鼎盛期，巧夺天工地造出了一种粥样的、被称为人脑的机器。这个机器发现并意识到了这个词的存在。

每次我这么想的时候，我的那个粥样的机器就翻腾个不停。地球有40亿年的历史，我却幸运地活在当今这个时代；地球上500万个物种，我却幸运地生为一个有意识的人；地球上60亿人，我却荣幸地生在发现了这个“词”的国家；在地球所有的历史、地理环境与生物环境中，偏偏就在我出生的5年前、距离我出生的地方只有200英里处，我这个物种的两个成员发现了DNA的结构，从而揭示了宇宙中最大、最简单而又最惊人的秘密。如果你愿意，你可以嘲笑我的激情，就当我是个可笑物质至上者

① Alexander Pope: 17、18世纪英国诗人、讽刺作家。——译者注

吧：居然对一个缩写词^①都肯倾注这么大的热情。不过，跟着我到生命的源头去看看吧，我希望我能够让你相信这个词是多么迷人。

1794年，博学的诗人、医生伊拉斯谟·达尔文（Erasmus Darwin）这样问道：“远在动物存在之前，地球和海洋就充满了植物；远在某种动物存在之前，其他动物就已存在。在这种情况下，我们能否假设：所有的有机生命都源自于，且仍然产生于，同一种有活性的纤维？”^[1]这样一个猜想在那个时代被提出来，让人惊愕。不仅仅是因为“所有有机生命都有共同来源”。这一大胆假说比他的孙子查尔斯^②有关这一题材的书还早了65年，也是因为“纤维”这一古怪的用词。确实，生命的秘密就是在一条纤维里。

但是，一根纤维怎么就能创造出有生命的东西？生命是不大好定义的，但是所有生命都有两种能力：复制自己的能力和制造秩序的能力。有生命的东西都能够造出跟自己差不太多的拷贝：兔子生兔子，蒲公英生蒲公英。但是兔子还会干一些别的。它们吃的是草，却能将其转化成兔子的骨与肉，不知怎么一来，就在混沌随机的世界里造出了有秩序有复杂性的身体。它们并没有违反热力学第二定律——在一个封闭的系统里所有事物都倾向于从有序变成无序。这是因为兔子不是一个封闭系统。兔子是靠消耗大量能量才得以建立一个有序的、复杂的局部结构——它的身体的。用爱尔兰·薛定谔（Erwin Schrödinger）^③的话说：生命是从环境里“把秩序喝进来”的。

指 DNA。——译者注

查尔斯·达尔文：19世纪英国生物学家，1859年出版《物种起源》，提出所有生命共同起源、变异和各自选择使得物种多样化的观点，是现代进化学说的奠基人。——译者注

爱尔兰·薛定谔：20世纪著名物理学家，写过一本更为著名的小册子——《生命是什么？》。这本小册子启发过包括沃森和克里克在内的对生命的各种现象作研究的人的思路。——译者注

生命的两种能力，关键都在于信息。复制的能力之所以有可能存在，是因为存在一种“配方”，里面有制造一个新的身体所需要的信息。兔子的卵就带有组装一只新兔子的指南。通过新陈代谢来创造秩序同样也靠的是信息——用来建造和维修制造秩序的机器的指南。一只具有生殖能力和代谢能力的成年兔子，是由它的生命纤维预先规划设计好的，正如一个蛋糕是在烘蛋糕的配方里就规划设计好了。这个想法可以直接追溯回亚里士多德。他曾说过，鸡的“概念”是隐含在鸡蛋里的，而橡树把自己的计划直接传达给了橡实。亚里士多德的这种模糊的信息学观念，在被物理学与化学埋没了多年之后，又被现代遗传学重新发现。麦克斯·德尔布吕克（Max Delbruck）^①曾开玩笑地说：这位古希腊哲人应该因为发现了DNA而被追授诺贝尔奖。^[2]

DNA的纤维就是信息，是一种用化学物质的密码写成的信息，每一个字母都是一种化学物质。而且，DNA密码事实上是用一种我们能够理解的方式写的，这真有点令人大喜过望。就像书面英语一样，遗传密码是写在一条直线上的线性语言；就像书面英语一样，遗传密码是数码式的，意思是说每一个字母都同等重要。更有甚者，DNA的语言比英语简单多了，因为它的字母表里只有四个字母，按惯例被称为A、C、G和T。

当我们现在知道了基因就是用密码写的“配方”之后，就很难想象在过去只有那么少的人曾经想到过这一可能性。20世纪的上半叶有一个没有被回答的问题在生物学里一再出现：什么是基因？当时，基因简直是神秘莫测。让我们回到——不是DNA对称结构被发现的1953年，而是此前10年——1943年。10年之后在破解DNA的秘密上做了最突出工作的人，那时候都在干别的。弗兰西斯·克里克（Francis Crick）当时在朴次茅斯（Portsmouth）

麦克斯·德尔布吕克：20世纪美国遗传学家，现代遗传学的奠基人之一。

——译者注

那边设计水雷；只有15岁的“神童”詹姆斯·沃森（James Waston）刚刚在芝加哥大学注册读本科，而且已立志用自己的一生去研究鸟类学；莫里斯·威尔金斯（Maurice Wilkins）在美国协助研制原子弹，罗萨琳·富兰克林（Rosalind Franklin）则在替英国政府工作，研究煤的结构。^①

还是1943年，在奥斯维辛集中营，约瑟夫·门格尔（Josef Mengele）^②正对孪生子们进行致命的折磨，他的“科学研究”其实是对科学研究的一种极其恶劣的嘲讽。门格尔是在试图理解遗传学，但他的优生人种论已经被证明不是正确的途径。门格尔的实验结果对他之后的科学家是没有用处的。

1943年，在都柏林，一个从门格尔那种手下逃出来的难民、大物理学家埃尔温·薛定谔，正在圣三一学院讲授一个名为“什么是生命”的系列讲座。他是在试图定义一个问题。他知道染色体载有生命的秘密，但是他不知道染色体是怎样储存生命秘密的。“就是这些染色体……以某种密码写就的程序，储存了每一个体发育的整个模式，以及发育成熟之后每一个体应有的功能。”他说，基因那么小，小得不可能是任何其他东西，而只能是一个大分子。他的这一见解影响了一代科学家——包括克里克、沃森、威尔金斯和富兰克林——去攻克一个顿时不再是无从下手的难题。但是，已经如此接近答案的薛定谔却偏离了轨道。他认为这个大分子之所以能够成为遗传物质的载体，是由于他心爱的量子理论。而他自己这个想法执迷的研究最后被证明是走进了一条死胡同。生命的秘密跟量子没有任何关系。关于生命的答案并不出自物理学。^{〔3〕}

这四人在1953年发现DNA结构上贡献最大的科学家。罗萨琳因罹患癌症于1958年去世，年仅37岁，另外三人于1962年获得诺贝尔生理学及医学奖。——译者注

约瑟夫·门格尔：第二次世界大战期间最臭名昭著的纳粹医生，在波兰的奥斯维辛集中营在犯人身上进行人体实验，对犯人进行了惨无人道的折磨，有“死亡天使”的称号。——译者注

1943年，在纽约，一位66岁的加拿大科学家奥斯瓦尔德·埃弗里（Oswald Avery），正在对一个实验进行最后的调整。这个实验将决定性地证实 DNA 是遗传的化学表现。这之前他已经发现，仅靠吸收一种化学溶液，一种肺炎菌就能从无害转变为有害。到了1943年，埃弗里已经总结出：发生了转变的东西就是 DNA。但是他在发表自己结果的时候，表达得过于谨慎，以至于一段时间之内都没几个人注意到他的成果。在1943年5月写给他兄弟罗伊的信里，埃弗里也只比以前稍稍大胆了一点：〔4〕

如果我们是正确的（当然，这一点还有待证明），那就意味着核酸（DNA）并不仅仅是结构上重要，而是功能上活跃的物质，能够决定细胞的生化活性与特性。那么，就有可能利用已知的化学物质去准确地改变细胞并使这种改变遗传下去。这是遗传学家长期的梦想。

埃弗里几乎已经走到这一步了，不过他仍然只是从化学的角度在思考。简·巴普提斯塔·冯·赫尔蒙特（Jan Baptista van Helmont）^①在1648年说过：“一切生命都是化学。”但这只是一种猜想。1828年，弗雷德里克·维勒（Friedrich Wohler）说：至少有些生命是化学。那时他刚用氯化氨和银的氰化物合成了尿素，从而打破了一直以来化学的世界与生物的世界之间不可逾越的界限。在他之前，尿素是只有生物体才能制造出来的东西。“生命就是化学”这句话是对的，不过也很煞风景，就像谁说足球就是物理一样。大概计算一下，生命可以说是三种原子的化学。生物体中98%的原子都是氢、氧和碳。但是，生命整体的特性，比如说遗传性，才有意思，而不是组成生命体的每一个零件。埃弗里

简·巴普提斯塔·冯·赫尔蒙特：17世纪比利时化学家、生理学家、医生。——译者注

想象不出来，是DNA的什么化学性质使它能够载有遗传性的秘密。这个问题的答案也不是从化学来的。

1943 年在英国布莱奇利 (Bletchley)，一位天才数学家艾伦·图灵 (Alan Turing) 正在眼看着他最有洞察力的一个想法在绝密环境下变成真实的机器。图灵论证过：数字能够自己进行运算。为了破解德国军队洛伦兹编码器的秘密，英国制造了一台建立在图灵理论上的计算机：克劳索斯。这是一台多功能机器，有可以修改的内存程序。当时没有人意识到图灵也许比任何人都更接近生命的秘密，图灵自己更是没想到。遗传，实际上就是一种可以修改的内存程序；新陈代谢就是多功能的机器。把两者连接起来的是一种密码，是以物理的、化学的，甚至是非物质形式存在的一种抽象信息。它的秘密就在于它能够复制自己。任何能够利用这世界上的资源把这密码复制出来的事物，就是有生命的东西。这种密码最可能的存在方式是数码方式：一个数字，一个短程序，或是一个词。^[5]

1943 年在新泽西州，一个有点与世隔绝的沉静的学者，克劳德·香农 (Claude Shannon)，正在琢磨一个他几年前在普林斯顿 (Princeton)^① 的时候想到的想法。香农的这个想法是说，信息和熵是一个硬币的两面，两者又都与能量有紧密的联系。一个系统的熵越小，它所含的信息就越多。蒸汽机之所以能够收集煤燃烧发出的能量并把它转化为旋转运动，是因为蒸汽机本身有很高的信息含量。人体也是如此。亚里士多德的信息理论与牛顿的物理学在香农的大脑中相遇了。像图灵一样，香农也根本没有想到生物学。但是香农这一深刻的想法，却比堆积如山的物理学与化学理论更接近于“什么是生命”这一问题的答案。生命也是数码信息，是用 DNA 写成的。^[6]

太初有“词”这个词却不是DNA。DNA的出现，是在生命已

经出现之后，在生物体已经把两种活动——化学反应与信息储存，新陈代谢与复制——分工进行之后。但是DNA一直存着这个“词”的一份纪录，在漫长的岁月里将其忠实地传递下来，直到今天。

想象一下显微镜下一个人类卵子的细胞核。如果有可能的话，你可以把23对染色体按大小重新排列一下，大的在左边，小的在右边。现在在显微镜下放大一下最左边的一根——纯粹是随意地，这根染色体被称为一号染色体。每一根染色体都有一条长臂和一条短臂，由一个被称为中心体的窄节所连接。如果你仔细地读，你会发现，在一号染色体的长臂上接近中心体的地方，有一串长约120个字母（A、C、G和T四种字母）的序列，重复出现了很多次。在每两个这种序列之间，是一些没有什么规律的“文字”，但这120个字母组成的“段落”却像一段耳熟能详的乐曲一样重复出现，总共出现了100次以上。阅读这种“段落”也许就是我们与最初的“词”最接近的时候。

这个短“段落”是一个小基因，它也许是人体内最活跃的一个基因。它的120个字母不断地被制成一小段RNA，称为5S RNA。它与其他一些RNA、一些蛋白质被仔细地缠在一起，住在一个名叫核糖体的结构里。核糖体是把DNA配方翻译成蛋白质的地方。而蛋白质又是使得DNA能够复制的东西。借用萨缪尔·巴特勒（Samuel Butler）^①的风格，我们可以说：蛋白质就是一个基因用来制造另一个基因的手段，基因就是蛋白质用来制造另一个蛋白质的手段。厨师需要做菜的菜谱，而菜谱也需要厨师。生命就是蛋白质和基因这两种化学物质的相互作用。

蛋白质代表的是化学反应，是生命活动、是呼吸、是新陈代谢、是行为——生物学家们称为“表现型”的那些东西。DNA代表的是信息，是复制、是繁殖、是性活动——生物学家们称为“基因型”的那些东西。两者都不能单独存在。这是一个经典的“先

有鸡还是先有蛋的问题是先有基因还是先有蛋白质 先有DNA是不可能的，因为DNA只是一件含有些数学信息的无生气的东西，不能催化任何化学反应，非得有其他东西帮忙不可。先有蛋白质也不可能，因为蛋白质虽然能进行化学反应，却不能精确地复制自己。这样看来，不可能是DNA创造了蛋白质，也不可能是蛋白质创造了DNA。如果不是最初的那个“词”在生命的纤维中留下了一点淡淡的痕迹，这个谜团也许会一直让人觉得奇怪和糊涂。正如我们现在已经知道的，蛋是在鸡出现之前很久就有了的（爬行类动物是所有鸟类的祖先，它们是下蛋的），现在也有越来越多的证据表明在蛋白质存在之前有RNA。

在当代，RNA是把DNA和蛋白质这两个世界联系起来的一种化学物质。它的主要作用是把信息从DNA语言翻译成蛋白质语言。但是，从它的行事特点看来，它几乎毫无疑问地是二者的祖先。如果DNA是罗马城，RNA则是希腊；如果DNA是维吉尔（Vivgil），RNA就是荷马。^①

RNA就是那个“词”。RNA留下了五条线索，使我们看到了它是先于DNA和蛋白质的。直到今天，要想改变DNA序列中的任何组成部分，都是通过改变RNA序列中相应的组成部分而完成的，没有更直接的办法。而且，DNA语言中的字母T是从RNA语言中的字母U造出来的。现代的很多酶，虽然是蛋白质，但它们要想正常发挥功能却离不开一些小的RNA分子。更有甚者，RNA与DNA和蛋白质还有不同的一点，就是RNA能够复制自己，不需要任何外界帮助：给它正确的原料，它就能将其织成一条信息链。不管你观察细胞的哪一部分，最古老最基本的功能都需要RNA的参与。基因中的信息是以RNA的形式被一种需要RNA才

历史上，古希腊文明出现在罗马之前。荷马是古希腊诗人，以史诗闻名，被认为是对欧洲文学影响最大的人。维吉尔是罗马诗人，亦擅长史诗，对于欧洲文学的影响仅次于荷马。——译者注

能正常工作的酶提取出来的。这个信息，是由一台含有RNA的机器——核糖体翻译出来的。而在翻译过程中需要的氨基酸，又是一种小小的RNA分子给搬运过来的。在所有这些之上，还要加上一条，与DNA不同的是，RNA可以做催化剂，可以把分子——包括RNA——打断或是连上。它可以把RNA分子切断、连上，造出RNA的组成成分，把一条RNA链加长。一个RNA分子甚至可以在自己身上做“手术”，把自己的某一段切除，再把两个自由端接在一起。^[7]

20世纪80年代早期，托马斯·赛克(Thomas Cech)和西德尼·奥特曼(Sidney Altman)^①发现了RNA的这些惊人特性，从而彻底改变了我们对于生命起源的理解。现在看来，最早的基因，“原基因”，很有可能是复制与催化合为一体的，是一个消耗自己周围的化学物质以复制自己的“词”。它的结构很有可能就是RNA。把任意一些RNA分子放在试管里，然后一遍遍地选出它们中间催化作用最强的成员，就可以重现RNA从什么也不是到具有催化作用的“进化”过程——几乎可以说是又进行了一次生命起源。这种实验最惊人的结果之一，就是最后得到的RNA往往含有一段序列，读起来酷似核糖体RNA基因——比如说，一号染色体上的5S基因——的序列。

在第一只恐龙出现之前，在第一条鱼出现之前，在第一条虫子、第一棵植物、第一种真菌、第一种细菌出现之前，世界是RNA的世界。这大概是40亿年前，地球刚刚形成不久，宇宙也仅仅有100亿年历史的时候。我们不知道这些“核糖生物体”是什么样子的。我们只能猜想它们是怎样“谋生”的——从化学意义上说。我们不知道在它们之前有什么，但从存留在今天的生物中的线索

托马斯·赛克和西德尼·奥特曼：当代美国生物学家（后者出生于加拿大）因他们在RNA功能方面的工作于1989年共获诺贝尔化学奖。

——译者注

看来，我们可以比较肯定地说 **RNA** 世界确实存在过。^[8]

这些“核糖生物体”面临着一个大问题。**RNA**是不太稳定的物质，几小时之内就可以解体。如果这些“核糖生物体”去了比较热的地方，或是试图长得比较大，它们自己的基因就会迅速坏死，遗传学家们称为“由错误而引起的灾难”。后来，它们中的一个从试验与错误中发明了一种新的、更“坚强”的**RNA**的变种：**DNA**。它还发明了一套从**DNA**复制**RNA**的系统，包括一种我们称为“原核糖体”的机器。这套系统既要快速又要准确，于是它把遗传信息连在一起的时候每次连三个字母。每个三字母的小组都带有一个标签，使得它能够更容易地被“原核糖体”找到。这个标签是氨基酸做的。很久以后，这些标签被连在一起，制成了蛋白质，而那些三个字母的“词”，则成了制造蛋白质的密码——遗传密码。（所以直到今天，遗传密码每个词都有三个字母，作为制造蛋白质的配方的一部分，每个词拼出20个氨基酸中的一个。这样，一个更复杂的生物就诞生了。它的遗传配方储存在**DNA**里，它体内的各种“机器”是蛋白质做成的，而**RNA**则在两者之间架起一座桥梁。

这个生物名叫露卡（**Luca**）——所有物种在分化之前最后的一个共同祖先。^①它长得什么样子？住在什么地方？传统的回答是：它长得像个细菌，生活在一个离温泉比较近的温暖的水塘里，或生活在浅海湾里。不过，在过去的几年里比较时髦的做法是给露卡一个环境比较险恶的住处，因为变得越来越清楚的是，地下与海底的岩石上存在着亿万种以化学物质为养分的细菌。现在一般认为，露卡存在于地下极深的地方，存在于火成岩的裂缝里，“吃”硫、铁、氢和碳为生。直到今天，生活在地球表面的生物仍然只是地球所有生物中薄薄的一层。地下深层那些喜热细菌——也许就是造就天然气的那些物质——体内含有的碳的总量，

原文是 **The Last Universal Common Ancestor** 缩写为 **LUCA**。——译者注