

基因熵

兼论物种的退化

约翰·圣弗德 著

繁星 流萤 译

基因熵
兼论物种的退化

Genetic Entory

第4版

ISBN-13 978-0-9816316-3-9
ISBN-10 0-9816316-3-0

Copyright © 2005, 2006, 2008, 2014
Dr. John C. Sanford

版权所有：约翰·圣弗德博士

非经圣弗德博士授权，或有明确的法律允许，不得以电子、机械、影印、录音或其他任何方式翻印、储存或散播本出版物的任何部分。

牧我羊出版社
牧我羊基金会之分支

作者简介

约翰·圣弗德博士任康奈尔大学教授已愈30载。他早年在威斯康星大学获得植物育种和植物遗传学博士学位。在康奈尔大学任教期间，他曾在纽约州立农业试验站培养研究生并开展遗传学研究工作。期间圣弗德曾使用传统方法培育新作物品种并随后投身于植物基因工程这一新兴领域。圣弗德已有100余篇科学出版物面世并取得数十项专利。他早期在科学上最重要的贡献包括三项发明：生物爆破（“基因枪”）法、抗源病原体 and 基因免疫。现今世界上一大部分的转基因作物（就数量和面积而言）使用的就是由圣弗德博士和他的同事们发明的基因枪技术。圣弗德同时也从他的研究中创建了两家生物技术公司，生物爆破公司和圣弗德科学公司。圣弗德仍然是康奈尔大学客座教授，但基本上已经退休。他现创办了一家小型非盈利机构，牧我羊基金会。

通过牧我羊基金会，圣弗德在过去的14年中从事理论遗传学和生物信息论方面的研究。本书就是从这些研究中得出的“初熟的果子”。

基因熵

献辞与致谢

本书的撰写完全是出于上帝的恩典，他是一切良善的源头。谨将此书献给我已故的研究生导师鲍勃·汉拿门博士。在科学上他曾激励我，在信仰和敬虔上他也是我的榜样。我要感谢我的妻子海伦一向坚定的支持。感谢在我之前勇敢地质疑过“基本公理”的许多科学家。特别感谢过去十年中曾与我在科学上合作的同事，他们用自己的研究有力地验证了本书中各个论点。这些同事包括沃特·赖敏、约翰·邦穆伽德纳、卫斯理·布鲁尔、保罗·吉布森、鲍勃·卡特、弗兰新·史密斯、柴斯·乃尔森和克里斯·路普等。感谢罗伊德·海特和克里斯·路普的艺术设计。

J.C. Sanford

译者引言		i
序		v
前言		ix
第一章	生命之书从哪里来？ 简讯：基因组是一套说明书	1
第二章	随机突变是好事吗？ 简讯：随机突变不断破坏信息	13
第三章	多少突变才算太多？ 简讯：人类基因突变率太高了	31
第四章	全能的选择来拯救？ 简讯：选择的能力非常有限	43
第五章	基因组选择可行吗？ 简讯：选择不能拯救基因组	65
第六章	进一步审视噪音 简讯：问题比你所想的要严重	81
第七章	假想的挽救机制 简讯：假想的机制经不住现实考验	95

第八章	人类可以自救吗? 简讯: 优生学/克隆不能阻止退化	109
第九章	自然选择能创造吗? 简讯: 突变加选择连一个基因也创造不出	115
第十章	果然是向下沉沦? 简讯: 一切证据指向基因退化	137
第十一章	综述近期主要科学发现 简讯: 基因熵的证据日益增强	139
后记	希望在哪里? 简讯: 有希望	165
更多资料		169
附录1	一个根深蒂固的观念	171
附录2	选择能同时作用于多少核苷酸?	193
附录3	合一现象和整合复杂性的概念	199
附录4	基因复制和多倍体能增加遗传信息吗?	203
附录5	可能存在的四条异议	209
附录6	学术界对这些发现有何反应?	219
词汇表		223
参考文献		233

译者引言

流萤

达尔文在《人的发源与性别选择》一书的结尾处写道：“人类为升到了有机界的巅峰而感到骄傲是可以原谅的，虽然这不是他自己努力的结果。既然他是如此升到了这个地步，而非一开始就被放在这里的，这一事实或许会给他一种希望，就是在遥远的未来要达到一个更高的境界。”英国剑桥大学的天文物理学家马丁·李斯爵士(Sir Martin Rees)在泰德网站(Ted.com)上谈到太阳末日的时候说：“(当太阳燃尽的时候)，世界上存在的任何生命和智慧将与我们截然不同，其差异就如我们与细菌的区别。”中国的文豪鲁迅、林语堂等都曾深受达尔文主义的影响而将希望寄托于子孙后代。许多人，包括我自己，也曾欣赏林语堂关于改祖先崇拜为子孙崇拜的革命性建议。可是在本书中，康乃尔大学杰出的遗传学家圣弗德博士却总结近几十年来遗传学的发现，尤其是最近对人类基因组（人体细胞内的全部基因）的直接观察，提出人类正在不可逆转地退化。在不久的将来，我们甚至可能面临彻底的绝种！

结论似乎惊世骇俗，证据却是显而易见的。首先，人类基因变异之快超出早年遗传学家的想象。半个世纪以前遗传学家还以为每代每人产生不到一个突变。但这个数目与时俱增，到2000年前后普遍认为每代每人新增的点突变为几百个。近年对人类基因组的研究更发现每个人都携带着与生俱来的成千上万的突变。如此高的突变率怎能不让我们为人类的基因组担忧

呢？其次，绝大多数突变是轻度有害的。由于遗传信息的精密设计，突变意味着对信息的破坏。随机突变从来不会产生有意义的信息。但由于基因组在设计上的冗余和生物体强大的缓冲能力，大多数的突变并不会立即影响机体的功能，而是在子孙后代中潜移默化。单个突变常常无关大局，但数目众多的突变却使我们的基因组逐渐朽坏。个体衰老就是突变的一个后果。正是由于人体80万亿个细胞中日积月累的突变，你我不知哪一天发现自己青春不再。由于生殖细胞中的突变代代相加，人类作为一个物种不知哪一天也将日薄西山！第三，自然选择对人群中数不胜数的突变基本上无能为力。选择是对抗突变的唯一力量，但所谓选择不过就是差异性繁殖，即有的人繁殖率高，有的人繁殖率低。选择排除一个突变的效率有赖于该变异对个体繁殖能力的影响。有多少突变真正决定着一个人所生子女的数目呢？那些子孙满堂的人的基因组就一定比别人的优秀吗？而且由于好基因与坏基因在基因组中紧密连锁，好坏之间就更难解难分了。每个人都有好基因，每个人都有坏基因，你我的基因组总的来说是“半斤八两”，生育子女的多少全看个人选择和机遇了。第四，即使有利突变会偶尔发生，也很难通过自然选择让它取代群体中的非突变基因。基因频率在群体中经选择而改变是如此缓慢，而有害突变的发生和基因频率的随机漂移是如此迅速，要靠选择有利突变来改善人类的基因组犹如螳臂挡车。

自然选择着眼于整个生物体的生育能力，而基因突变却是在分子水平上，两者之间相隔着重重的缓冲屏障。本书中有

一个“制造加工厂”的比喻，自动组装生产线起始处是一个高度近视错误百出的“程序誊写员”（细胞核内的基因复制），组装线末端有一个高度远视大大咧咧的“质量控制员”（自然选择），全靠这两位老兄怎能不让“进化”出的产品质量每况愈下？除了致病、致畸、致死的少数突变，大部分突变因损害轻微（接近中性）而为选择所忽略，在人群中代代积累，逐渐地、不可逆转地侵蚀着人类基因组中的有用信息，使基因组的混乱程度增加，这就是本书作者所称的“基因熵”。如热力学中的熵一样，基因熵在自然过程中只会增大，不会减小，基因组一步步走向无序，信息趋向于零。正如汽车外壳的锈蚀和零件日复一日的磨损导致整部车子迟早寿终正寝，基因组信息消失的过程也就是一个物种灭亡的过程。

本书的逻辑特点是定量分析。除了突变率、突变的“适合度效应”（即突变对繁殖率的影响）、基因频率等，作者还分析了选择的代价，而这一点正是进化理论家所经常忽略的。选择的代价就是过剩的人口。不要以为当今世界最不缺的就是人口，切记自然选择的对象不是整个人群，而只是有繁殖能力的年轻人和他们的后代。为了淘汰有害突变，携带突变基因的人口必须夭亡、独身、不生育或者少生育。不说许多国家因生育率低已经出现人口结构的老龄化，全世界的平均生育率也不能提供足够的盈余人口以淘汰如洪水猛兽而来的众多突变。因为突变数目大但盈余人口有限，针对每一个有害突变的实际选择力度（淘汰人口的百分比）只能很小，否则就难以维持繁殖人群的数量，反而会加快人类灭绝的步伐。基因组的迅速变异和自然选择的相对松弛最终势必导致生育能力的下降，而低生

育能力又使选择淘汰成为奢谈。这种恶性循环叫做突变熔消，这正是目前濒危动物所处的尴尬境界。谁知人类离突变熔消还有多远？作者在“人口爆炸”的今天谈人类灭绝，听起来好象痴人说梦，但遗传学的事实不容忽视。

此外作者在书中还考察了各种环境和生物学噪音对选择的干扰，剖析了旨在挽救达尔文主义的选择理论，并指出了人工选择的不可行和克隆人的无济于事。圣弗德博士以自己多年科学研究中练就的敏锐目光和严谨的思维分析，捅破了“突变加选择”作为人类基因组的创造者兼守护者的神话。

人类基因组的来源和维持机制，是基因组研究中最大的奥秘，但很多科学家被新达尔文主义蒙蔽而对此视而不见。所以并不是所有的遗传学家都赞同本书提出的观点（其实一些人感同却没有这个勇气），但基于圣弗德博士对遗传学和农学的突出贡献，基于本书所列举的事实、数据、和严谨的逻辑，每一位生物专业人士，每一位学过进化论的人，每一位对人类未来感兴趣的学者，都应该仔细阅读本书中关于基因熵的概念和基因组退化的现象。读过此书以后，您对现代生物学的重大发现、对遗传性和退行性疾病的发展趋势、对一些耳熟能详的达尔文主义教条、以及对人类的未来走向，必会有一个新的视角。

流萤曾获上海医科大学医学硕士和美国俄亥俄大学生物学博士，现在佛吉尼亚州一所私立高校任教。

序

约翰·邦穆伽德纳博士

在过去半个世纪中，科学界为人类开启了一扇门，进入了一个梦幻般的小人国。在这个国度里有一个自我复制的自动化工厂。工厂里各部件以每分钟上万转的速度运行，有自动化包装标记系统，有运输和分配系统，更有复杂的监测和反馈控制系统。这当然就是细胞和分子生物学的世界。在活体细胞中，数以万计不同类型的复杂纳米机器上演着令人难以置信的化学壮举。在如此复杂的细胞之上，还有同样复杂的由万亿细胞惊人地协同运作的有机体。在高等生物中还有一个由万亿神经交错连接的大脑。面对如此令人瞠目结舌的复杂性，善思考者自然会问：“这一切是怎么来的？”这个问题的标准答案被本书的作者称为“基本公理”（那就是，生命不过是偶然性的突变经过自然选择筛选而来）。

《基因熵》对于“基本公理”的基础做出了探索性的分析。它着重讨论了决定生命之惊人复杂性的基因软件。作者指出，在高等生物，当然也包括人类，基因信息的总体（称为基因组）是巨大的。基因组不仅规模庞大，而且也极其复杂，充满了环路和分支，基因与基因层层控制。在许多情况下，同一个基因字符串在不同环境中可能传达完全不同的信息。如此惊人的信息结构是怎样发源的显然是个重要问题，但作者提出了另外一个问题，即面对每一代人中新产生的大量有害突变所造成的不断侵蚀，人类基因组何以竟能保存下来？

本书作者是一个职业遗传学家，但他承认有许多年他并没有发现基本公理的理论基础存在严重的问题。他认为自己在做研究生的时候接受了这一理论，主要是源于对权威的信任，而不是真正个人的理解。作者觉得自己当时别无选择，因为那时候他认为这个抽象而且高度数学化的领域已经超出了自己的评判能力。直到后来，他才开始意识到公理所依赖的关键假设在严谨的分析面前是多么的不现实和脆弱。作者认为，今天大多数职业生物学家们正像从前的他一样，多数人并没有意识到基本公理存在严重的问题。这是因为公理的基本假设无论在研究生课堂里、研究生教科书中、还是在专业文献中都没有经过认真的评判。

群体遗传学家为生物学界提出的理论模型虽然披着高雅的数学外衣，但在其根本上却有很多不合理的假设。基本公理正是依赖于这些假设的支撑，可是大多数职业生物学家并没有认识到这一点。

群体遗传学主要涉及复杂的数学模型，试图说明基因突变是如何在其出现后一代代遗传下去，以及突变如何影响每一代中个体的生存。自然地，这些理论模型的现实性主要取决于它们所依赖的假设的真实性。在本书中，作者揭露了大部分已经使用了75年的最关键假设都具有明显的不现实性。就如作者在他初期的职业生涯中一样，大多数职业生物学家对于基本公理的信心正是来源于这些理论模型的推论，尽管这些模型是基于不合事实的假设。今天多数生物学家们并不知道他们在研究生院里学到的这些群体遗传学的结论已经经不起科学的推敲了。因此大多数人很难想象，如果从现实的假设出发，群体遗传学

实际上恰恰否定了基本公理。

《基因熵》精辟地解析了基本公理的不实。它的写作风格既有挑战性又简单易懂，对于遗传学或生物学略有了解的非专业读者也能理解。与此同时，书中翔实的内容和完整的文献足以让受过最高训练的生物学家严肃地反省他/她对于基本公理的一贯信念。我认为此书值得全世界每一位生物学专业人员和生物学教师阅读。该书有可能让整个学术界的观点发生深层次的变革。

约翰·邦穆伽德纳在加州大学洛杉矶分校获得地质物理学博士学位，曾在劳山拉莫斯国家实验室理论部从事了20年的研究工作。他还获得了普林斯顿大学的电机工程硕士学位，在那里他第一次接触到信息理论并认识到信息理论对生物系统的影响。他是复杂数值模拟的专家，曾参与设计电脑软件《孟德尔的会计师》，这是目前对突变/选择过程的最切合实际的模拟软件。

前言

回顾过去，我意识到自己曾为了争辩一些无关紧要的事情而浪费了很多时间。所以我由衷地希望在本书中能探讨一些真正有意义的问题。我们究竟是什么？从哪里来？到哪里去？我认为这些话题至关重要。这就是本书的真正主题。

现代思想有一个核心前提，即人类只是无意义的自然过程（无指导的进化）的产物。这种学说被广泛传授，而其逻辑结论使我们相信，我们只是些毫无意义的装满分子的皮囊。分析到最后，万事都无所谓。如果这一学说是假的，那么它就是人类设计出来的最阴险和最具毁灭性的思想体系。然而，如果它是真的，它也是毫无意义，正如其他一切事物一样。在当今知识界中广为流传的整个思想体系，其理论基础就是达尔文主义所认为的无指导、无意义的进化过程。

现代达尔文主义的根基，我把它称为“基本公理”。基本公理认为，人类不过是随机突变加上自然选择的产物。在我们的学术界，基本公理被广泛地传讲、广泛地接受。它是不断念诵的经文，在每个大学校园里无尽地重复着。很难找到哪个大学的哪个教授会考虑（或者说敢于）去质疑基本公理。正是由于这个原因，绝大多数曾经相信生命不仅仅是化学反应的年青人在大学里失去了信仰。我相信这也造成了我们在整个社会文化中所看到的普遍的自我毁灭和自我抹黑行为。

如果基本公理是错误的呢？如果有人能证明基本公理是错误的，那就必然会深刻地影响到我们的文化，我相信也会深刻地影响到亿万人的生活、深刻地改变我们对自己的看法。

在我职业生涯的后期，我做了一件对于康奈尔大学的教授来说似乎不可思议的事，我开始怀疑基本公理。我这样做的时候怀着极大的恐惧和不安。我知道自己是在反对现代学术界的至圣神主牌。其后果可能不堪设想，甚至可能导致自己被驱逐出学术界。我虽然在自己的专业领域（应用遗传学）已取得相当的成就和认可，但这样做将意味着冒险跨出我自己的那一小块安全属地。我将不得不开始探索一些重大的问题，包括我以前一直凭信心而接受的理论遗传学的某些领域。我觉得自己不得不去做这一切，但我必须承认，我已做好了可能碰壁的准备。然而我逐渐认识到，在基本公理周围构筑的看似强大而无懈可击的堡垒其实是座纸板房，这是我始料未及的。基本公理其实是一个极其脆弱的理论，事实上，它根本是不攻自破的。它表面的无敌大多来自于恫吓、烟雾、和镜影。基本公理赖以立足的很大一部分原因是“真正的信徒”们对于无所不能的自然选择的一种近乎神秘的信仰。此外，我开始看清，这种对于自然选择的根深蒂固的信仰，往往伴随着一定程度的对某种意识形态的认同，而这种认同只能被称为宗教。我开始意识到（再次伴随着惶恐），我可能会冒犯一大批人的宗教！

要质疑基本公理，我需要重新审视自以为了解的遗传学的几乎所有内容。这是我一生中最困难的心智上的努力。根深蒂固的思维模式改变起来是非常缓慢的（而且，我还要加上“痛苦”二字）。最终我彻底推翻了自己以前的理解。挣扎数年之后，我内心深处产生了一个新的、十分强烈的信念，即基本公理是绝对错误的。更重要的是，我开始相信，对于任何有理性、思想开放的人来说，基本公理可以被证明是错误的。认

识到这一点即让人振奋又让人害怕。我意识到，我有道德上的义务公开挑战这一至圣神主牌，但我也意识到这会让我遭受学术界大多数同僚的强烈蔑视，更不用说来自其他权威的激烈反对和愤怒了。

我该怎么办？我已确信基本公理的阴险是无人能及的，它对无数人的生活具有灾难性的影响。此外，我进行的每种客观分析都让我确信公理显然是错误的。所以现在，无论会有什么后果，我都要大声地说：**皇帝没有穿衣服！**

我敬请读者仔细思考这个重大问题，您真的是毫无意义的一大包分子吗？仅仅是随机分子突变和繁殖过滤的产物吗？当您读这本书时，我请您对一些非常有挑战性同时也让人兴奋的问题开放您的心智。我认为，如果您愿意稍费脑力，随我思索一些简单的论据，我就能说服您，基本公理是假的。您能想象比这更激进更能解放思想的事情吗？如果基本公理可以被证明是假的，这对您自己的生活乃至您周围的世界都会产生重大的影响。正因为如此，我才斗胆写了这本书，有些人将视之为大逆不道，而另一些人将视之为醍醐灌顶。

如果基本公理是错的，就必有一个出人意料但又非常现实的后果。仅在自然力量的作用下，人类基因组必会随时间而退化。认识到这一冷酷的现实，其意义将不会限于学术上，这也不仅是个历史问题。这应当促使我们每一个人反省：自己对未来的希望是基于什么。

补充：自从这本书最初写成以后，一系列重大的新进展陆续出现，所有这些新成果都有力地支持本书的主题。这些进展包括：基因组非线性性质的

发现、高等基因组中许多核苷酸的多功能性、基因组所包含的信息远超出我们想象的事实、以及关于人类基因组中大多为“垃圾”DNA和“沉默”DNA这一谬论的破产。在各章之后，我添加了简短的“补充”以说明这些新论据。另外我又加上了第十一章，对重大的新进展进行了总结。



生命之书从哪里来？

简讯：基因组是一套说明书

基因组是一个生物体全部基因元素的总和，包括所有的染色体、基因和核苷酸。它是一套说明书，决定着一种生命形式。人体细胞之所以成为人体细胞，人体之所以成为人体，都是由人类基因组规定的。基因组的复杂性是任何人工设计出来的信息系统所无以比拟的。

要想理解基因组的复杂性，首先必须弄清它包含着多少信息。不妨先打一个比喻，比如你给孩子买了一辆红色的四轮小拉车，有一本小册子告诉你如何把它组装起来。这小册子看起来不过几页纸，但它并不包含制造小拉车所需要的全部信息。这本小册子没有告诉你如何制造各个部件，也不包括生产钢材、橡胶和油漆所需要的知识。如果把所有的信息加进去，这本说明书将会很厚。如果你要编辑一套制造一部现代汽车的指导说明，这将会是一个小图书馆。如果把制造所有部件，以及制造自动组装生产线所需的信息也包括进去的话，这会是个相当大的图书馆。同样地，制造一架战斗机及其所有部件，加上制造它的电脑和组装线所需的信息也会是一个庞大的图书馆。要制造一架航天飞机及其部件和所有支持体系所需的信息更是

不计其数。然而要说明即使一个最简单的生命形式，其复杂性与一架航天飞机也不相上下。试想从一个细菌到一个人体，其复杂性的跳跃就如从一辆红色小拉车到一架航天飞机的跳跃一样。事实上，没有一种人为的技术可以与人体的复杂性相提并论。人类基因组就是包含人体生命所需的全部信息的说明书。

迄今为止科学家才发现了这“生命之书”的第一个维度，一个由4种被称为核苷酸的极其微小的分子所组成的线性序列。这些小分子构成了DNA螺旋梯形结构的每一个阶梯。这些分子是基因密码的字母，以A、T、C、和G表示。这些字母如直行文字串联在一起，它们就是我们的说明书的字母。这四个分子字母的组合形成了说明书的单词，这些单词结合形成了基因（说明书的篇章），基因联合形成了染色体（说明书的卷册），染色体统合形成了整个基因组（一个图书馆）。

一个完整的人类基因组由两套染色体组成，每套染色体含有30亿个字母。这其中只有很小的一部份被用来直接编码大约数十万种不同的人体蛋白质和细胞内无以计数的功能性核糖核酸分子（RNA）。每个蛋白质和RNA分子都是由上百个部件组成的精致复杂、设计巧妙、功能完善的微型机器。然而，基因组中的线性信息虽然相当于许多套大百科全书，还是不足以解释生命的复杂性。

尽管DNA序列的线性信息令人惊叹，但这只是基因组复杂性的第一个维度。基因组不是一串字母组成的简单的直线序列，而是包含着相互重叠的多层线性密码，是一个极其繁杂的“数据压缩”体系（见第9章）。

基因组信息除了多层相互重叠的线性语言形式之外，它还

像计算机程序一样，有无数环路和分支。一些基因控制着另一些基因，环环相扣。有些基因能感受环境的变化并启动复杂的串联事件，进而指导其它基因的反应以适应环境。还有些基因能够自行重排或改变其它基因的序列，也就是说，说明书自己能修改自己的部分内容！

最后，有证据显示线性DNA能够折叠成二维或三维结构（就像蛋白质和RNA分子一样），这种折叠可能还包含着更深层的信息。有理由相信，在典型的非分裂状态（静止期）的细胞核内，存在着精致复杂的三维DNA排列，其三维架构控制着高级生物功能。

总而言之，基因组的指导说明书不是简单、静止的线性字符串，而是动态的、自我调节的、多维的。没有任何人造信息系统可与之相提并论。基因组最高层的复杂性和互相作用也许超出我们的理解力，然而至少我们可以认识到这种更高级的信息是必定存在的。人体基因组中的线性信息已经令人惊叹不已了，而其非线性信息必然更加玄妙。考虑到生命无以复加的复杂性，这一点是无可争议的。

所有这些信息都包含在一个基因组中，而基因组又蕴藏在一个比灰尘还小的细胞核内。人体内的细胞数量好比一个星系（超过100万亿）。每个细胞都有同一套完整的指导说明书，却执行着高度特异的职责。人类基因组不仅决定了我们的细胞和躯体的复杂性，而且也指导我们大脑的功能。要形成人脑内的各种结构，其组织水平完全超出我们的想像。

当我们意识到基因组的高级维度，相信我们可以接受卡尔·萨干经常讲的：每个细胞所包含的信息比国会图书馆还多。

确实，人类的生命比所有人类技术的总和更复杂！**所有这些信息都是从哪里来的？它是如何维护的？这就是基因组的最大奥秘。**

关于生物信息起源的标准答案是：突变和自然选择产生了所有的生物信息。这就是新达尔文主义的理论基础。该理论认为世界上所有的基因组（说明书）都是从某些最简单的原始基因组经过一系列的突变（印刷错误）和多次自然选择（选择性复制）而来的。这就是生物进化的基本公理。**生命之所以成为生命是因为发生在分子水平上的随机突变经历了一个作用于整个生物体水平上的繁殖筛的不断过滤。**

什么是公理？公理就是一个不可测试而要凭信心接受的概念，因为它在一切有理性的人看来是不证自明的。在此基础上，它被认为是绝对真理。在本书中，我将敦促读者发问：我们应该接受今天的这个基本公理吗？如果基本公理被证明是错误的，那就意味着我们需要重新审视许多其他的流行理念，因为基本公理是奠定现代思想的根基。这将导致一个重大的模式转变（模式转变即统领群体思维的基本理念的改变），从而使人们对构架现代思维的许多根深蒂固的概念得以重新评估。

我们应该从一个诚恳和现实的角度去思考基本公理。我想提一个坦率的比喻来准确地描述今天我们这个基本公理的特点。我的比喻讲的是交通方式的演变。

在我们的红色小拉车的比喻里，最原始的基因组编码了组装第一辆小拉车的说明书。这个简单的基因组说明书由一个隐形的机械记录员抄写，目的是复制更多的说明书。每本新复制的手册被用来制造新的红色小拉车。然而这位记录员并不完

美，他会抄错。所以每一辆新出的小拉车都不同，而且各有一本独特的说明书粘在车底。当第一批小车废旧了，它们的说明书也被丢弃。新的说明书只能用还没有废弃的前一代小车的说明书不完美地复制。因为说明书的复制是接续性的（而不是从最初的原始版本而来），每一本手册都含有长期积累的误差，据此而制造的小车也开始变化，形态各异。当然，累积的误差就是突变的比喻。

也许你对这个比喻有些不安？无疑你意识到我们所说的的是一个衰退的情形。信息不断流失，说明书不断降级，小拉车的质量也无疑越来越差。最终整个系统将崩溃，手册完全无法读通，有用的小拉车也会绝迹。我们将在第二章和第三章详细考察突变在这一方面的问题。

现在要介绍我们的英雄：自然选择。自然选择像一位法官，或质量控制员，他来决定哪些小拉车是适合进一步复制的。作为法官的自然选择告诉记录员不要复制较差的小拉车的手册。这就是选择性繁殖（繁殖筛）。但值得一提的是他从来不直接选择好的说明书，而是只选择好车子。由此看来，问题就出现了。突变是复杂的，且发生在分子水平，但选择只能在整个生物体的层面上运作。记录员和法官各自独立工作。记录员几乎是瞎眼的，在分子水平上工作，因为高度近视，他只能看到他所抄写的一个个字母。这个法官也是几乎瞎眼的，因为高度远视，他从来看不见手册的字母，甚至看不见小拉车的各个部件，他看到的只是车子的整体相对性能。

记录员坐在自动组装流水线的开端。他盲目地、不精确地为自动生产线从旧程序复制出新程序，一个字一个字地抄。而

质量控制员检验装配完的拉车，根据整辆车的性能判断好车劣车。他选中的拉车的程序被交给记录员抄写，用作下一轮组装的蓝本。

用这种方法，许多有缺陷的拉车可以被排除了，说明书中的大多数错误也可以认为被排除了。更令人兴奋的是，有些罕见的拼写错误可能会造成更好的拉车，因而法官可以指示记录员选择性地复制这些更好的说明书。进化的过程开始了！

现在，让我们分析一下选择过程作为改善基因组信息之机制的可行性。说明书里的信息不仅要被这一过程所改善，而且也会被扩展。如果这位不完美的记录员偶尔多抄写了一页（重复），那就开始增加信息量了。当然重复的一页并不是什么新信息，它肯定会混淆和干扰手册的阅读。然而，法官只允许从好的拉车那里复制其手册。所以，坏的重复应该被排除了，而无害的重复可能被保留。这些无害的重复也将有复制的错误在里面，有些错误或许能够产生有用的新信息，比如为拉车制造新的功能性部件。只要稍作想像，不同的重复内容通过拼写错误可以渐渐进化，并编码全新的东西，比如一个内燃机、机翼、或电脑导航系统。这样我们就可以预测一辆红色小拉车能够通过一系列抄写错误，逐渐演变为一辆汽车，一架飞机，甚至一架航天飞机。

但这个比喻还不够，因为一个人的身体要比航天飞机复杂得多。事实上，我们的表型组（整个身体包括我们的大脑）之复杂程度无法估量，远超过任何人为的技术。如果我们想像小拉车转变为一艘带有超光速引擎和模拟环境舱的星际飞船（图1a-d，第10-12页），这个比喻或许会更贴切些。基本公理教导

我们，拼写错误和选择性复制能够同时演变出一套装满图书馆的制造说明书（基因组），和一个星际飞船（表型组），如图1d所示。

现在我们必须问：拼写错误和选择性复制真的有此魔力？要评估基本公理的可信性，首先必须准确地理解选择的含义。整个过程并没有智能的参与。这个记录员不过是一套复杂但毫无理智的分子机器，只会盲目地复制DNA。法官也不过代表着某些个体比其他个体更能繁殖的趋势。许多人下意识地把自然选择看作一种超自然的智能。但自然选择只是描述一个盲目和漫无目的的过程。要明白我们的记录员和法官既没有远见也没有智力。他们的智商总和为零。

生物进化的基本公理宣称在没有智能、目的、或设计的情况下，抄写错误和有限的选择性复制能够把一辆小拉车改变成为一艘宇宙飞船。这多么不可思议！你认为可信吗？更令人惊奇的是，在基本公理之下，这艘宇宙飞船根本没有经过预先设计，甚至上帝也没想过。整个过程纯系偶然发生。宇宙飞船只是一个突变的拉车。然而这是基本公理最好的比喻。唯一的不足是人为的技术难比生命的复杂性，而人造信息体系也难比人类的基因组。

这个比喻与道金斯(Dawkins, 1986)所描述的假像截然不同。著名的道金斯争辩，用一个像儿童游戏一样简单的过程，得出“我觉得那像一只黄鼠狼”这句话（“Methinks it is like a weasel”，出自莎士比亚《汉姆雷特》）。但这个比喻是用一种预先设定的语言把一个预先设定的信息系统地显示出来。在道金斯的模式里，要表达的信息和得到信息的方法都是经过智慧

设计、巧妙构思、仔细琢磨而且目标明确的。不仅如此，他的选择方案能够直接选择基因型（拼写错误）而不是表现型（拉车性能）。简而言之，道金斯编写了一个简单的计算机程序，从一串随机组合的字母开始，字母数量与“我觉得那像一只黄鼠狼”这句话的字母数量相同。他的程序随机“突变”（更改）字母。当某个位置上出现一个新的字母与“我觉得那像一只黄鼠狼”里的字母吻合时，程序便会选择出这个“改进”了的信息。显然，不用多久这个程序就能制造出所要的这个句子。然而，要使用这个简单的程序，道金斯还得小心设计“复制速度”，“突变率”，以及其他的参数以便达到预期的目的。此外，他还需要假设每一代在每一个字母上都具备完美的选择效率。这个程序是为了证明经突变/选择而进化的必然性（不需要任何智慧设计）。道金斯显然使用了一个智慧设计的计算机，又以他自己的智慧来编程，来优化程序，甚至来设计一个预先选定的句子。基于这些原因，道金斯的说法不能实实在在地用来为基本公理辩护。（因为基本公理不允许使用任何智能、目的、或先见，而且也不允许直接选择拼写错误）。

在本书中，我们将考察遗传学的一些基本层面，来判断关于人类基因组的已知事实是否与基本公理相符。读下去，如果你感觉到基本公理不再是一个显而易见的真理，那你就应该义不容辞地否定它的公理性。如果基本公理不再成其为公理，它就应该被视为尚未证明的假说，是可以证伪的。

2014年补充：2013年有一本划时代的书出版，书名为《生物信息——新视角》（见BINP.org）。这是在康奈尔大学举办的一个学术会议的论文汇编，主编为马克斯等。作者包括29位享有盛誉的科学家，代表了众多的科学

领域。他们对于生物信息的本质有大致相同的见解：**生物信息系统在本质上更像一个严密的计算机系统，而不是一本书。**DNA就像细胞的硬盘。数以百万计的RNA和蛋白质分子之间相互作用，就像细胞的活动内存或RAM。每一个基因就像一个可执行的计算机程序（事实上每个基因包括多个程序）。细胞内的每一个蛋白质分子和RNA分子就是一个简单的程序（算法）。DNA、RNA、蛋白质和无数其他分子不断地相互交流，在每个细胞内形成一个巨大的互联网。有人用信息显像技术显示，高级基因组在结构上与可执行的计算机程序惊人地相似。

有关基因熵的问题的最新发现，可见geneticentropy.org网站。

图1a 需要组装……



一个红色小拉车不是信息，但它需要信息指导其组装。一本典型的组装小册子并不包括生产拉车所需的全部信息。真正完整的生产手册应该是一本非常庞大的书，详细说明生产所有部件（轮子、拉手等）和所有原材料（钢材、油漆、橡胶）的过程。

图1b 一个图书馆



详细说明制造一部现代汽车的完整的说明书可以组成一个很大的图书馆。如果组装完全由机器完成（无需“智能”），机器人的制造和编程所需的信息更是相当可观的，可以汇编成数量繁多的书。

图1c 多层信息

详细说明制造一架喷气战斗机的完整的指导手册，包括机上计算机系统，以及所有制造、支持、和维护系统，将是一个庞大的图书馆。想像一下如果每个部件都用机器人来完成，这套指导手册的内容该有多大！

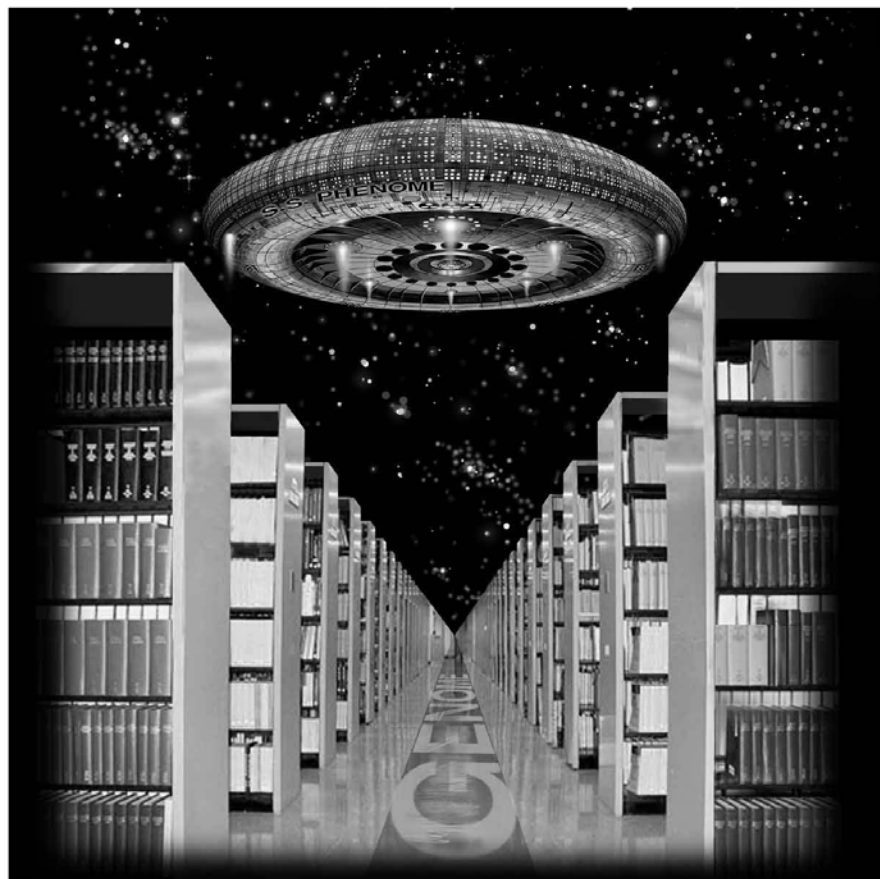


图1d 一个星系的信息……

上图中的图书馆代表人类基因组（我们所有的遗传信息）。宇宙飞船代表人类表型组（我们的整个身体，包括我们的大脑）。我们想像不出到底这个图书馆的容量要有多大才能解释这个带有超光速引擎和模拟环境舱的科幻星际飞船的构造细节。会不会比国会图书馆大得多？然而，可以说人体还是比一个假想的星际飞船更复杂。哪种星际飞船会自我繁殖？

atcg

第二章

随机突变是好事吗？

简讯：随机突变不断地破坏信息

突变是生命之书抄写错误。探讨人类基因组突变这个话题的时候应该带着恻隐之心，因为这关乎人的生命，而生命正受到突变的伤害。先天性缺陷所殃及的家庭很多，所以这不仅仅只是个“统计数据”。广义上的遗传性疾病是灾难性的。如果考虑到所有的有遗传倾向的疾病，结论必然是：我们都是高度突变体。此外，几乎每个家庭都受到癌症的困扰，而癌症本质上是体内细胞突变的恶果。不仅如此，越来越多的证据还显示衰老本身就是我们的细胞内不断累积的突变带来的。突变是让人痛心疾首的祸源。事实上，正是突变在无情地杀死我们每一个人。所以突变决不仅仅只是个学理话题！

我们能说突变有益吗？几乎所有的保健政策都旨在减少突变。大多数个人卫生措施也是为了减少突变，以降低癌症和其他退行性疾病的风险。怎么能说突变是好的呢？然而根据基本公理，突变是有益的，因为能产生变异和多样性，促成选择和进化，进而产生生命所需的信息。

在做进一步阐述之前，我们首先要认识到有两种变异：随

机变异和设计变异。随机变异就像我的汽车随着岁月而发生的变化，即是那些锈迹、撞痕、刮伤、和失灵的部件。这些都使汽车与汽车之间出现变异，但能否因之产生好车呢？打印错误真能改善一个学生的期末论文吗？向玻璃房子丢石头能让它增值吗？除了这类的偶然事件（随机变异）之外，还有一种变异称为设计变异。我买车的时候有很多选项：漆色、轮胎的类型、引擎的规格等等。这些对我选车有帮助。部件的多样设计即使在买了车以后也还有用，比如添加或减去一些功能，更换损坏的零件等等。设计的多样化是有益的，不仅方便汽车的维修，甚至在一定限度之内能改进汽车。然而，即便这类智慧设计的变异也永远不会将我的车子改造成一艘宇宙飞船。

根据基本公理，基因变异是没有经过设计的，所有的变异必须来自随机突变。然而，基因工程时代已经开始了，这个公理性的假设已明显地不再正确（因为现存许多物种内已经引入了人为设计的基因变异。）这一事实也许能够打开我们的眼界，让我们想到被设计的基因变异在人类以前就存在的可能性。若不是执著于基本公理的教条，我们应该有充足的理由认为随机突变从来不是有益的。就汽车而言，随机变异更像岁月带来的撞痕和擦伤，而不是备用零部件。

突变能产生新信息的例子少之又少，让我们看到突变的本质是极其有害的。要知道科学家们有一个极其敏感而广阔的网络，他们不会放过任何一个能产生新信息的突变，多数遗传学家一直积极关注此事。这类研究已经进行了大约100了。科学家的观察网络是如此敏感，即使一百万个突变中有一个确实产生了新信息（微调不算），文献中就会有铺天盖地的报道。但我还从未见到

一个令人信服的例证，能说明突变无疑地创造了新信息。当然有许多突变被认为是有益的，但这些有益突变多数并未创造信息，而是破坏信息。比如说，有些人（像我）会认为汽车报警器失灵是“有益”的。然而这些有时看来是“理想”的随机改变并没有创造新的功能，反而代表了信息的流失。举个真实的例子，当细菌因为染色体突变而对抗生素产生耐药性时，总是伴随着细胞功能的丧失。耐药菌并没有进化，事实上它的基因出现缺陷，偏离正常。只要把抗生素拿掉，突变菌株很快就会被正常菌取代，因为正常菌比耐药菌有生长优势。还有一个就是赤花花狗的例子。形体变小和毛发脱落也许有助于适应酷热的环境，但这也是退化，并没有产生新的信息。在这些情况下，信息是流失而不是增加。然而基本公理仍然认定突变是有益的，是进化形成基因组内数不清的信息的必要步骤。下面我们进一步讨论这个概念。

基因突变引起的退化在前述说明书的比喻中显而易见。假如一架喷气式飞机的组装说明书中有这样一段话：

“步骤6. 完成上一个步骤后，重复步骤3，直到B部件厚度达到10.004 mm。在进行下一个步骤前，至少等候3小时。”

如果我们只考虑简单的点突变（拼写错误），说明书可能受到三种不同程度的影响。首先，理论上，有些拼写错误可能对文本信息没有影响（在上面这句话中我看不到这样的例子）。其次，大多数拼写错误将对文本信息的清晰度和连贯性有非常轻微的影响（所有没有划线的词语中发生的拼写错误）。最后，少数改变（在划线部分）将产生重大（几乎致命的）的影响。在以上例子当中，我们看不到拼写错误可能改善说明书并得到一架更好的飞机的情况。虽然这样的改变是可以想像的，但不大可能发

生，其原因难以完整地描述。任何因拼写错误可能带来的好处都是微乎其微的（例如，厚度小数点后第三位的4改成8）。这种改变并没有增加信息，只是调整，或微调了系统。任何有理性的人都应该明白，我们不能指望任何拼写错误会给飞机制造技术带来重大改进。比如，拼写错误不会产生一个新的可以申请专利的部件。这样重大的改进显然需要智慧设计。即使有假想的拼写错误能轻微地改进（更准确地说，调整）一架飞机的蓝本，它必将被掩埋在大量有害的拼写错误之中。这些有害的改变可以是几个致命的错误，也可以是数量众多的几近中性的改变。

上述比喻可延伸到基因组（图2，第25页）。在人类基因组内有超过30亿个潜在的突变位点。其中只有一小部分在突变时产生重大影响。虽然没有人能证实大多数的潜在突变都是零作用的，但这些核苷酸的位点应该是“接近中性”的（以下会更详细地讨论）。生命说明书中的拼写错误有时是极其有害的，但在绝大多数情况下，其危害只是轻微的。不能期望突变产生新的信息，虽然现有信息会在一定范围内进行调整。生物调整好像动用细胞的“变阻器”。比如，突变会调整一个基因启动子或酶的活性，有时上调，有时下调。可是当我们用一个变阻器使灯光暗下来时，我们并没有创造出新电路或新信息。我们只是对现存系统进行了微调，事实上这种**可调性是预先设计的**。

我刚说过绝大多数突变应该是接近中性的。这一点所有的群体遗传学家都会同意，其原因是多方面的。首先，这是任何书面语言拼写错误的通性（试想一下改变这本书的任何一个

字母）。其次，基因组中核苷酸的总数是巨大的。平均来说，每个核苷酸位点所包含的信息是总数的30亿分之一。第三，大量关于特定编码序列、基因启动子、增强子突变的研究均得出类似结论。实验表明，大多数核苷酸位点对于细胞功能只有非常轻微的作用，只有少数突变能真正称得上是基因功能的“杀手”（切记任何一个基因的功能只是整个细胞体系极微小的一部分）。最后，在全基因组范围内有些特征性模式（密码子选择、核小体接合区、GC等值区、物种间不同的“词汇构成”等等）。这些模式涉及到散布在基因组中的亿万核苷酸。单个核苷酸位点对维持大范围重复性模式所起的作用微乎其微。这种作用虽然接近无限小，但还不等于零。有了每一个核苷酸的参与，这些模式才得以存在。每一个核苷酸仍然携带信息并发挥影响。但不论我们如何分析，我们都会看到大多数核苷酸位点一定是接近中性的。

有没有真正中性的核苷酸位点？真正的中性不可能通过实验来证明（那将要求无限的敏感度）。然而，由于我们接下来将要讨论的一些原因，一些遗传学家曾急于压缩功能性基因组，把基因组的一大部分归类为“垃圾DNA”。“垃圾DNA”内的突变被看作是**完全中性的**。然而实际研究结果却不断扩展着基因组中的功能性DNA，而假定中的“垃圾DNA”不断地减少。早先曾经有许多遗传学家相信功能性DNA不到全基因组的3%，短短的几年之内这个比例就增加到30%以上，而最近的证据提示基因组中80%以上的DNA是有功能的（2012年ENCODE项目报告）。随着功能性基因组的扩大，中性突变的可能性相应地减小。除此之外，还有重要的理论考量使我们相信不存在完全中性的核

核苷酸位点。核苷酸位点在基因组中占据空间，影响其他位点的间距，也影响区域性核苷酸组成、DNA折叠和核小体的接合。如果一个核苷酸绝对不携带信息，根据定义它就必然轻微有害，因为它会延缓细胞的复制过程而且消耗能量。正如一部大百科全书中没有真正中性的字母，基因组中也没有真正中性的核苷酸位点。所以不可能改变一个位点而不带来生物效应，不论这效应多么轻微。虽然大多数位点可能是近乎中性的，很少或者没有一个位点是绝对中性的。艾尔沃克 (Eyre-Walker) 和凯特里 (Keightley) 早在2007年就清晰地驳斥了关于绝对中性突变的谬误。他们写道：

“要说任何一个突变是真正中性的，对适合度毫无影响，这是不可能的。凡是突变都会有某种效果，不管其效果是多么微乎其微。然而确实有一类突变，我们可称之为实际中性……在这里，中性的定义是在运作上，而非在功能上。这里讲的是自然选择对某个突变是否有效。选择的效力不仅在于突变对个体适合度的影响，还在于它在群体水平上的效应，也要考虑到该基因在分离过程中的基因组背景。”

那么基因突变的效应分布究竟是什么样子的呢？图3a (第26页) 显示一个不切实际的钟形突变曲线，一半突变有利，另一半有害。新达尔文主义理论的创始人之一，费歇尔 (R. A. Fisher)，曾认为这就是突变效应的实际分布，他还在这错误的假设之上建立了他的“关于自然选择的费歇尔基本定理”。既然前提是错误的，他的一整套数学分析也是无稽之谈。若是根据这样一个突变曲线，选择和进化就不难想像。任何水平的选择，不管多么微弱，也会保留下一些有利突变，淘汰掉一些不利突变。如果这个分布是正确的，上升性进化就不可避免。这种观点显然是错误的。事实上，有利的突变是如此之稀少，

在这种图上根本显示不出来。图3b (第27页) 则更合乎实际，突变分布于致死性 (-1) 和中性 (0) 之间。然而这还不完全正确。曲线还要大幅度地偏向中性值 (零效应点)。如前所述，大多数突变是接近中性的。图3c (第28页) 是根据木村资生 (Kimura, 1979) 的描述修正扩展而来的。该曲线很近似地代表了突变的实际分布情况。

从木村的曲线可以看出，多数突变呈负效应，并且紧靠零值。换句话说，绝大多数突变是有害的，而且压倒性地接近中性。木村因提出“近中性区域” (这里显示为一个方格) 而闻名。木村称接近中性的突变是“实际中性”，意思是它们的作用很轻微，不受选择的作用。然而，我们可以看到木村并没有显示任何突变是绝对中性的。他的曲线趋近、但没有达到零效应点。在木村的计算中，“不可选择区”的界限 (图中方格的大小) 是 N_e (一个繁衍群体中有生殖功能的个体数量) 的函数。

值得注意的是木村根据群体数量计算出来的方格大小仅为实际中性突变之分布范围的最小估计。任何影响生殖几率的非遗传性因素都会扩大方格的实际面积。在第六章我们将要看到，非遗传噪音极大地增加了方格的面积 (图9, 第94页)。任何会降低“信噪比”的因素都将按比例增加基因组中不可选择的核苷酸数量。著名遗传学家穆勒早就意识到了非遗传性因素对自然选择的重大影响 (Muller, 1964)。

在木村的图中，零点的右边没有任何突变，即没有有利的突变。他显然认为有利突变太罕见，无需考虑。看了这个突变分布图，你自然会问：“理论家们是如何解释进化过程的呢？”回答是：所有在“近中性方格”之内的突变都被重新定

义为完全中性，所以不予考虑。然后假定在近中性方格之左的突变都可以用自然选择来彻底淘汰。以这两种思维方式排除了全部有害突变之后，理论家们可以自由地辩称无论（在方格右边的）有利突变数量是多么稀少，都应该有足够的时间和足够的选择力来挑选有利突变，并使其成为进化的基础材料。我们很快就会发现，他们的说法无论如何都是错误的。在方格之内的突变不能被忽略，在方格之左的突变不一定被选择过程完全淘汰，也没有足够的时间或选择力来筛选极其稀少的可能出现在方格之右的有利突变。

既然有利突变在所有的进化理论中扮演着如此重要的角色，我怀疑木村为什么没有将之显示在图中。为公平起见，我觉得应该把它画出来。对实际发生的有利突变来说，这个曲线应该是有害突变曲线的反转。正如有害突变一样，绝大多数有利突变应该是近乎中性的，密集于中点的右侧。克劳(Crow, 1997)清楚地指出了这一点。绝大多数有利突变的作用是微乎其微的（在附录5中有详细论述）。然而，有利突变的数目比有害突变少得多，其分布范围和曲线下的面积也要相应地小许多。我看到一些估计，有害突变和有利突变的比例在一千比一到一百万比一。最好的估计应该是一百万比一（Gerrish and Lenski, 1998）。有利突变的发生率极低，低到难以实际测量（Bataillon, 2000; Elena et al., 1998）。所以，我不可能在零点右边画出一条足够小的曲线来准确地表示有利突变是多么稀少。我只能把它极大地夸张，使人能够看出有利突变可能的分布（图3d，第29页）。图3d忠实地代表了突变的自然分布，只是有利突变过于明显。最有意思的是（我自己也受震动），

该图让我们认识到，假设中的有利突变分布范围完全坐落在木村的“实际中性”区域内。这就是说所有的现实发生的有利突变是“不可选择的”。所以选择过程不可能筛选出这类有利突变，因而这些突变会在群体中一一漂移流失。我们以后会讨论，也有罕见的有利突变效应很强，但那是特殊的例外。图3d中“非选择区”的右侧有两个箭头，用以代表这种情况。

图3d生动地说明了为什么突变不可能导致信息的净增加。我们在接下来几章中将会更清楚地看到，选择过程不能触及在近中性区内的任何突变。所以，在方格中占绝对多数的有害突变必将造成信息的净流失。再者，当突变是隐性的时候，或者当突变率高而繁殖率一般或低下时，选择过程甚至不能完全排除在方格之左的有害突变。最后，我们将会看到，制约选择的众多因素，还限制着我们挑选可能出现在近中性方格之右的极少数有利突变的能力。**突变之真实分布的每一个特点都否认突变会带来信息的净增加，否认正向进化的可能性。**

因为有利突变是基本公理的活力核心，我需要在这方面多谈一点。在上个世纪，人们耗费大量人力物力，试图利用突变来产生有用的变异。在我的领域（植物育种）里面尤为突出。当人们发现有些放射线和化学物质是强有力的诱变剂时，数以百万计的植物被诱发突变，以筛选可能被改良的品种。根据基本公理，这显然要引起农作物的快速“进化”。有几十年，这曾经是作物改良研究的主流方法。通过亿万万个突变实验，科学家们制造并遴选了无数的突变株，“发明”了许多矮小、不育、病态、畸形、异常的植物。然而，这一切努力的结果是，几乎没有制造出什么有意义的改良作物。所有的努力基本上可

说是一个大失败，这种做法后来几乎完全放弃了。为什么经过众多博士科学家的煞费苦心，这个突变/选择的大实验还是失败了呢？原因是，即使在那亿万个突变中也没有一个有意义的有利突变。几个例外的突变也证明了这一点。低植酸玉米是最著名的突变育种成功的例子。这种低植酸玉米作为动物饲料有些长处。低植酸玉米是通过诱导玉米变异，然后挑选植酸合成基因被破坏的突变株而来的。虽然得到的突变株有利于农业上的某种特别需求，但这是经过信息的净损失（如失灵的汽车报警器）和生物功能的丧失来实现的。其他的突变育种成功的例子大多是在观赏植物的范围内，功能不良的畸形物被认为是新奇和悦人眼目的。“有用”的观赏植物突变包括不育、矮小、斑点或杂色的叶子、或变形的花朵。

如果在大型科学技术指导下的过程都不能获得真正的正向突变（信息的净增），那么我们为什么会认为一个没有智能指导的、与此相当的过程会在自然界结出硕果呢？然而，正是那些做突变/选择试验失败的科学家，当他们放弃了诱变育种而改为利用植物种属内预先存在的自然变异后，取得了在农作物改良方面的极大成功。这就提示我们，这些自然界预先存在的变异主要不是通过突变而发生的，而是当初经设计而形成的。

伯格曼 (Bergman, 2004) 综述了有利突变这个题目。在这篇论文中他通过生物学文摘 (Biological Abstracts) 和医学在线 (Medline) 做了一个简单的文献搜索，发现453732条“突变”索引，但其中只有186条（万分之四）提到“有利”的字眼。当他对这186条索引进行审查时，发现所谓的有利突变只不过在很狭隘的意义上有利而且无例外地源于功能丧失（信息损失）。

他没有找到一个确定产生了新信息的突变。尽管几乎整个科学界都认同产生新信息的有利突变必然存在，但这一观念似乎是基于对基本公理不加辨别的接受，而不是基于事实证据。我不怀疑有些突变是有利的，并形成新信息，但它们的发生率显然极低，不足以创造基因组。

结论：突变是压倒性地有害的，即使在某些特别意义上被归类为有利的例子，也通常伴随着总体的损坏和信息的腐蚀。下面我们将会更详细地考察，即使伴随着选择，突变也一般地不会产生新信息。通过突变而来的变异更像是生命的撞痕和擦伤，而不是生命的备用零件（备用零件是设计出来的）。突变是个体衰老的根本原因，正在导致你我的死亡。除非选择过程能够阻止人类基因组内的信息腐蚀，突变不但要导致我们个人的死亡，而且将导致整个人类的灭绝。我们将很快明白，自然选择必须有效地淘汰大量的近中性核苷酸突变才能防止基因组的退化。

2014年补充：本书的一个目的就是要表明，正向进化需要高频率的有利突变，而且这些突变的效应必须强大得足以被选择出来。这类有利突变会很常见吗？简单的逻辑就能排除这一点，数十年来大量的科学观察也得出了同样的结论。逻辑告诉我们，遗传信息作为一种特别指令，是细致而具体的，随机突变会破坏其特异性，从而系统地损害有用的信息。然而偶尔有人宣称在实验研究中发现惊人的有利突变率。我和同事们在最近发表的一篇科学论文中有力地驳斥了这些说法（见Montanez et al., 2013）。要理解这些所谓的高频率、高强度的有利突变，关键是要明白，任何微生物的基因组中都有一大部分（可高达50%）基因是备用基因，而这些备用基因需要精确调控。备用基因在野生环境中很重要，能使生物体耐受一系列特别的生存挑战。但是在人工的、恒定的实验室环境中，这些基因不过是累赘。

破坏或删除这些备用基因常会节省能源，并能在人工环境中加快生长速度。类似地，破坏这些备用基因的正常调节机制也可能在人工环境中有利。在人工环境中，通过信息的删除、功能的丧失和基因的退化，会形成各种各样的“有利”突变（就像一辆赛车在比赛前拆除不必要的设施一样）。把这些突变为“有利”是不确切的，其实应该称之为“适应性退化”。这在林斯基等所做的著名的大肠杆菌长期培养实验中可以清楚地看出（Barrick et al., 2009; Lenski, 2011）。培养后的大肠杆菌确实适应了某种人工环境，而且在特定的人工培养基中其生长速度加快了。然而当人们分析相关的“有利”突变时，发现都是由于功能缺失（不是基因破坏就是启动子破坏）。基因组无一例外地缩小了。基因组显然不是这样建立起来的。相反，这样的发现直接指向基因组的退化（我们正在撰写一篇专门讨论这个问题的论文）。

最近，巴太郎和贝利(Bataillon and Bailey, 2014)也考察了突变效应的分布。他们特别研究了有利突变的分布，得出结论说：“直接实验证据证实了之前对有利突变适合度效应的预测，即突变效应大致呈指数分布，且右侧有边界。”换句话说，他们的分析与我的看法一致，并支持图3d。他们接着写道：“估算结果一致提示形状参数在韦伯域内”。我和同事们在数值模拟中通常采用的就是韦伯类分布。最后，他们讨论了有利突变的稀少：“常规的假设是……群体一般地都很接近最佳适合度，因而有利突变极罕见，可以忽略不计。所以，一般的模型中所使用的假设根本不允许对有利的适合度效应进行评估……然而，若是在进化实验中寻找有利突变，它们也不是少得找不到；这要看特定的选择环境，有时候有利突变还挺多……可以说，这类资料代表完全没有适应环境的群体，只在实验室内的人工环境中有意义。”这两位作者显然明白有利突变是极罕见的例外。总体上讲，基因突变系统地破坏生物信息。

```
atcgtacgtacgaggctatgcatgcaatgcatgctgctatatcgcatcgatatcggagatct
caccgtacgattccgagagttaccaatcgataggctatatccgccttaggcgctacac
atatttcgtacgaggctatgcatgcaatgcaatgctatatcgcatcgatatcgggac
gggacgatccacactcggagagttaatacgataggctataccggccttaaagcctaca
atatatttcgtacgtagaaaggctatgcatgcaatgcatgcttatatcgcatcgtaat
tcgggaattgcccataatacgataggctataccgccttaagcgtaactatcattcaactt
atcgacgtagcgaagctatgcatgcaatgcatgcttatcgatactatcgggagcta
cgtacgtatcggagagttaatacgataggctatcctccgcttaagcgggctaacatat
attgtacgtacggccccctaatgcatgcaatgcatgcttatcgatactatcgatacga
atcgtagcgtacgaggctatgcatgcaatgcatgcttatatcgcatcgatatcggagatct
caccgtacgattccgagagttaccaatcgataggctatatccgccttaggcgctacac
atatttcgtacgaggctatgcatgcaatgcaatgctatatcgcatcgatatcgggatt
gggacgatccacactcggagagttaatacgataggctataccggccttaaagcctaca
atatatttcgtacgtagaaaggctatgcatgcaatgcatgcttatatcgcatcgtaat
tcgggaattgcccataatacgataggctataccgccttaagcgtaactatcattcaactt
atcgacgtagcgaagctatgcatgcaatgcatgcttatcgatactatcgggagcta
```

图2 遗传信息的特点

基因组呈现为A、T、C、G字母的线性序列。实际的基因组比以上序列大3百万倍。假如你要看自己基因组的一半，如果你每秒看十个核苷酸，每星期看40小时，那就需要连看40年才能看完！这个语言系统并非像图上这么简单。一个高级的基因组必然包含大量的数据压缩（参看第九章）和大量非线性信息。除了某些简短的序列，我们不能把基因组简单地看作一个线性文本，如一本书。大量的信息内容可能存在于三维结构，如蛋白质的折叠一样。

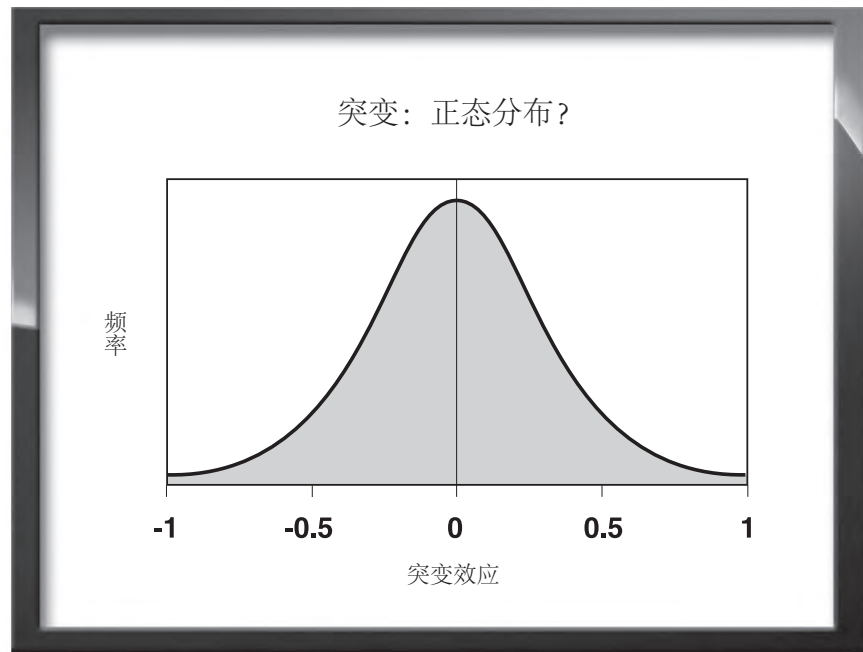


图3a 突变适合度效应的分布——不现实的看法

对突变效应的天真看法是一个钟形分布图，一半突变对适合度有害（中线左侧），另一半对适合度有正面影响（中线右侧）。照这个分布，不难想像自然选择过程会排除一些有害突变而保留部分有利突变，结果无疑地就是向前进化。然而，我们知道这个图像是不对的。

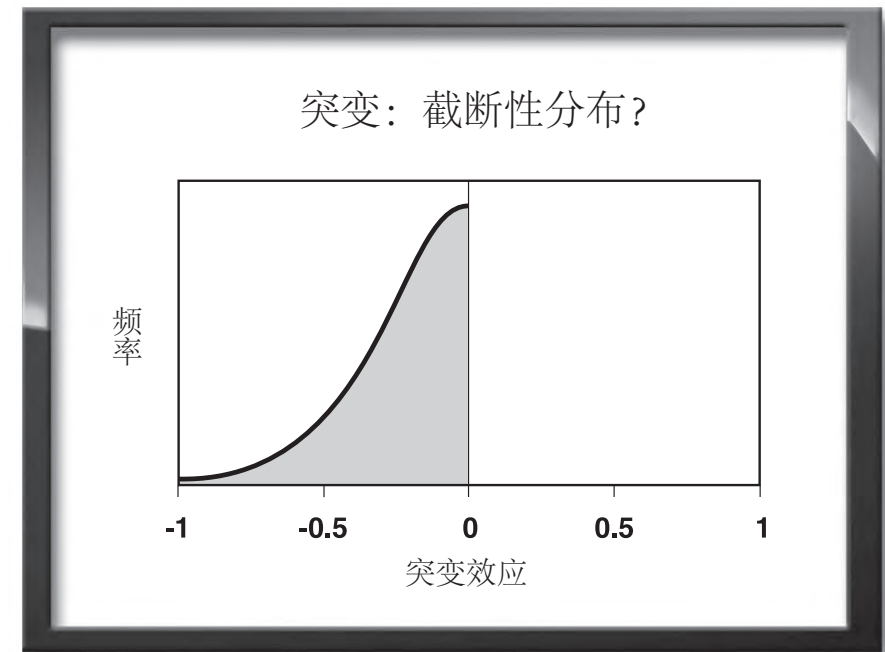


图3b 突变对适合度的效应——几乎没有有利的

群体遗传学家明白几乎所有突变都是有害的，而对适合度有正面作用的突变是如此罕见，在这种分布图中一般地不显示。这给进化理论造成很大的问题。然而该图还是过于乐观。

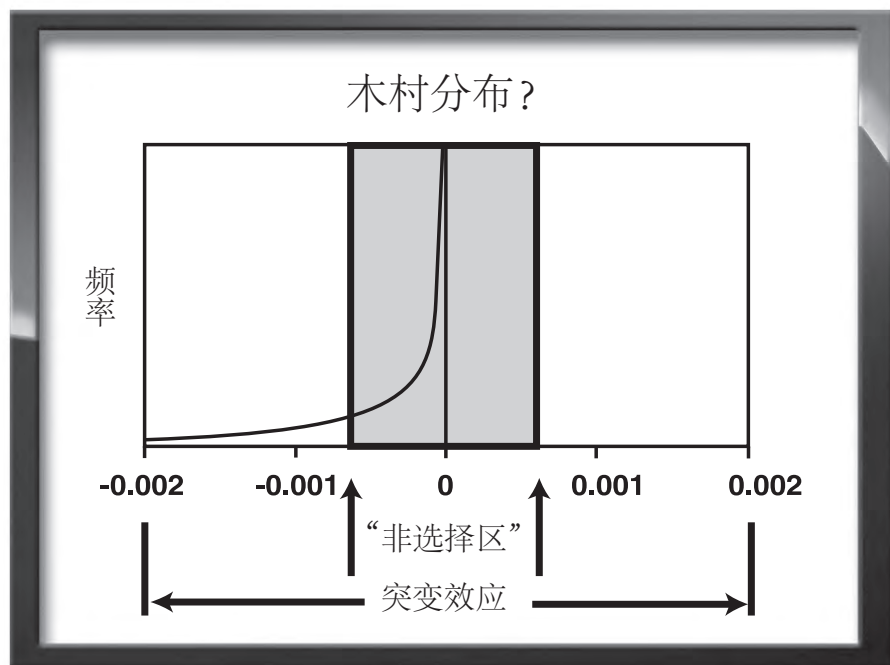


图3c 突变的效应——有害效应通常很微弱——不受自然选择的作用
群体遗传学家明白突变的分布高度偏向中性。正如一本说明书，少数拼写错误是致命的，但大多数拼写错误无关紧要。接近中性的突变对于进化论来说是最大的难题。这是从木村资生 (Kimura, 1979) 所画的图修改而来。注意该图的比例与前两个图不同，不是从-1到+1，而是从-0.002到+0.002。木村和他的同事太田以指出大多数突变是近乎中性的，因而不为选择所作用而闻名。木村的“不可选择区域”就是图中的灰色方格。

该曲线的大体形状是重要的，但其具体的数学内涵无关紧要。太田觉得突变分布是一个指数曲线，但木村认为它是个“伽马”分布 (Kimura, 1979)。然而，不管哪一个数学公式能最准确地描述突变效应的自然分布，它们都与上图相似。

遗传学家们认为非常有害的“主要”突变的频率几乎为零（不在图中），“次要”突变频率中等。次要突变是主要突变的10-50倍 (Crow, 1997)，但是近乎中性的突变在数量上大大超过这两种突变。

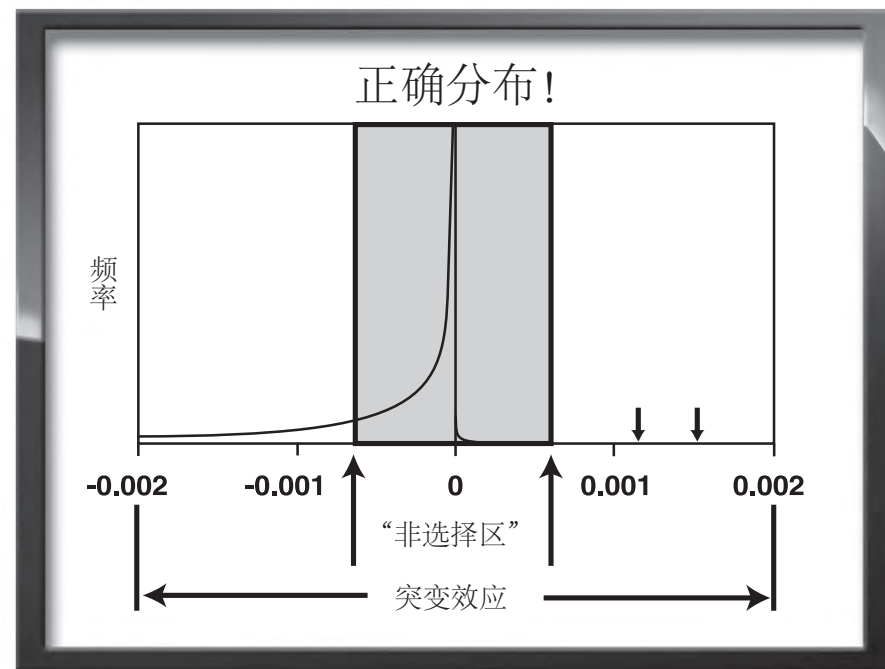


图3d 突变效应——罕见的有利突变虽然会发生，却通常不会被选择

木村的图 (3c) 还不完整。要使图解完整，我们必须显示出有利突变发生的区域，因为这对进化论至关重要。有利突变的分布应该是木村曲线的反转，但在区间和高度上要小一千倍到一百万倍。限于全图的尺寸，我不能把这部分突变分布画得足够小，所以只能画一条相对放大的曲线。尽管有利突变被高度夸大了，但显然几乎所有的有利突变都存在于木村的“非选择区域”之内。这个完整而正确的图说明了上升性进化在基因组水平上是不可能的。在特定环境中适应性的改变仍然会发生，这是由于极稀少的高强度有利突变——这属于孤立的例外（显示为“非选择区”外的箭头）。这些罕见的有利突变几乎全部是由于功能的丧失，所以无助于“向前进化”。

atecg

第三章

多少突变才算太多？

简讯：人类基因突变率太高了。

几十年来遗传学家一直担心突变对人类造成的影响 (Muller, 1950; Crow, 1997)。最早出现这类担忧时，人们对有害突变发生率的估计为每代每人0.12-0.30个 (Morton, Crow, and Muller, 1956)。从此科学界一直密切关注人类基因突变的积累，恐怕导致高“遗传负荷”和群体的普遍衰退。有很长时间人们一度认为如果有害突变率达到每代每人一个时，远期基因退化将成为必然 (Muller, 1950)。这种想法符合逻辑，因为选择过程淘汰突变的速度必须跟上突变的发生率。我们需要阻止突变个体的繁殖，但也必须留下足够的人口来繁衍下一代。这样一来，若要淘汰全部突变而仍然维持种群繁衍，人类每个孩子中必须少于一个突变。这是因为现在全球生育率为两个成人养不到3个孩子，所以理论上只允许三分之一的孩子被选择性地淘汰。基于上述原因，遗传学家们自然迫不及待地要找出人类的突变率到底是多少！

近年来遗传学界最惊人的发现之一是人类突变率（仅限于生殖细胞中）大约是每代每人75-175个核苷酸置换（拼写错

误) (Nachman and Crowell, 2000; Kondrashov, 2002; Xue et al., 2009; Lynch, 2010; Campbell and Eichler, 2013)。高突变率在遗传学界已经成为广泛共识。不仅如此, 上述作者之一的康卓寿(Kondrashov)博士, 曾向我表示100只是他的估计下限。他个人认为实际的点突变率(拼写错误)可能高达每人300个。即使文献报导中最低的数字, 每人75个, 也已经高得惊人, 而且含义深远。早先当聂耳的研究显示人类突变率可能高达每人30个时, 这位久负盛名的作者已经认识到, 这样的突变率对于进化理论的意义非同小可(Neel et al., 1986)。即便我们假定基因组中的一大部分是完全中性的垃圾, 这些数字仍然意味着每代每人带来多个新的有害突变。所以我们每个人都是突变体, 而且是多重突变体! 什么样的选择方式可以阻止信息这样流失呢? 下面我们将会看到, 没有任何现实的方法可以阻止基因组的退化。随着基因组内被认为真正有功能的部分在不断增加, 被认为实际有害的突变的数目也会迅速增加。如果基因组的大部分被证实是有功能的, 那么每个人这75-175个突变中也就有一大部分是有害的。由于对突变率的估计有很大的误差范围, 在本书以后的章节里我会使用每人100个作为人类突变率的保守估计。然而这个数字的确太低, 因为它只考虑点突变(字母置换), 还未包括其他多种类型的常见突变, 如串联重复的改变、缺失、插入、重复、易位、倒置、转换及线粒体突变。

为了理解上述数字确实低估了突变问题, 首先有必要介绍一下点突变以外的突变类型。然后我们再考虑整个基因组中哪

些部分确实有功能。

细胞内有一种称为线粒体的亚细胞结构, 线粒体中有一个属于自己的小基因组(大约16500个核苷酸), 这个小基因组只能通过母体遗传。这个DNA分子虽然不大, 但是因为线粒体基因组是高度的多倍体(每个细胞有好几百个拷贝), 又因为线粒体基因突变率极高, 每一代人中还带有大量的线粒体突变。这些需要统统排除, 才能阻止人类的退化。人类线粒体突变率估计为每个核苷酸位置上每一百万年发生2.5次突变(Parsons et al., 1997)。假设一代为25年, 线粒体基因组有16500个核苷酸, 那么每代每人生殖细胞中就有1.0个线粒体突变。仅线粒体突变本身就可能超过了选择淘汰的理论限度。即使突变率只有每人0.1个, 仅仅为了阻止线粒体基因的退化, 每一代都需要淘汰掉相当一部分(10%)的人口。这样的话, 每代每人还有100个或更多的细胞核突变在无约束地积累。线粒体的高突变率对于选择来说是个大问题(第四章和第五章)。这是因为线粒体DNA缺乏重组(“穆勒的棘轮”——Muller, 1964), 而且有效群体小(这种DNA只通过母体遗传, 所以选择过程只作用于群体的半数)。

人类基因组中突变最活跃的区域存在于微卫星DNA带。这些独特区带的突变率是其他序列的一百万倍, 一般的突变率估计都不包括这里的突变。然而这些序列已发现是有生物作用的, 其突变导致许多严重的遗传病(Sutherland and Richards, 1995)。据估计每发生一个“普通的”点突变, 就至少伴随一个微卫星突变(Ellegren, 2000)。这样每代每人的突变数实际上就

多了一倍，从100个增加到200个。

除核内点突变、线粒体突变、和微卫星突变外，还有种类繁多的更严重的染色体突变，通称为“宏观突变”。这包括缺失和插入。根据康卓寿(Kondrashov, 2002)的估计，点突变与宏观突变的比例大致是100:4。他的估计似乎只考虑了最小的宏观突变，而不包括大段的DNA插入/缺失。虽然宏观突变相对较少（每代每人也许只有4个），但这些“主要”突变将无疑对基因组产生更大的破坏，所以如果要设计一个阻止基因组退化的选择方案，则必须优先考虑这类突变。宏观突变所影响到的核苷酸数量不一，少至1个，多至100万个，就如我们可能偶然删掉这本书里的一个字母、一个词、或者整章内容。这些宏观突变所造成的序列改变比所有点突变的总和还要大3-10倍(Britten, 2002; Anzai, 2003)。算上了宏观突变，总突变计数就提高到每代每人208个以上。但如果考虑到宏观突变所造成的核苷酸改变比所有点突变之和多3-10倍，每人携带的核苷酸改变总数就多得多了！然而这些数字可能还是太低。我们还未考虑倒位和易位。还有，进化理论学家正提出极高的基因间转化率，可能使这些数字再加一倍。现在想想，本章开始我们提到诺贝尔奖获得者、遗传学家穆勒，他曾认为，如果每代每人突变率为1.0或更高的话，人类必将迅速退化。我们虽然不知道人类突变率的确切数字，但还是有理由相信每代每人的核苷酸改变远超过了1000个（参见表1，第40页）。

在所有的突变之中，有多少是真正中性的？关于大部分基因组的功能性，人们的认识在过去的几年里有一个戏剧性的转变。“垃圾DNA”的概念正在迅速消失(Slack, 2006)。事实上，

所谓“垃圾DNA”（不编码蛋白质的DNA）正是决定生物复杂性的关键 (Taft and Mattick, 2003)。塔夫特(Taft)和马蒂克(Mattick)的研究强烈提示一个生物体的“垃圾”越多，这个生物体就越高级。所以在“垃圾DNA”内的突变不能假定为中性。

早在2005年，人们已经发现人类基因组中有50%以上的DNA被转录成RNA(Johnson et al., 2005)，而且被转录的DNA中至少有一半是双向转录的(Yelin et al., 2003)。所以这些DNA序列不仅是有功能的，而且其中很大一部分或许有双重功能。虽然基因组中只有一小部分直接编码蛋白质，但每一段蛋白编码序列的周围也都是功能性序列，它们的作用是调节蛋白质的表达。这些调节性序列包括启动子、增强子、内含子、前导序列、尾随序列、以及影响局部折叠和DNA结构的序列。我不相信至今还有哪一位严肃的生物学家会认为内含子（一个典型基因的大部分序列）真是中性的“垃圾”。事实上，许多已知最保守（必要的不可改变的）的序列就存在于内含子中(Bejerano et al., 2004)。虽然一个典型的蛋白编码序列可能只含有3000个核苷酸或者更少，但控制蛋白表达的序列则长达50000个核苷酸。人类基因组中有超过20000个蛋白编码基因（有不同的估计），如果我们把所有相关的核苷酸（每个基因50000）加起来，全部基因的总长度很容易就达到15亿个核苷酸。这正好是基因组的一半。另外，已发现一类全新的基因不编码蛋白质，只编码功能性RNA。在用计算机搜索蛋白编码序列时这类基因被忽略，所以一直没有被重视。然而它们是真正的基因，而且可能构成基因组的一大部分 (Mattick, 2001; Dennis, 2002; Storz, 2002)。在从前一直被认作垃圾的区域里，人们最近发现

了这些基因。此外，有两组科学家研究了基因之间的大段DNA序列，都发现这些序列有着广泛的功能 (Koop and Hood, 1994; Shabalina et al., 2001)。这些区域也曾经被认作垃圾。假基因曾一度被认为是僵死的重复基因，但最近也发现是有功能的 (Hirotsume et al., 2003; Lee, 2003)。假基因的功能似乎是制造调节性的RNA分子 (参见Chen et al., 2004)，而不是编码蛋白质，所以它们不是“死化石”。正如我将在后面的章节要详细讨论的，就连那些被人丑化的“自私基因”也似乎有许多不同的生理功能，这些基因有时被称为“寄生性DNA序列”，也称为“转座因子”。这些因子在细胞内发挥着多种极其重要的功能，包括控制染色体配对 (Hakimi et al., 2002) 和DNA修复 (Morrish et al., 2002)。长期被认为是垃圾的重复性DNA，包括卫星DNA，已证明对维持基因组功能起着必不可少的作用，并组成象着丝粒和端粒这样重要的基因组结构 (Shapiro and Sternberg, 2005)。最后，有些基本的结构模式贯串于全基因组的每一部分，比如GC等值区 (Vinogradov, 2003)、全基因组范围的“组词法” (Karlin, 1998)、以及核小体接合点 (Tachida, 1990; Segal et al., 2006)。这些大范围模式对于细胞功能至关重要，这也提示整个基因组都有功能。例如，核小体接合 (染色体结构和基因调节的关键) 似乎是由每十个核苷酸重复一次的双核苷酸模式来规定的 (Sandman et al., 2000)。所以，仅为了规定核小体接合点，就可能有五分之一的基因组是功能性的和必要的 (Tachida, 1990)。我们日益清晰地看到，大部分或全部基因组都是有功能的，所以多数基因组突变肯定是有害的。

就每个人来讲，100个突变相对于我们庞大的基因组内的

总体信息来说只是一个极微小的损失。然而，如此高突变率的真正影响表现在群体的水平上，而且主要靠着时间的推移而显现出来。考虑到全世界的人口已经远远超过了60亿，每人平均带来100个新突变，仅我们这一代就为人类增加了6000亿个新突变。如果我们记得人类基因组只有30亿个核苷酸位点，就不难明白，在我们有生之年人类基因组中每个核苷酸位点就已经发生了200次突变。所以，仅在我们有生之年，人类基因组中可能发生的任何点突变都已经发生了许多次！由于现今人口数目巨大，突变对人类的冲击是史无前例的。大多数突变的后果不会立即彰显，而是要在后代中表现出来。

我们将会看到，即使我们能够遏制进一步的突变，也没有什么选择方案可以逆转在我们这一代已经发生的破坏。这些突变中相当多的一部分将会向群体深处漂移，进而为人类造成永久性的基因破坏，再强大的选择也无力阻挡。然而我们的下一代，下下一代，还会增加新的突变。在可预见的未来，这一退化过程将会持续下去，我们正在止不住地下滑。

当选择不能消除因突变而带来的信息损失时，一种称为“误差灾难”的情况就出现了。如果得不到迅速纠正，这种情形将导致物种的死亡，即灭绝。在其最后阶段，基因组退化导致繁殖力减弱，繁殖不足又阻碍进一步选择 (选择总是需要过剩的群体，每一代都要淘汰掉一些)。最终选择完全被近亲繁殖和基因漂变取代，群体迅速消失。这个过程是不可逆转的沉沦。这个基因组退化的晚期阶段称为“突变熔消” (Bernardes, 1996)。大家普遍认为突变熔消是今天所有濒危物种所面临的直接威胁，同样的过程对于人类也似乎是潜在的理论上的威胁。

谁能防止突变熔消？

2008年补充：2007年6月，一个大型国际性基因组科学家联合体（称作ENCODE，“DNA元件百科全书”）以29篇系列论文的形式发表了他们的最新发现。这些研究结果轰动了遗传学界(Kapranov et al., 2007)。他们显示人类基因组比预料的要复杂得多，几乎全部基因组都被转录，而且大多数是双向转录。他们的结论是大多数核苷酸不但有功能而且有多重功能，扮演着多种角色。这意味着基因组的功能性超过100%（大多数DNA的双股都有功能）。鉴此，没有突变应该被看作是“完全中性的”，几乎所有突变都必须认为是有害的。这就是说人类真正的有害突变率必定是惊人的，大约每代每人100个，比之前的估计高出一百倍。

2014年补充：自2008年本书再版之后，薛 (Xue et al., 2009) 和林奇 (Lynch, 2010) 进一步证实了人类极高的突变率，大约是每人100个。林奇强调人类基因组绝大部分都是垃圾DNA, 所以几乎所有的突变都是无害的。不过，他认为人类基因突变的积累仍然是个严重的问题（附录1）。

有关人类生殖细胞突变率最新、最完整的分析见于坎贝尔和艾克勒的一篇综述(Campbell and Eichler, 2013)。文章的表1总结了关于人类基因突变的多项研究结果。表中的有些估计只涉及一个家庭，所以不很可靠。根据其中样本量超过20个家庭的六项研究，平均点突变率是每代每个性细胞每个核苷酸 1.56×10^{-8} ，据此可以算出每个新生儿有93.6个突变。但是这篇综述的作者指出这只是一个低限。他们写道：“……值得注意的是，所有这些研究都为了避免假阳性而滤掉了很多新变异，而且这些研究通常排除了基因组中的高度重复区域……针对外显子或其他区域的序列分析曾得出更高的突变率……”不仅如此，突变造成的核苷酸改变大部分不是点突变。所以每代每人100个突变仍然是对人类突变率的很保守的估计。

其他类型的突变呢？这篇综述的作者也指出，除了点突变以外，基因组中还有其他类型的破坏性更大的突变，包括微卫星突变、中心粒重复序列突变、插入/缺失突变和拷贝数目变异(CNV)等。他们还没有提到线粒体突

变、倒置、易位和转换。

这篇最新综述的作者证实，宏观突变是基因组分化与退化的最重要来源。譬如，他们指出拷贝数目变异在每一个新生儿中带来16000到50000个碱基对改变。类似地，他们发现插入/缺失突变率比先前认为的要高。之前曾认为插入/缺失突变率只有点突变的6%，最近的估计表明这个数字应该更靠近10%，甚至20%，而且会随着测序方法的改进而升高。这就是说每个新生儿可能有20个插入/缺失突变。由于插入/缺失都涉及许多核苷酸，从几千到几百万，所以有理由认为，仅插入/缺失突变就造成了每代每人千千万万的核苷酸改变。

2013年魏尔士(Wells)详尽地综述了近期的相关文献，显示人类基因组大部分是有功能的，这意味着有害突变不少于每代每人100个。2012年ENCODE的一项大型报告也支持这一论断(The ENCODE Project Consortium, 2012)。

最近有一篇文章(Rands et al., 2014)宣称人类基因组的8.2%在进化上是“受限”的（就是说自从最早的哺乳动物出现以后就没有太大变化）。这是基于人类基因组与其他动物基因组的相似性。这倒没有什么，但是进化论者现在据此宣称其余的91.8%的基因组必定是“垃圾DNA”。这一结论就很不理智了。人类基因组中有些部分与大多数哺乳动物相同，这并不足为奇——基因组中的这些区域一定是编码大多数哺乳动物所共有的功能（比如乳腺、共同的生化功能等等）。然而，显而易见地，基因组中的其他部分必定编码每种动物所特有的功能。人、鲸、蝙蝠，各有不同的功能。哺乳动物基因组中大部分序列之所以各不相同，是因为其中的信息各具不同的功能，让人去做科学，让鲸能潜入一英里的深处，让蝙蝠能飞行并进行超声定位。

突变类型	人均突变数	人均核苷酸改变
1. 线粒体突变 ^a	<1	<1
2. 核苷酸置换 ^b	75-175	75-175
3. 微卫星突变 ^c	75-175	75-175
4. 缺失 ^d	2-6+	1-3000+
5. 复制/插入 ^e	2-6+	1-3000+
6. 倒位/易位 ^f	许多	数千?
7. 转换 ^g	数千?	数千?
每代每人总数^h	100++	几千!

表1 人类基因突变种类繁多、频率惊人

突变有很多类型，各类突变都导致可遗传的基因改变。不幸的是，每种突变都造成信息的净损失。在人类，线粒体突变最稀少，人均不到一个。然而每人不足一个的线粒体突变已经引起一位进化论者的警惕：“我们应该进一步注意到一个更广义的问题，即进化而言，生物体如何（能否）容忍其遗传系统内存在如此重的突变负担。”(Howell et al., 1996)。现在，所有各种突变加起来又要如何！

a 对线粒体突变率的估计说法不一，从每人0.1到每人1.0(Parsons et al., 1997; Carter, R., 2007)。

b Kondrashov (2002), Xue (2009), Lynch (2010), Campbell and Eichler (2013)。

c 对核苷酸置换率的估计通常不包括突变热点如微卫星区。一般认为，微卫星突变率与点突变率大致相等。

d, e 康卓寿 (Kondrashov, 2002) 估计缺失与插入发生的几率各有点突变率的2-6%，二者加起来即点突变的4-12%。然而，他考虑的只限于小范围的插入和缺失，所以实际的插入/缺失数目可能更高。坎贝尔和艾克勒(Campbell and Eichler, 2013)提出插入/缺失的频率可能比康卓寿的估计要高三倍。因为缺失和插入的片断可能很大，就核苷酸数量改变而言，它们的效应总和是所有点突变的3-10倍。

f 染色体重排的实际发生率未知。进化论认为人类是不久前从黑猩猩中分化出来的，据此他们说这类变化的发生率一定很高。这些改变能影响很大的DNA片段，所以要使进化论者臆想的情况行得通，每代必须有平均数以千计的核苷酸改变位置。

g 基因间转化的发生率未知，但进化论的断假设要求不同位点上的基因之间转化率极高，每代每人数以千计。

h 突变总数的估计只能非常粗略，但很清楚，各类新突变，包括转换，其总数必定远超过每代每人100个。这些突变，包括许多宏观突变，无疑使每代每人发生产成千上万的核苷酸改变。

atcg

第四章

全能的选择来拯救？

简讯：选择的能力非常有限。

人类遗传学家一致认为，由于突变的快速累积和自然选择的松弛，现今人类正经历着基因退化(Crow, 1997; Lynch, 2010)。这些遗传学家意识到突变正在群体中累积，而且其发生的速度比以往所想像的要快得多。遗传学家普遍认为，这些突变的效应不是中性就是有害（就算存在有利的突变，因其过于罕见也往往不予考虑）。他们进而认识到遗传信息正在流失，这将最终导致人类适合度的下降。适合度下降的速率为每代1-2%(Crow, 1997)（参见图4，第60页），但也可能高达5%(Lynch, 2010)。这一切都发生在基因水平上，虽然医学和科技的发展在社会水平上可以延长我们的平均寿命。所以大多数人类遗传学家也许会同意，如果我们要阻止基因退化，就必须增加选择强度。然而，基本上没有人公开这样说。试想，这样的说法可能带来什么样的政治影响！

这就带来一个有意思的问题：“需要多大的选择强度才能制止基因退化？”真正的问题也许是：“选择能够制止退化吗？”

对很多人——包括许多生物学家——来说，自然选择就像

一支魔棒。它似乎无所不能。这种对自然选择抱天真想法的人随处可见。即使作为一个植物遗传学家，我也曾对自然选择究竟如何运作存有不切实际的理解，对于选择如何在全基因组水平上发挥影响，也曾抱极其幼稚的认识。真正分析过选择在全基因组水平上能有多大作为的科学家是为数不多的群体遗传学家（一群非常专业化的人）。群体遗传学是一个高度理论化和数学化的领域。理论数学家们完全被他们的公理（假设）所束缚，他们的方程式只能基于这些公理之上。整个群体遗传学是由一个小圈子里的人建立起来的。这个小圈子对基本公理是完全和极端地忠诚。今天，这仍然是一个很狭小的领域，其成员仍然全部是基本公理的“真实信徒”。这些人智能极高，但被彻底地无条件地束缚于基本公理之上。大多数时候，其他生物学家甚至不明白他们的工作，而是靠信心去接受他们的结论。然而，正是这些遗传学家们自己透露了自然选择存在的最大限制（参看附录1）。因为自然选择不是一支魔棒，而是一个真实的现象，所以它既具有实际的功效，又受到实际的限制。它不是“全能的”。

最基本的问题

公主和核苷酸的悖谬

自然选择面临着一个根本性的难题，即在基因型的改变（分子突变）和表现型的选择（在整个生物体水平上的选择）之间存在着一个巨大的鸿沟。在分子水平上有数十亿种不同的突变，其间的差别既微乎其微，又细致复杂。自然选择必须从

中作出挑选，而实现这一目标的唯一途径只能是在整个生物体的水平上控制其繁殖数量。当自然界选择或淘汰群体中的某一个体时，它必须一次性地接受或拒绝其中的一整套60亿个核苷酸！在基因组这套说明书中，要么所有的字母统统接受，要么一个字母也不要。事实上，大自然从来看不见单个的核苷酸，它所看见的只是整个有机体，它没有能力看清并选择某一个核苷酸。大自然通过选择或淘汰整个有机体，竟然能够精确地控制一个群体中数十亿拼写错误的命运，我们看到要相信这种理论需要何等大的信心。

基因型改变和表现型选择的问题很像儿童故事《公主和豌豆》里的情形。公主的贵胄品性表现在，即使隔着13层床垫，她还是能感觉到床底下有一颗豌豆，因而无法入睡。孩子们喜欢这个故事，因为它太滑稽了。不论是否皇室贵族，无人能感觉到13层床垫之下的豌豆。然而我们的遗传学问题实际上比这更困难。我们的公主就是在整个有机体水平上的自然选择，她的工作就是决定让谁有机会繁殖。为此，她需要隔着一系列床垫来阅读大量用盲文写的书，以准确地查出哪些书中的错误最少！这听起来是个绝妙的神话，但谁会相信它，并用它来解释生命的形成过程呢？这个问题可称为公主和核苷酸的悖谬，就是图5所示的怪诞图画（第62页）。

公道地讲，有少数突变的效应比这个比喻中的盲文字母大得多。有些罕见的突变会导致重大的生物效应，就像床垫子下面的一个保龄球。自然选择显然能够淘汰掉这些主要的突变，但“保龄球”（半致死）突变的数目极其稀少，这些核苷酸位点所携带的信息只是基因组总信息中极微小的一部分。基因组

中大部分的信息，是存在于效应比上述比喻中的盲文字母更微小的核苷酸里。我们要了解的正是那些弱效应核苷酸的起源和维护。所谓公主和核苷酸的悖谬，实质上就是说大自然（自然选择）在撰写和不断地校对生命之书的过程中，根本看不见书中的字母。用专业语言来说，自然选择是基于表现型，而非基因型。

从一个核苷酸到一个生物体的距离是难以想像的，这种差异的一个方面就是在体积上。如果我们把一个核苷酸放大到豌豆那么大，按比例，公主就大约有10000英里之高。不仅如此，在核苷酸和生物体之间还隔着许多的组织层次。比如，一个核苷酸可能影响到一个基因的转录，继而影响mRNA的加工，进而反映在某种酶的多寡上，某个代谢途径因而受到影响，然后细胞分裂受影响，然后影响某种组织，组织又影响整个机体，从而改变机体繁殖的机率，最后才落实到这种突变遗传给下一代的机会。每一个组织层次都带来大量的不确定性，使效果被淡化，同时“噪音”大量积累，分辨率逐步丧失。一个核苷酸（一个分子）和一个生物体成功繁殖的机率之间的关联小到几乎不存在！核苷酸和生物体简直就是两个世界。我们的公主（整个生物体水平上的自然选择）必须察觉到仅比原子水平高一点的差异。

通常我们看不到电视机上的像素，所以想像一下，试图通过用肉眼检查每个像素的质量，在种类繁多的电视机中挑选特定的一台，这是多么地不可思议。然是事实比这还要糟糕。在一个生物系统里，我们讲的是像素，像素里的像素，就这样一层一层地下去。我们所说的是从一个核苷酸的直接效应到其对

整个有机体的远距离影响的一长串事件。每一步都要呈对数地将信息稀释。在每个层次上，相关性都要损失一个数量级。这就好比要测量一只蝴蝶的扇动对一个一千英里以外的飓风系统的影响，也有点像根据一支军队的表现，来选择它里面的某个特定士兵，这是一幅完全颠倒了图画。然而这正是基本公理的实质！基本公理的倡导者看到的是一个人类基因组（60亿个核苷酸），而想像其中每一个结构单位，都可以仅靠整个生物体水平上有限的繁殖筛作用而被独立地选择（或淘汰）。我们下面将要看到，这是完全不可能的。

为了更好地理解公主和核苷酸悖谬的本质，让我们想像一种改进教科书的新方法。从一本高中生物化学教科书开始，假设它相当于一个简单的细菌基因组。现在开始引入随机拼写错误、重复和删减。全国的中学生每人将得到一本稍微不同的教科书，每一本里都有一系列独特的随机错误（每本书大约有100个新错误）。学年结束时，我们对所有学生进行测试，然后只保留成绩最好的前100名学生的课本，将这些课本用作下一轮复制的蓝本，复制时引入新的“错误”，如此循环下去。这样我们能期望教科书持续改进吗？当然不能！我们能期待平均学生成绩不断提高吗？当然不能！

大多数人都能明白，在以上的例子中，教科书里的拼写错误几乎无一是有益的。更重要的是，课本中的细微差异和学生的成绩之间并不存在有意义的关联。为什么？因为每本书的缺陷大致相当，与其他事情相比，课本之间的差别微不足道。

“其他事情”指的是什么？我指的是一个学生的成绩取决于许多别的因素，包括个人能力和个人情况（老师、课室、其他孩

子、学习动力、家庭生活、浪漫生活、睡眠不足、“倒运”等等)。所有这些其他因素(我将称之为噪音)将抵消课本中少数拼写错误的效果。如果一个学生在测试中取得了高分,并不是因为他的课本错误较少一点,而主要是因为以上所说的各种原因。

如果这个突变/选择循环持续下去将会如何?教科书显然会退化,而学生的平均成绩也终将下降。然而,这个荒唐的突变/选择体系,却是生物学基本公理的一个很合理的比拟,它肯定不会提高甚至不能保持学生成绩。这种选择失败的最基本原因,是课本中的个别字母与学生的整体表现之间关系不大,相关性基本为零。所以,这是公主和核苷酸悖论的极佳比喻。你若觉得太荒唐,还有一点要明白:基本公理声称全本生化教科书最初正是由这一突变/选择体系所撰写的,从来没有任何智能中介作为作者或编辑。

当我们理解了**内环境稳定**现象时,公主和核苷酸悖论的问题就显得更为严重。内环境稳定是所有生物随环境改变而自我调节的现象,这种调节机制进一步中和了大多数突变的效果,使得更多的突变为选择所忽视。一个很好的例子是寒冷气候下的温血动物。内环境稳定是通过每个细胞内极其复杂的感受和调节网络而实现的。虽然很难详细解释,但大家都同意它在所有生命体系中都在运作,而且很有效。内环境稳定现象有点像一个超级的、自我调节的床垫。如果一个网球被放在床垫底下,床垫会自动地通过无数复杂的机械机制来自我调节,以求表面平坦。然而在真实的生物界,情况更像是13张层层叠加的能够自我调节的床垫(内环境稳定运作于生命组织的每个层面

上)。这样就让我们的公主更为难了,因为她需要感觉到垫子底下的豌豆(或读盲文)。

当孟德尔的遗传定律在达尔文发表《物种起源》将近50年以后被“重新发现”时,遗传学家们意识到在任何特定群体中必然有大量可分离的遗传单位。不久他们又认识到如果遗传单位数目很大就遇到一个问题。虽然他们没有用类似的语汇,但基本上就是我所称的公主和核苷酸的悖论。早期的群体遗传学家都笃信达尔文主义哲学,他们意识到必须找到一个办法克服公主和核苷酸悖论,以使达尔文理论在遗传学上讲得通。^{*}所以他们很精明地将选择单位从整个生物体转移到遗传单位(如基因或核苷酸)。为此他们必须重新定义群体的概念,使其只不过是“基因库”而已。这样一来他们就可以声称真正的选择是在基因库中的核苷酸水平上而不是在个体身上进行的。人们就可以认为,每个核苷酸可以独立地被选择、被淘汰、或者既不被选择也不被淘汰。这样就很容易想像任何选择和进化过程,不论生物学状况如何复杂。公主底下的床垫子被彻底抽走了,好像她突然能够感觉到每颗豌豆了,甚至可以直接读每个盲文字母了!整个问题就这样被掩盖,有效得令人瞠目。事实上,如果没有这一重大智慧发明,达尔文主义此时已寿终正寝了(Provine, 1971)。

如此重新定义存在一个严重的问题:新图像绝对是假的。

^{*}“霍尔丹…费歇尔…和赖特…试图…打消孟德尔主义已经推翻了达尔文主义的信念…费歇尔、霍尔丹、和赖特将孟德尔遗传学和自然选择量化结合起来建立了群体遗传学。”(Provine, 1971)。

群体与基因库相去甚远，而选择也从来不作用于单个核苷酸。为了使这个偏激的新生命图像合理化，这些理论学家们就必须堂而皇之地做出几个显然是错误的假设。例如，他们必须假定所有遗传单位都能够自由地组合，以便每个核苷酸可以独立地被遗传，好像没有基因连锁块的存在（完全错误）。同样地，他们必须假定没有上位效应，好像核苷酸之间没有相互作用（完全错误）。他们还通常假定近乎无限大的群体（显然是错的）。他们常常暗示无限的选择时间（显然是错的）。而且他们一般地还假定无限多的性状可以同时被选择（我们将指出其错误）。从群体遗传理论一开始，就需要很多不切实际也不合理的假设来支持这一模型，使它貌似可行。

现代群体遗传学的理论柱石就是建立在这个虚假的基础之上的。尽管其模型与生物学现实不符，但这些人有着不可思议的学术权威光环，他们的论证很抽象，使用的是让很多生物学家望而生畏的高度数学化的公式。不仅如此，大多数生物学家也是忠实的达尔文主义者，在哲学上与群体遗传学家是一致的。即使故事不太合理，他们也欣然紧跟。事实上，早期群体遗传学家迅速被奉为“科学的宠儿”。我还记得自己做研究生时在群体生物学课堂上的情形，自己天真而谦卑地接受和默认了将生命非常不自然地重新定义为“基因库”。我记得自己不是很明白，以为这是我的问题（虽然我们班上所有其他同学也好像有同样的问题）。既然我“知道”进化论是正确的，我对生命是核苷酸库的概念理解不透并没有关系。如果群体遗传学界的泰斗说它是正确的，我哪有资格提出疑问。即使他们的前提是错误的（例如核苷酸的自由组合），他们的结论无疑也还是正确的，他们是天才！虽然我事

实上是个思维比较自由开放的学生，我还是全盘而无抵抗地接受了公主和核苷酸的故事。

在教条之外，生物学的事实到底如何？事实是选择过程作用于生物体水平，而非核苷酸水平（参看Crow and Kimura, 1970, 第173页）。人类基因从来不存在于“库”里，而是存在于真人体内的大型基因组合中。在特定人体内每一个核苷酸与其它核苷酸是密切联系的，整套60亿个核苷酸一起被选择或被淘汰。我们将看到，**连锁现象**是高度的和广泛的。没有一个核苷酸会独立遗传。每个核苷酸与它周围的核苷酸是紧密联系着的，就如这页书的每一个字母与一个具体的词、句子、段落和篇章的联系一样。这本书不是一个字符库，不是由一个像大型游戏机一样的选择系统通过独立地、随机地选择每一个字母而制造出来的。每个字母都是根据设计而设置的，作为一个词、一个句子、一个段落和一章书的一部分。核苷酸也是一样，它们只存在并作用于其它核苷酸背景之中（这就是我们所说的上位效应）。我们现在知道人类核苷酸是以连锁串或连锁块的形式存在的，每个连锁块大小不等，由1万到100万个核苷酸组成。每个连锁块就是一个遗传单位，从不分裂。这就完全否认了理论学家们最基本的假设之一，即每个核苷酸可以看作一个独立的选择单位。既然群体遗传学的生命模型（基因库）是绝对错误的，那公主与核苷酸的悖谬还是完全没有解决。这对整个群体遗传学科应该是极其尴尬的。从实际的角度来讲，这意味着自然选择从来不能创造出一长串核苷酸序列。

不但公主与核苷酸的悖谬没有得到解决，我们现在知道问题比早期群体遗传学家所想像的更糟糕得多。我们现在知道基因

组的规模和复杂性（盲文书的规模）是庞大的，也知道内环境稳定（床垫的厚度）是多么深广，而且基因型和表现型之间有更多的组织层次（床垫的数量）。我们要等到第六章才可以完全了解生物“噪音”的问题（原来床垫本身就充满了豌豆大的硬块）。所以我们应该明白公主与核苷酸的悖谬是整出戏的终结点，基本公理在这第一层最基本的水平上失败了。连孩子都应该能够看得出，虽然很多成年人“受太多教育”而看不到。然而公主与核苷酸的悖谬只是基本公理所有问题的开始。为了进一步的讨论，为了本书后面的内容，我将欣然容忍理论学家们以“基因库”作为生命的模型，以及选择过程作用于单个核苷酸水平的观点。我同意这么做是因为，我们将看到，还有许多其它问题能完全推翻基本公理。切记，公主与核苷酸的悖谬本身就足以反驳基本公理（参看附录3）。

选择的三个具体问题

为了理解基因组选择的基本问题（你很快就会发现这已经不简单），让我们举一个简单的例子来说明。假设只有一个突变（而不是几百万个），而且允许我们进行人工选择（而不是自然选择）。假设这是一个点突变，已经在人群中累积，以致于有50%的人口携带这个突变。什么样的人工选择可以排除这个突变呢？有哪些关键因素会影响选择呢？为使问题简单化，我们假设这个突变是显性的（几乎所有突变都是隐性的，选择起来困难得多）。乍一看，这个问题好像很容易解决。我们能在一代之内排除这些突变，如果我们可以：1）牺牲50%的繁殖人

口；2）确认携带突变的每一个人；3）100%地阻止突变携带者婚配。那么问题在哪儿呢？

1. 选择的代价

“选择的代价”问题最初是由霍尔丹(Haldane, 1957)描述，后来被木村和太田 (Kimura and Ohta, 1971; Kimura, 1983) 证实和发展。赖敏(ReMine, 1993, 2005)又作了进一步的阐释。所有选择都涉及到一个生物代价问题，即选择必须排除(或“耗费”)部分能繁殖的群体，选择性排除就是选择的实质。在目前的例子里，我们应该问：“我们真的可以排除50%的人口，即阻止一半的人群繁殖，以加快选择的速度吗？”依照现今人类繁殖的水平（全球每两个成人有不到3个孩子），如果为了选择除掉50%的繁殖人口，人口总数会（在一代内）下降25%。显然，每对成人必须有至少两个能繁殖的孩子以保持群体的数量。然而，并不是所有的孩子都会继续繁殖（由于意外死亡和个人选择等因素）。所以，每两个成人需要生育大大超过两个孩子才能维持群体的生存。如果每两个成人需要三个孩子才能保持人口数目的话，考虑到我们的低生育率，可容许的选择为零。由于这些原因，可供为选择所“耗费”的孩子远少于三分之一。霍尔丹(Haldane)相信，典型的自然人群中实际上只有10%可供选择所用。如果为了选择的目的每代除掉50%的群体的话，世界人口会迅速缩减，导致人类灭绝。所以，在以上例子中，在一代之内排除所有突变个体是不合理的。然而，在两代之内作这样的选择可能是合理的，因为每代只有25%的群体被选择过程耗

费。

举这个简单的例子是为了说明选择虽然有用，但其力度受到明显的限制，而且我们要明白每次选择都有一个生物学上的代价。要同时选择许多不同的突变就成大问题了。很明显，对于人类群体来说，能供选择“耗费”的群体数量远低于33%。根据霍尔丹的现实估计，也许在10%上下。相比之下，我在康奈尔大学作植物育种的时候，可以轻而易举地为选种而“牺牲”掉99%的种群，因为植物繁殖力极强。

“选择的代价”这个概念极为重要，需要进一步论述。一个物种的正常繁殖率显然应该是每对个体至少两个后代，不然物种就迅速灭绝。然而，仅仅为了生存，每个物种的繁殖率都要比这高许多。必须有过剩的群体来预备许多情况，包括遗传因素和非遗传因素的折耗。例如，成功繁殖的偶然性很大。群体中的许多个体死亡或不育，其原因与基因无关。车祸或阵亡与一个人的基因“适合度”几乎没有关系。随机死亡的代价是绝对的，在考虑选择之前就必须“支付”。在有些物种，这个代价可达总群体的50%或更高。在这种情况下，我们需要每对个体至少生育4个后代才可以避免绝种。此时选择还没有开始。当我们实际考虑决定繁殖成败的基因因素时，有些重要的基因特征并不遗传给后代。这些基因特征是不具“遗传力”的。例如，很多基因在某些组合中发挥作用，而作为单个的基因就毫无益处（这就是杂种优势或上位效应的情况）。选择了这些基因组合其实是“假选择”，因为它对于后代没有好处。这些基因组合在减数分裂时被分离而不能进一步遗传。然而这样的“假选”也是有代价的，如此以来，对繁殖率的要求就更高

了。而我们还没有开始为“真”选择付代价！真正的选择可以有几种形式：稳定性选择、性别选择、渐进选择等等。每种选择形式都有一个繁殖的代价。繁殖代价是相加性的，所有的代价都得付。总繁殖代价绝不能超过一个物种的实际繁殖能力。只有当一个物种有充足的繁殖力，也有充足的过剩群体来支付所有其他代价时，任何形式的选择才有可行性。换句话说，只有首先付上所有其他代价后，仍有盈余群体时，才有选择的可能性。选择有点像一个预算紧张的家庭的额外花费。问题最终总是回到：“我们付得起吗？”

适合度（总体生物活力）实际上是让自然选择发挥作用的真正性状，而这个基本性状实际上遗传力很差。这种说法有悖于主流思想。根据木村资生的论述，适合度的遗传力可能低至0.004（Kimura, 1983, 第30-31页）。我们在第六章会详述遗传力的概念。目前只要知道低遗传力意味着环境因素对性状的决定作用比遗传因素更重要。木村讲总体适合度遗传力很差是非同凡响的认知。这就是说即使在强选择压力下，几乎全部过剩群体都被“耗费”在排除非遗传性变异上，因而**多半繁殖淘汰都无济于事**。换句话说，选择总体适合度对下一代基因组成的影响微乎其微。木村的结论暗示着，过剩群体中只有一小部分真正用于排除突变（这正是我一直在说明的）。暂且把这个重要的事实搁置一旁，我现在要极其大方地允许将所有“选择资金”（多余群体）都花在排除突变上。除非另有说明，我将假设全部过剩群体统统用于选择以排除突变。然而，读者要明白，事实上，只有一小部分过剩群体可供排除突变（参看第六章，图8a-c，第91-93页）。

我认为理论学家们在发明他们的各种方案时，经常犯的最基本的错误之一就是忽略选择的代价。他们花费选择资金就好像一个少年人拿着一张信用卡。他们做推想时好像永远有无穷无尽的过剩群体。由于理论家们通常不受实际代价的限制，他们的头脑中想像着同时能付得起任意数目的选择方案。所以他们能推出一个又一个的选择方案。这让我想起了老西部影片中牛仔把“六轮子”连放几十枪竟不用装子弹，或者像电影《指环王》里的莱戈拉斯永远有射不完的箭！电影毕竟是离奇的，影片制作人可以自由地宣称“艺术的真实”。遗传理论学家们没有艺术执照，应该对他们如何花费选择的资金负责，就像一个会计对钱花在什么地方负责一样。应该根据特定物种的实际繁殖情况，给理论学家分配一个现实性的选择资金数额。然后他们应该理智地“花费”这部分多余的群体，而不要出现赤字。如果诚实地采用这个原则，会发现对选择的期望将大打折扣。

2. 识别隐藏的（“看不见的”）突变

若要淘汰一个突变，就必须首先能在群体中认出它。但除非用非常昂贵的DNA测试来筛查整个群体，我们不可能识别一个典型的点突变的携带者。在现实中，只有那些极其罕见的、造成严重畸形的点突变才能够被确认。对于大多数点突变来说，即使它们是破坏性的并造成信息流失，但它们是如此微弱，基本上是“看不见的”，它们不会产生明显的可识别的效果。要想人为地淘汰某个典型的点突变，我们需要对地球上的每个人

进行昂贵的实验分析，而这根本是行不通的。所以，我们看到为了选择的目的是要确认“好的”和“坏的”个体，这里有一些根本性的问题。显然，当考虑到同时选择几百万个突变时，这个问题变得不可思议。试想你要买一套大百科全书，并且知道每一套都有数千个独特的拼写错误。你能够站在书店里检查所有书卷，期望找到错误最少的那一套吗？就说一共有两套吧，每一套各有10000个独特的拼写错误，你怎么才能知道哪一套错误更严重呢？选择成了完全的随心所欲了！这个数量众多的沉默的或“近中性”突变的问题是由木村（Kimura, 1968）和太田（Ohta, 1973, 1974, 1992, 2002）首先提出的，而其含意已为康卓寿（Kondrashov, 1995）进一步论证。

3. 系统性的繁殖淘汰

选择的代价和“沉默”的突变是大问题，但要阻止突变个体婚配问题就更大了。在这个星球上还没有一个社会制度能够精确地控制人类的繁殖。最臭名昭著的例子是在希特勒领导下的纳粹德国，但那个实验惨痛地失败了。计生站和现代生育控制手段，虽然对于降低平均家庭成员数目有效，也还不能淘汰突变。除了极少数因突变造成显著遗传缺陷的病例，一般情况下，人类的婚配和繁殖基本上还是随机的。

从上述很简单的例子可以看出，即使要人工选择人群中的一个点突变也不可行，这是很发人深省的信息。当我们进一步考虑多个突变时，问题还要按指数增多。即便我们可以确认各种突变的所有携带者，并有效地阻止他们婚配，迟早我们还

是会遇到选择代价的问题，因为要同时淘汰多个突变，我们会面临一个急剧缩减的群体的内在问题。所以我们看到选择并不像我们所想的那么容易！即使最简单的选择模式也需要几个重要的因素：1) 维持一定的群体数量；2) 清楚地鉴别突变体；3) 从繁殖群体中有效地排除突变体。我们在下一章将看得更加清楚，当我们同时考虑基因组中的所有突变时，这三个要求一个也不可能达到。

如果人工选择都不能清除一个典型的点突变，那么自然选择会来拯救吗？考虑到以上问题，一个可能的结论是，不用人类费心，让自然来为我们做吧（让不适者自然消亡吧）！问题是自然选择跟人工选择一样，也会遇到一模一样的问题。由于选择的代价，自然选择也不可能同时淘汰太多的突变，否则不是彻底失效就是造成群体数量锐减（或两者同时出现）。再者，自然选择也要能识别众多基本上是“看不见”的突变。最后，自然选择也需要设法从繁殖群体中，同时排除大量的突变体。同时淘汰在逻辑上是不可行的，因为有选择干扰。自然选择的这些真实瓶颈，将限制我们期望选择所能起到的实际作用。事实上，人工选择比自然选择有效得多。人工选择是基于人为的斩截，将“不适合者”全部排除，而自然选择是基于“概率选择”，只是降低了不适者繁殖的成功率（不完全消除——某些不适者仍将生存）。

基因选择仍然有效。请不要误解我的意思，我并非说选择根本不起作用，在一定水平上它是有作用的。作为一个植物育种专家，我的工作就涉及到人工选择。我和同事们经常培育出对现代农业有重要意义的植物和动物种类。后来我参与了植物

的基因工程研究，我们常规地用选择技术来获得转基因（基因被人工改造的）植物。同样地，自然选择已经排除了最有害的人类突变。如果不是这样，人类种族早就退化衰变了，我们也不会还活在这里讨论所有这些问题。然而自然选择和人工选择的作用范围都是很有限的，两者都不像传说的那样无所不能。选择不是一支魔棒！我绝对同意选择可以改变某些基因的频率，与此同时，我要指出没有一种选择可以维持（更不用说创造）高级基因组。总而言之就是：**选择有时能够在基因水平上发挥效力，但在基因组水平上一概是失败的。**

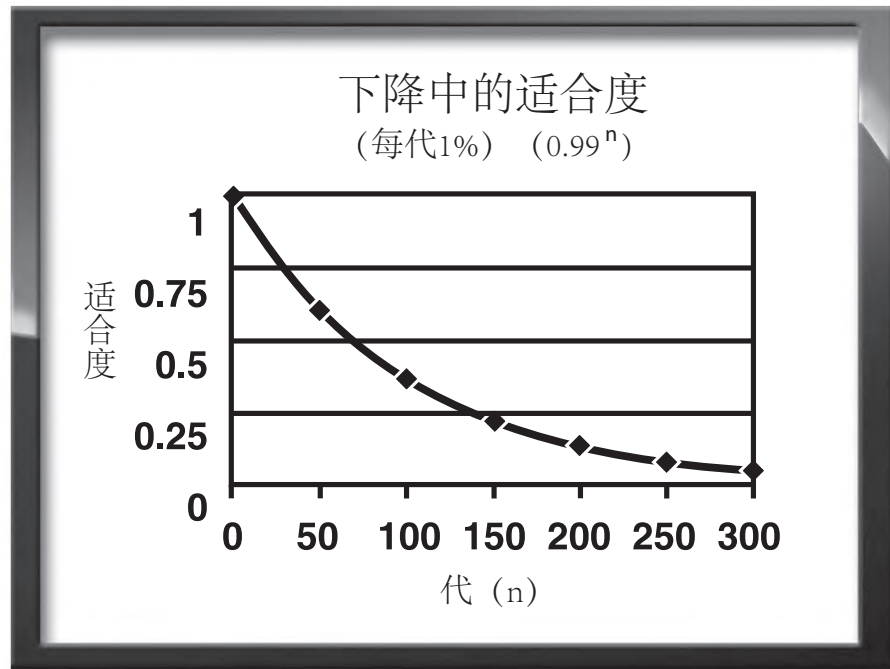


图4 基因熵的效果

克劳博士 (Crow, 1997) 指出他认为人类的适合度因突变的积累正在以每代1-2%的速度下降。图中所描述的是一个假想人群的适合度在300代 (6000年-9000年) 内以每代1%衰减的曲线 (起点适合度为1)。我们看到的是一个典型的生物学衰减曲线。适合度的这种进行性衰减无疑将在可预见的未来导致人类的剧烈退化。



图5 公主和核苷酸的悖谬

基本公理要求作用于个体水平上的自然选择（公主）必须识别分子水平上的几十亿个核苷酸的效应（豌豆）。在公主和核苷酸之间存在着一个巨大的鸿沟。这个鸿沟部分地是由于大小比例悬殊（试想：如果一个核苷酸有如豌豆般大小，那么一个人体就大约有10000英里高）。这个鸿沟也是由于多种组织层次所致（核苷酸与人体之间至少有十几层生物组织形式）。没有一个核苷酸会直接影响人体，所有的核苷酸都要通过精细复杂的多重系统而发挥作用。所以在公主和核苷酸之间的床垫多而且厚。不仅如此，生命的自我校正机制（称为内环境稳定）在每层生物组织上发挥作用，有效地缓和了大多数核苷酸的效应，就如现代消声技术使用负反馈来抵消声音。在我们的比喻里，我们需要将高科技的“自动校正”机器安装在每层床垫里面。然而我们还要再加上一点说明以使比喻准确。基本公理不但让公主去感觉她厚厚的床垫底下有没有一粒豌豆，而且通过这些床垫，她必须“感觉”到用分子大小的盲文写成的百科全书，并决定哪一卷的错误最少！

atcg

第五章

基因组选择可行吗？

简讯：选择不能拯救基因组。

上一章我们谈到选择人群中的单个突变时所存在的问题。传统上，遗传学家研究突变问题时，一次只考虑一个突变。大家普遍假定一个突变的情况可以拓展应用于所有突变。这显然是无稽的推测！就好像说如果我买得起一辆车，多少辆车我都可以买，或者说假如我能耍三个球，我就可以耍三百个！我们已渐渐明白真正棘手的问题并不在于单个基因或单个核苷酸，而是在于当我们将所有基因单位（整个基因组）一起考虑的时候。为了弄清楚如何才能阻止基因组的退化，我们必须在基因组水平上考虑突变的问题，而且必须实施我所称的“全基因组选择”。

人体基因组中有30亿个核苷酸位点，每个位点都有两个拷贝。所以共有60亿个可能的点突变（拼写错误）。过去只有几个有远见的遗传学家（例如霍尔丹、穆勒、木村、康卓寿）认识到了突变加选择在基因组水平上运作的深层困难，但这个棘手的问题屡屡被彻底藏在台面以下。这是因为它为进化理论造成了不可克服的难题。在过去几十年里，我们认识到基因组中有功能的部分比任何人所预测的要大得多，也复杂得多。我们知道人类突变率比以前认为物种生存所允许的最大值高得多。

我们意识到真正中性突变的百分比在不断地缩小，而能够增加信息的突变少到微乎其微。所以不仅突变率很高，而且几乎所有的突变都是有害的。鉴于所有这些原因，我们再也不能忽视基因组退化的问题了。我们必须问：“基因组选择能解决问题吗？”

说到底，我们每个人都是高度的突变体。这个破坏性的突变过程已经持续了很长时间。除了每个人为人类基因库新添加的约100个新突变外，我们还从祖先继承了大量的突变。为了正确地看待基因组选择的问题，我们要重温上一章开始提出的选择问题（单个核苷酸位点上的单个点突变），然后把问题放大至少十亿倍来看。你现在是否开始明白，在基因组水平上淘汰突变与前述淘汰单个突变的例子是有根本区别的？

1. 选择的代价

所有人都是突变体的事实使得选择更加困难。若要淘汰一切突变，没有人可以繁殖，马上就会绝种！显然，这种选择策略造成的繁殖代价太大了。大家都承认我们每个人继承了数千个有害的突变。作为一个群体，我们一共携带着数万亿有害突变。然而，为了把问题简单化，让我们将注意力只放在我们这一代人带进人类基因库里的6000亿个新突变（每人100个新突变乘以60多亿人口）。既然我们不能简单地淘汰一切“突变体”，我们只能在“多重突变体”和“少重突变体”之间作出选择。我们将看到的是，辨别“多重突变体”与“少重突变体”本来就是个大问题。考虑选择的代价，选择过程中每一代

所排除的人口必须远少于33%，否则我们会很快灭绝。

让我们来描述一下，选择6000亿个突变的代价究竟有多大。如果我们有60亿人口，最多只能“淘汰”（阻止其繁殖）其中的三分之一，就是20亿人口。想象一下，这样的念头足以让最冷血的优生主义者不寒而栗。然而采取这样严酷的措施有什么益处呢？阻止20亿人婚配只能淘汰 $100 \times 20 \text{亿} = 2000 \text{亿}$ 个新突变。这还将留下4000亿新突变作为新增基因负担传给下一代！每一位幸存者仍然比父母亲多带着100个突变。即使我们假设这些突变有三分之二是完全中性的，还有1333亿个有害的突变进入人群。我们无法除掉足够的突变而同时保持群体的数量！即使三分之二的突变是中性，而且让选择压力加倍（虽然我们肯定不能“耗费”每一代人口的三分之二），还将有667亿个新的有害突变留给下一代。选择的代价显然限制了我们的每代能够淘汰的突变数量，而人类的突变率已高到任何选择都无以抵消的程度。因此，突变会继续积累，而种族必将退化！你现在能否看到，在基因组水平上的选择，其代价确实是个令人头痛的问题？

2. 模糊或“看不见的”突变

致死和近乎致死的突变对于选择来说不是大问题，至少从整个群体的角度上来看是如此。这类突变很罕见而且会自我淘汰掉。同样地，绝对中性突变（如果存在的话）也无关紧要。短期内造成最大破坏的是次要（弱效应）突变（Kimura and Ohta, 1971, 第53页）。选择必须阻止次要突变的大量累积，不

然物种将很快衰退，适合度会下降。然而，即使选择能控制次要突变，它对于数量最多的一型，近中性突变，也显得无能为力。因此，不管有没有选择，高级基因组从长远来看终将退化。

2a) 近中性突变

单个近中性突变对于整个基因组的效应趋于无限小。在近中性核苷酸位点上的一切突变只会随机漂变（见词汇表“基因漂变”，*Genetic drift*），也就是说它们基本上不受选择的影响。它们的适合度效应非常微弱，以致于生物系统中最微小的随机波动（噪音）也能将其掩盖（Kimura, 1968, 1983; Kimura and Ohta, 1971）。它们是数量最多的突变，显示于图3d的“近中性方格”以内（第28页）。

既然大多数突变是近中性的，而且其效应都轻微到无法选择，那么这些突变有什么意义呢？它们的重要性就在于这些核苷酸位点含有信息，其突变最终将导致信息的腐蚀。综合起来，近中性核苷酸携带着基因组中大部分的信息。这就好像本书中所有看起来不起眼的字母，集合起来便形成清晰的信息。如果我们从一个非常长且复杂的书面信息入手（例如，一本大百科全书），开始制造印刷错误，大多数单个的错误对于整体信息来说影响甚微。单独来看，它们确实不重要。但是，如果不停止这个过程，整个信息终将遭到破坏，最后完全丧失。另外一个例子是一辆生锈的汽车。随着汽车老化，我们可以修理较大的问题，更换轮胎，修补撞痕（类似于选择作用

于主要的和次要的突变），但我们不能阻止一个原子接着一个原子地生锈（类似于近中性的突变）。每个氧化的铁原子看起来实在微不足道，但在整辆车上加起来，这个过程是严重的和无可挽救的。第三个例子是我们身体的衰老，我们可以补牙，整容，甚至心脏移植。但是一个一个细胞的老化（主要是因为突变）日积月累最终决定了我们的寿限，虽然每个细胞都微不足道，弃之无碍。正如人的身体由于无数极微小的错误（每个错误都无关大局）而不断地“朽坏”，人类基因组也必因近中性突变而“朽坏”，没有什么选择方案可以中止这个过程。这就是近中性突变问题的实质，这个问题使得一位杰出的群体遗传学家写了一篇题为“为什么我们还没有死过100次？”的文章（Kondrashov, 1995）。近中性突变之不可选择性确实是个问题。

一般来说，均一环境中的大而均一的群体（比如一个典型的细菌样本）比较能够抵抗基因漂变，因为噪音少而且选择有效。这样的群体通常有比较简单的基因组，每个基因组突变较少，而基因间作用更少。再者，群体中的个体数量庞大而且繁殖率高。最重要的是，由于细菌是单细胞生物，在每次细胞分裂的时候，每个细胞都会被独立地选择。在这样的系统中，选择过程更有效，更精确，而且分辨率高得多。这就是说在细菌基因组中，近中性和不可被选择的部分要小得多。这就是为什么理论学家们一般地偏爱使用微生物的例子。

不幸的是，像我们这样的哺乳动物没有以上所说的优势。我们受到高水平生殖噪音的影响。我们的基因组大，突变率高，基因间相互作用强，选择过程在我们身上严重地受限。这就是为什么人体中实际“不可选择”的突变比例应该是很大

的，人群中这类突变的频率完全受控于随机基因漂变。所有这样的核苷酸位点都将随机突变，而它们所编码的全部信息将随时间而退化。

2b) 次要突变和选择失效

根据定义，次要突变对繁殖能力有轻度但明显的影响。这就是在图3d（第29页）中近中性方格以左紧邻着的那些突变。虽然动植物育种专家很难看到这些轻微的变化，但是只要群体足够大，时间足够长，自然选择（其实不过是繁殖差异的另一种说法）通常能“看到”它们。再者，这类突变的效应有一定的相加性，所以自然选择能够同时选择多个次要突变。事实上，从自然选择的作用方式看，它好像是被设计出来用以稳定生物界的，不然生命会快速衰退。这似乎是个很奇妙地设计出来的系统。

然而，选择对于次要突变的作用有很大程度的局限。有序地淘汰次要突变的过程受到严重的噪音干扰。自然选择必须能看见这样的突变体，把它视为影响繁殖几率的因素。但“大自然母亲”监查次要突变常常有困难。这是由于次要突变所引起的繁殖适合度差异很轻微，而其他因素的影响可能很大。这有点像把小石子儿投进波涛汹涌的大海里，要看它所激起的涟漪一样。

所有其他影响繁殖的变量加起来，会严重干扰自然选择对任何特定次要突变的淘汰作用。例如，如果一个群体中的随机死亡率高，它将超过并掩盖次要突变的轻微效应。同样地，对

致死性和近乎致死性突变的优先选择，必然压过次要突变的微弱效应。多数突变为隐性这一事实，极大地掩盖了它们的负面适合度效应，严重地妨碍选择和淘汰。同样地，所有的基因间相互作用（“上位效应”）也将干扰选择性地排除次要突变。在小群体中，两性重组的随机性（染色体分离和配子结合都是随机的，因而也是波动的）常规性地压制选择。这些效应导致基因漂变这一基本现象。基因漂变已经被广泛地研究。众所周知，基因漂变在小群体中能压制除了最严重的突变以外的绝大多数突变的选择。植物育种者（像我自己）都知道，所有影响个体活力的外在因素都会干扰选择的有效性。群体中某一突变的频率常会随机漂移；每当所有其他因素综合起来对于繁殖机率的净效应超过核苷酸本身时，该突变将免于选择。

形象地说，选择对于很微弱的遗传效应的作用，就如在一个嘈杂的屋子里听细声细语。声音微弱、信息复杂、背景噪音强等因素都导致信息的流失。当一个次要突变的适合度效应还比较“大声”，且没有什么生物“噪音”的时候，选择最能发挥作用。尽管受到很大限制，选择对于某些次要突变还是起作用的。感谢上苍，选择在此还用得上，不然我们也不会在这儿。

虽然选择绝对可以作用于某些次要突变，但当这些次要突变数量增多时，每个突变基因的适合度效应对整体繁殖的可能性显得越来越不重要。随着次要突变数量的增多，及每个突变效应的减小，选择对各个突变的功效趋于零。木村指出了这个问题(Kimura, 1983)。我在附录2中用数学的方法论证了这个问题，结果显示在图6a和6b中（第78-79页）。每增加一个需要选

选择的性状，针对每个性状的最大选择压力就必减少。随着要选择的性状数量增加，选择对于每个性状的功效趋近于零，达到任何选择目标所需的时间将趋于无穷大。根据我的计算（参看附录2），像人类这样一个群体，可同时选择的突变最多约700个。木村提出同样的问题（Kimura, 1983, 第30页），虽然他没有给出具体的计算方法，但他指出即使对一个群体的总选择压力非常强（50%的淘汰率），而且对每个性状的选择系数很低（ $s = 0.01$ ），也只有138个位点可以同时被选择。试图同时淘汰多于几百个突变，必将导致选择进程的停止。然而即使在一个很小的人类群体中，每代都产生数百万个新突变，而且都必须被淘汰！在全人类，我们要选择的不是几百个突变，而是几十亿个。即使假设选择仅作用于几百个基因位点，这虽然在理论上是可能的，但值得指出的是，针对每个性状的选择压力被高度削弱了，选择的速率受到严重影响，使选择进程几乎成为原地踏步，需要几十万代的选择才能改造这些数量有限的核苷酸位点。

从某种意义上讲，当我们选择越多次要突变的时候，每个突变就成了其他突变的噪音。从远处听，一间充满低声细语的房间就是充满了噪音，毫无信息可言。随着每个突变的效应越来越弱，它的声音越来越低，要压过噪音就越来越难。即使是在最理想的选择状况下，每个微声发出的信息也不能被分辨。任何特定突变基因的效应，相对于其他繁殖因素来说变得毫无意义。当一个突变的生殖效应低于一个阈值（门槛）时，突变实际上变为中性，所有选择趋于停止。当我们需要选择越来越多的突变时，我们终将达到这个临界点，届时选择过程基

本上瓦解。当次要突变超过一个特定的数量时，即使在理想的低噪音条件下，也都会变得不可选择。也就是说，同时淘汰太多的次要突变必将导致选择停滞，被基因漂变取代。大体上，同时选择太多次要突变使它们都表现为近乎中性，就像前一节所述的一样。对此我已经用数值模拟的方法加以显示（见第11章）。

霍尔丹(Haldane, 1957)和木村(Kimura, 1983, 第26页)都认为由于选择代价的问题，同时选择大量性状是不可能的。这个简单的事实也使得任何方式的全基因组的选择都成为不可能。

3. 繁殖淘汰

我们已经知道，由于人类群体中存在大量的突变，而且淘汰每个突变的繁殖代价太高，我们不可能阻止基因组的退化。不仅如此，我们又探讨了由于大多数突变是近乎中性的，它们因效应不明显而被埋在生物噪音之中，选择对此无能为力，这也导致了基因组的漂变和退化。再者，我们也认识到如果同时选择太多的次要突变，它们都会表现为近乎中性，变得不可选择而进入随机漂变。最后，我要论证基因组的退化还有一个原因，就是在实际操作中我们不可能对多个突变基因同时实施有效地繁殖淘汰。我称该问题为选择干扰。这个问题还没有引起足够的重视，只是简单地被看作一个干扰选择的因素(Haldane, 1957; Lynch, Conery, and Burger, 1995; Kondrashov, 1995)。试图同时选择基因组中成千上万乃至数

百万个不同的突变，选择干扰绝对成为一个重大的问题。

对一个性状的选择会妨碍对另一个性状的选择，这就是选择干扰。例如，在同一个体中，理想性状与非理想性状经常同时存在。选择淘汰非理想性状意味着相关的理想性状会在无意之中自动地被选择淘汰（我们只能接受或淘汰整个人）。两个性状之间的联系也许很紧密（为同一个基因所编码或两个基因并排存在于同一染色体中），也许不太紧密（同一个体中两个基因不相干）。即使同一个体中的两个突变联系不紧密，对于这一代来说，那两个性状还是相连的（只能同时被选择或淘汰）。在每一代的每一个体中，任何突变基因肯定总是临时与所有其他好几个突变基因相联系着。所以，选择从来不能单独作用于某个突变。每选择一个有利突变总是会带入一系列与之关联的有害突变，这个问题是不可避免的。

为了解释这个问题，让我们想象一下在一个群体的两个个体中选择一个。因为两个个体的基因都是来自同一个“基因库”，所以他们体内的突变数量大致相等。虽然每一个体都有自己的一套突变，但两套突变有着几乎相同的总体有害效应。将这两个个体相比较（看哪一个适合度较高），我们可能发现每个个体有大约1万独特的突变，所以两个体之间有2万个基因差异。每个个体有约1万个较“坏”的基因单位（突变），也有约1万个较“好”的基因单位（非突变的核苷酸）。由于效应的平均化，两个体之间实际的基因适合度差异将很小。如有差异，那主要是由于少数强效应核苷酸改变的影响。哪一个将更有繁殖优势呢？因为整体基因适合度的高度相似，繁殖成功率将更多地取决于偶然机会和噪音因素，而不是真正的基因适合

度。即使“较好的”个体真有繁殖优势，选择过程也几乎没有进步，被选择的个体和被淘汰的个体有着同等数量的突变。我们淘汰了一个个体中的1万个突变，但同时给群体基因库加上了另一个体中的1万个突变。这样几乎所有选择都相互抵消了，前进1万步，又后退1万步。唯一的结果是淘汰了少数对适合度产生真正影响的‘主要’的变异。在高级基因组内，选择只对数量有限的“重要”核苷酸差异起作用。绝大多数突变是次要的或近中性的，它们在个体之间将互相抵消，而成为“不可选择的”。在基因组水平上，即使我们能彻底地控制个体的繁殖，还是无法有效地阻止绝大多数有害突变的传播。根据我本人的知识，这个问题虽然经常被群体遗传学家们提到，但还未得到足够的重视。

4. 连锁基因的选择干扰问题

当有利突变与有害突变在染色体上紧密连锁时，选择干扰最明显，也最严重。这会导致一个被称作穆勒的棘轮的不可调和的问题。对自然选择的一个最明显的要求是要能够分离好突变和坏突变。当有利突变与有害突变紧密连锁时，这种分割就是不可能的。基因组中几乎所有的序列都是以大连锁块的形式存在(Tishkoff and Verrelli, 2003)，所以这个问题实际上存在于基因组的各个部分。如果我们回过头来参看图3d（第29页），可发现突变绝大多数是有害的，但还应该有一些非常罕见的有利突变。这些罕见的有益突变似乎给正向进化留下一丝曙光，但这种希望并不理智，因为这些有利突变绝大部分是近

乎中性的（因而不可选择），而且不断累积的有害突变总是多于有利突变。然而对于乐观主义者来说，只要那些罕见有利突变还在那个图上，就好像还有希望。然而连锁的问题将那些有利突变从我们的图上抹去了，因为包含有益突变的连锁块也包含着许多有害突变，整个连锁块的净效应是负值（图7，第80页）。由于这个问题，正向进化再也不存在任何理性的指望。

在任何一个连锁单位中，平均来讲，在出现第一个有利突变之前应该已经有成千上万有害突变的累积。所以，在基因组中，任何时候都不会出现一个信息净增的连锁块。任何一个有利突变总是与大量的有害突变密不可分。这可以从图7中看出来。在图3d中，我们画出了单个突变效应的分布情况。我们也可以对连锁突变群簇的效应画出类似曲线。因为这些突变群簇从不分离，其遗传方式就好像是一个突变，其效应就是连锁群中所有突变相加的净效应。当每个连锁块中积累了至少两个突变时，几乎每个有利突变都被一个相连锁的有害突变抵消。这时，分布图中就已经没有信息净增的连锁块了（参看图7）。为了进一步解释这一点，假如一百万个突变中有一个是有利的，一对连锁的突变都有利的可能性就低到可以忽略（ 10^{-12} ）。随着时间的推移，每个连锁群簇内的平均突变数目增加，其净信息损失也增加，信息净增连锁块的完全消失很快成为百分之百的事实。人类基因组是由大约10万到20万个连锁块组成。根据以上推理，我们可以很有把握地讲，每一块“进化的砖瓦”都在不断地退化。

根据大量独立的证据，我们不得不认为人类基因组退化的问题确实存在。选择虽然对于延缓退化是重要的，却没有一种

选择形式可以阻止退化。就好像知道人人都会死一样，我不再留恋于这个问题。人类基因组的灭绝就象生物体要死去、恒星要燃尽、宇宙要热死亡一样必然无疑，势不可挡。

2008年补充：2006年有一篇论文表明，从理论上讲，在进化论的时间范围内，“穆勒的棘轮”在自然界的运转对人类是致命的。值得指出的是，该论文的作者只考虑了一个连锁单位，即线粒体的染色体。这只是人类基因组约20万个连锁群簇中的一个，所以该文作者将连锁的问题看轻了20万倍（Loewe, 2006）。

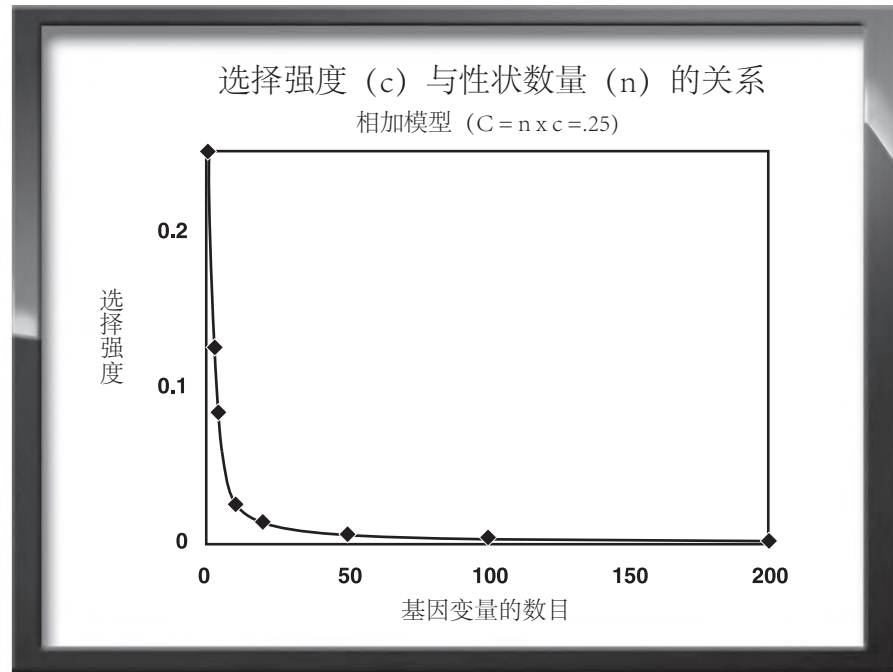


图6a 选择干扰

选择总是涉及繁殖代价 (C)，意思是有些个体不能繁殖。总的选择代价必须大大地少于一个生物种的过剩繁殖量，否则群体将快速萎缩并面临绝种。随着选择作用于更多的性状 (n)，花在每个性状上的代价必须相应减少，以保证总代价不超过该群体繁殖的潜力。为描述选择门槛的问题，我们假设一个群体允许损失25%的个体以排除突变 ($C = 0.25$)，上图显示作用于每个性状的最大选择强度 (c) 与被选择性状总量 (n) 的关系。可以看出，随着性状数量增多，作用于每个性状的选择压力迅速下降，趋近于零。当选择压力低于一个临界点时，性状实际上显为中性。这个曲线是根据一个相加性模型，依据公式： $C = n \times c$ (参看附录2)。附录2还显示，用相乘模型进行类似的分析会得出几乎同样的结果。

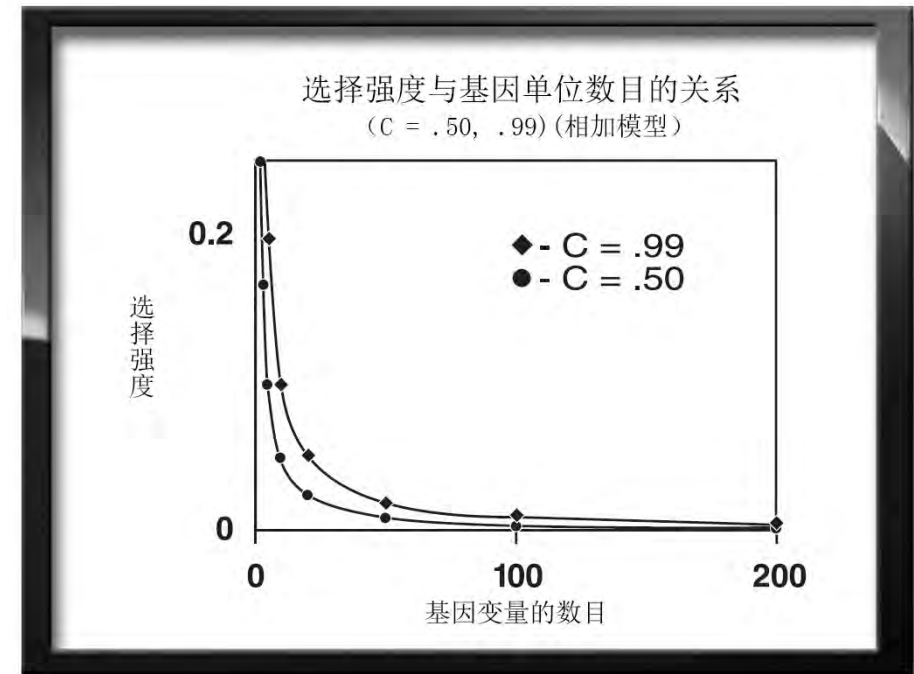


图6b 繁殖力极强的群体的选择门槛

假设一个群体的繁殖力极强，以致于每代可以淘汰50% ($C = 0.5$)，也不能有效地缓解图6a中的问题。即使繁殖力极强的物种，比如植物，C值可能达到0.99 (每株植物100个后代，选择淘汰99%)，选择门槛的问题还是存在。同时选择多个性状降低了对每个性状的选择效率，直到一个临界点，届时选择完全无效。这组曲线使用的是相加模型公式： $C = n \times c$ 。

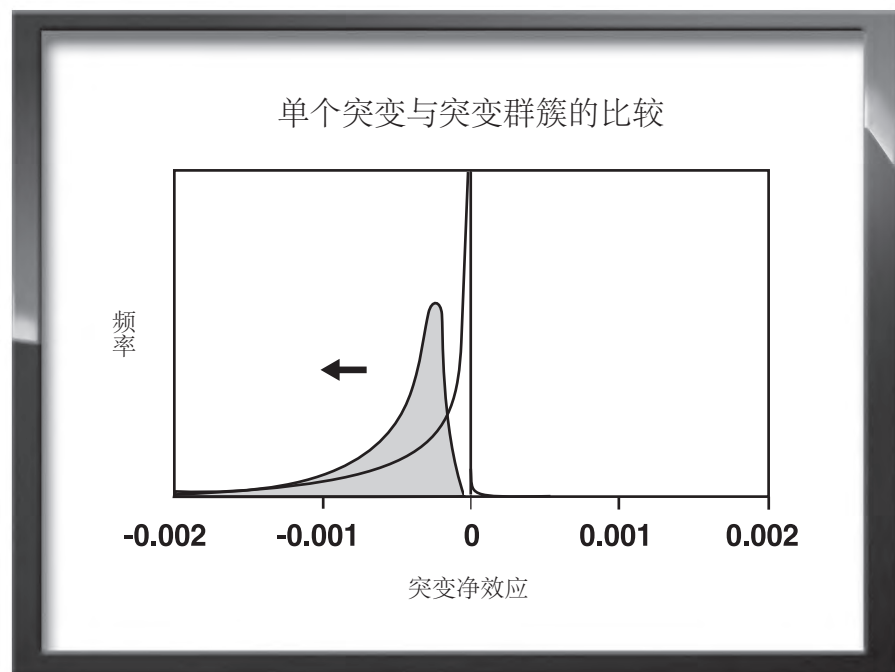


图7 突变出现于连锁群簇中

连锁突变群簇几乎100%是有害的。在图3d中，我们看到单个突变的分布趋于中性，虽然有很多有害突变，但也有极少数有利突变。当我们把突变放在连锁群中分析时，这些罕见突变的有利效应就消失了。与单个突变比较，连锁突变群簇的实际效应分布（阴影部分）将强烈左移。任何连锁突变群簇将作为一个基因单位被遗传，其效应就是它里面所有突变的净效应。任何群簇的适合度效应可计算为单个突变的平均效应乘以簇内突变数量。任何罕见有利突变势必被数量更多的有害突变抵消。因为绝大多数突变是有害的，一个突变群簇对于适合度的负面效应每一代都会增加。当每个连锁群只有两个突变时，几乎每个有利突变都被一个相连的有害突变抵消。随着每个群簇内突变数目代代增加，其净效应必然越来越有害（在适合度分布图中向左移动）。

atecg

第六章

进一步审视噪音

简讯：问题比你所想的要严重得多！

如果你想接收无线电信息，就必须控制“噪音”或静电干扰。如果存在一个很强的干扰信号，讯息就会丢失。正是由于这个原因，有些政府才可以“堵塞”那些他们不希望国民收听的无线电节目。同样地，假如政府不限定每一个电台所允许使用的波段范围，那么，各电台很快就会成为相互之间的噪音。静电干扰的来源有很多：其他电台、太阳热风波动、宇宙射线、被云层反射的无线电波、附近正在转动的电动机和当地的步话机等。无论来自哪里，噪音总会使信息丢失。

一个很弱的信号很容易就会被任何强度的噪音所破坏。信号越弱，噪音越强，信息损失就越肯定。低信噪比总会使信息丢失。当我们同时收到信号和噪音时，放大也无济于事。把你的收音机音量调大也不能克服静电干扰，因为增强信号的同时也同样程度地增强了静电干扰。为了确保把信息损失减到最低，就必须有一个适当的信噪比。

大部分核苷酸必然不可以被选择的原因，是一贯的低信噪比。同样地，不可以同时选择多个核苷酸也是由于信噪比随着被选择的核苷酸数目增加而迅速降低。实际上，选择不能拯救

基因组的一个主要原因正是由于无处不在的噪音。当我们让选择作用于整个基因组时，信噪比迅速趋近于零。因此，从整体上看，噪音一贯强于单个核苷酸的效应。这就是为什么选择在基因水平上有效，但在基因组水平上无法作用。

在遗传学上，信噪比通常用“遗传力”来表示。遗传力是基因型和表现型之间的重要区别点。如果一个生物性状有高遗传力，那么，大部分所观察到的在该性状上的变异都可以传给后代，对这个性状的选择也会容易。遗传力这一概念的本质，简单来说就是可遗传变异与不可遗传变异之间的比率。不可遗传的变异很大程度上是由一个生物体的特殊环境引起的变异，是表现型噪音的来源。所以，遗传学上一个性状的遗传力实际上等同于信噪比。例如，在两个人之间，任何可观察到的智能差异一部分是由于可遗传的基因差异（可以从父母传给子女的基因差别），而另一部分则是由于环境因素（如营养供给和教育质量等）。所以，有一部分智力取决于先天本性（经遗传而来的潜能），另一部分则取决于后天培育（非遗传的特性）。这也同样适用于身高、速度和体重等性状。从某种意义上来说，遗传力可被理解为先天遗传与后天培育之比率的反映。当一个性状的遗传力是“1”的时候，那个性状100%都是可遗传的（如血型），完全不受环境的影响。而当一个性状的遗传力是“0”的时候，那个性状是不可遗传的；它本质上完全是由环境所决定的（如纹身）。像身高这样的简单性状其遗传力通常较高（ $h^2 = 0.3$ ）。这就是说对于这样的一个性状，有30%的表现型变异是可遗传（可选择）的变异。但可惜的是，对于像适合度这样的相当复杂的性状来说，遗传力值却很低（可低达

0.004），甚至接近于零（Kimura, 1983, 第30-31页）。这是因为总体适合度综合了个体和环境方方面面不同类型的噪音。

近期的研究进一步表明，总体适合度的遗传力太低，在天然群体中一般地测不出（Merila and Shelton, 2000; Kruuk et al., 2000）。这些报告也指出，遗传力数值之所以低，主要是因为非遗传性噪音的干扰。在这之前，人们曾经想当然地认为，适合度的遗传力数值接近于零是由于自然选择极其高效，造成群体内的基因背景趋于一致，所以就没有可遗传的变异。这种单纯的想法今天已经不成立了——我们已经知道高等生命形式都有很高的突变率，而且选择过程无力消除各种各样的有害突变，所以基因差异还是相当大的。无数的分子生物学研究已经验证，天然群体中的基因一般地是高度多样化的。所以适合度的遗传力之所以总是低得无法检测，并不是由于缺乏基因变异，而是由于无处不在的生物学噪音。

当木村称适合度的遗传力通常都非常低的时候，他是讲个体间适合度的差异，几乎全部是由于非遗传的效应。因此，几乎所有对适合度的选择都是浪费。低遗传力意味着剔除不良表现型对于淘汰不良基因型的作用甚微。比如一棵树上掉下来的种子，有些种子会落在适合生长的肥沃土地里，但大部分种子会落在过干、过湿、杂草丛生、没有阳光、或是人口密集的地方。结果长出来的树苗的健康和活力多种多样，其存活和繁殖的机会也大不相同。但几乎所有这类“自然选择适者生存”，其实不过是选择最幸运者，而不是选择优良基因。因此，在大多数的自然状况下，适合度的表现型变异大部分只是源于非遗传性噪音，而不是由于任何可遗传的差别。这就是我们所说的

低遗传力。低遗传力在很大程度上缓冲了选择的效应。好像一辆汽车装有赛车的引擎，但传动装置坏掉了，可以有大量的选择，但基因并没有什么进化。

让我们进一步探讨适合度的非遗传性变异是由哪些因素造成的，参考图8a（第91页）。表现型变异的最重要来源是不同个体所处的不同环境。很明显，环境的差异可造成不同个体之间的重大差别。一般估计，大部分的表现型变异（超过50%）仅仅是由于环境的差异。如果同一块田里一株植物长得比另一株好，很可能仅仅是因为它所在的那一小片土壤稍微好一点。这方面的表现型变异显示在图8a的区域1里。这类变异增加了非遗传性噪音，降低了信噪比率，因而强烈地干扰着有效选择。

表现型变异的第二大部分被称为“环境与基因型相互作用”，一般估计占全部表现型变异的25%。考虑同一块田地中的两株植物，如果我们通过浇灌土壤来改变植物的生存环境，增加的水分可能对其中一株植物是有益的，但对另一株不利。这取决于它们不同的基因型。这类变异并不总是可以遗传给后代的。正像环境变异一样，这方面的表现型变异增加了噪音，因而妨碍选择，见饼图区域2。

表现型变异的第三大部分是非遗传性基因变异。这听起来来自相矛盾，其实不然。大部分基因变异是源于不能持续地代代相传的因素，这包括表观遗传效应（区域3）、上位效应（区域4）、显性效应（区域5）和循环选择的基因效应（区域6）。简而言之，大部分的基因变异都不可遗传，至少不能直线遗传，因而没有选择效果。

基因变异中唯一可遗传的（因而有选择潜能的）部分被称为相加性基因变异（区域7）。很简单地讲，相加性基因变异是指群体中的某一个基因（或核苷酸）明确地且持续地优于决定同一性状的其它基因（核苷酸）。对于适合度这样一个复杂的性状来说，这类相加性基因变异在全部表现型变异中仅占很小的一部分。如果木村所认定的适合度遗传力可低至0.004是正确的，那么在表现型变异中仅有0.4%是可选择的。这代表了约为1:250的信噪比。换句话说：有99.6%的对适合度表现型的选择将会完完全全被浪费掉。这说明了为什么对总体表型适合度的简单选择，基本上不会带来基因改进。

如果我们有一个低遗传力的性状，如适合度（图8a），又有一个低生育力的物种，如人类（图8b，第85页），我们可以看到群体中只有极小部分可用于有效的选择（图8c，第86页）。假如我们只能排除约16.7%的群体，而其中只有0.4%是实际上有效的选择，那么群体中只有0.07%可真正用于有效的选择淘汰。换句话说，1000人中不到1人可用于有效地排除所有的有害突变，并用于有效地固定任何可能发生的有利突变。

像总体适合度这样一个性状的遗传力可能会非常小，然而一个典型核苷酸的“遗传力”更是无穷小。让我们考虑一下一个普通的单核苷酸突变的遗传力，这不是遗传力概念的正统用法，但有助于说明问题。这种“信号”（就是这个核苷酸可遗传的相加性适合度价值）本质上微弱到不可测量的程度，但“噪音”却是庞大的。那是所有的非遗传性变异因素的综合效应，再加上所有其它多样性核苷酸位点的效应！在一个典型的人群中有数百万个多样性核苷酸。由于信噪比率总是低于

1:1000000, 一个普通的核苷酸的效应必将消失在噪音的海洋中。这样一个核苷酸的遗传力几乎为零, 这就是为什么大多数核苷酸实质上是不可选择的, 根据木村和太田的定义都必属于近乎中性(参见第2章)。

噪音的另一个主要来源是自然选择的方式, 即**概率选择**, 而不是**斩截性选择**。作为一个培育植物的人, 我常按照表现型(产量、存活力和抗病能力等等)给数百株植物打分, 然后将它们从优到劣排序。我会决定要淘汰群体中多大的百分比, 在想要的水平上划一条界线, 只保留界线以上的那些植株。这是一种人工选择, 称为“**斩截性选择**”, 因其特别有效, 常常被育种者所用。然而, 这种类型的选择在自然界永远不会发生。自然选择总是基于繁殖机率的差别。大自然并不会综合所有性状的表现型, 并据此为群体中的每一个成员计算一个“**适合度总值**”, 也不会将所有个体排序。它更不会随意作出一条线并剔除所有在界线以下的个体。实际情况是, 表现型劣等的个体只是较其它个体繁殖概率稍低而已。十分常见的是, 劣等个体也有机会继续繁殖, 优势个体反而没有机会。事实上, 表现型优势与繁殖成功率之间的相关系数并不很高。繁殖噪音越多(即随机生存/交配), 相关性就越弱, 优势个体的实际繁殖优势就越不确定。这就是自然选择真实的运作方式, 称为“**概率选择**”, 很模糊, 也很低效。如果我们在进化模型中使用符合现实的概率选择, 而排除各类斩截性选择方式, 结果将是增加了新的一层噪音, 从而进一步降低了选择的功效。

真正的自然选择(概率选择)总的来说是疲软无力的, 我可以信手举几个例子。想像有一大群虾正在海里结队而游, 一头鲸

过来一口吞下虾群的一半。难道这是一个虾类适者生存的例子吗? 有没有一个从最适合到最不适合的精确排名, 然后通过一种严格的选择进行一刀切? 再想像一条河里有一大堆蛙卵。在孵化以前很多卵就已被鱼给吃掉了。大量的蝌蚪被飞鸟啄食, 一只小船驶过又撞死几百条, 还有更多的被瀑布冲走。很多发育中的成蛙钻进泥岸里, 但很快就被航道疏浚作业清理了, 剩下的成蛙又大多被其他动物所捕食。几乎所有这些淘汰都是随机性的。我们又一次看到最幸运者生存。所以哪里有系统的表现型分类? 基本上没有! 不仅在决定谁有最佳表现型的过程中有大量的噪音, 在差异化生存和繁殖的层面上也是一样。无论什么表现型, 噪音都会影响繁殖成功率, 而这层噪音是在上文论及遗传力时所考虑的噪音之上。可能有50%的繁殖失败与表现型无关(图 8b, 第85页)。

还有**第三层遗传噪音**, 称作“**配子取样**”(在精子或卵子形成过程中的**随机基因抽样**)。这是一个与小群体相关的统计学波动, 是随机基因漂变的原因。假如你多次抛掷一个硬币, 你可预见有50%的机会是正面朝上, 50%反面朝上。然而, 如果你只投掷10次, 很可能正反面不各占50%, 你可以用概率表很容易地查到各种可能的正反面组合出现的概率。同样的统计学变化也会在群体中的某个基因或核苷酸分离的过程中出现(在配子形成过程中成对的基因随机分离, 进入不同的配子)。基因分离在十分庞大的群体里很恒定可测。但是, 在较小的群体里就会发生明显的波动, 就如反复投掷硬币一定次数的结果一样。这样的统计学波动就导致了基因漂变。这意味着不管有没有选择, 基因频率总是会改变。基因频率波动的这种单纯概率

因素已被研究得很透彻。

经典的群体遗传学家研究遗传噪音的时候，其注意力仅限于最后这一个噪音层次上，也就是配子取样（概率波动），这类噪音对群体大小很敏感。在小群体中随机基因漂变很强，甚至可以压过重大突变所造成的影响。这就是为什么濒危物种的群体特别容易发生“突变熔断”。在小群体中，自然选择基本上已经停止。木村首次对近中性突变作出定义的时候，正是基于配子取样这个问题上的。也正是由于同样的原因，木村的非选择区域（图3d，第29页）的大小是群体大小的一个简单函数（选择系数在正负 $1/2N_e$ 之间）。把噪音仅限于配子取样问题的做法对遗传理论家们很有吸引力，因为只要想象着一个大的群体就能很方便地让配子取样的噪音消失。同时，理论家们也能借助于小群体瓶颈时期的高噪音，让选择周期性地“消失”（例如人类起源的“走出非洲理论”）。

配子取样仅仅是总遗传噪音的一小部分，遗传噪音的另外两个重要层面（影响表现型的噪音和影响繁殖成功率的噪音）在大群体中只是部分地被消除。我们不能借助大群体来让噪音“消失”。噪音始终存在，且其水平比群体遗传学家一般所承认的要高。实际上，非常大的群体总是会增加噪音。其原因一部分是由群体分支结构引起的（许多小规模亚群，每一个都有自身的配子取样波动）。另一部分原因是大群体跨越了更广的环境范围，更易于发生环境变异。大群体并不能减少影响繁殖的随机因素，也不能减少表现型对基因型的遮蔽。噪音总是会对自然选择产生严重的约束作用。在人工条件下，动植物培育者可以很成功地选择有限数量的性状，他们靠的是智慧设计来有意识地将噪音降到

最低。他们使用了屏蔽技术、复制、统计分析、斩截性选择和高度受控的环境。但自然选择却一样都不会。根据定义，自然选择是一个盲目和没有控制的过程，受制于各种噪音及无限的随机波动。

所有这些噪音会产生怎样的遗传学后果？当我们意识到高水平的噪音是不可避免的，我们就会明白图3d(第29页)中的非选择方框是无法消除的，即使借助于大群体也做不到。当我们开始理解在配子取样之外还有两层完全不同的噪音时，我们便意识到这个非选择方框需要修正（见图9，第94页）。由于基因型和表现型的不完全相关，表现型和繁殖成功率的不完全相关，木村的非选择方框必须大幅度拓宽。这两个不完全相关的结果就是近中性（不可选择的）核苷酸位点的数量远远多于我们通常所认识到的，而且该数目并不严格地依照群体的大小而变化。小群体当然会加重噪音问题，但大群体也不能避免这个问题（更详细的讨论见附录5）。

无处不在的严重遗传噪音，强化了我在前面几章中关于选择局限性的所有论述。由于大量的次要突变在噪音的干扰下变为接近中性的不可选择的突变，木村的非选择方框被扩大了（图9）。选择门槛问题，也就是同时选择过多性状时导致选择进程完全停止的问题，会因高度的噪音而提早出现。最后，噪音问题会导致大部分的“选择资金”被彻底浪费（见图8a-c，第91-93页）。这大大增加了选择的实际代价，因而严重加剧了“成本”问题为任何选择方案造成的真实局限。

为了取得对噪音问题更直觉的了解，我们可以回到一些能见的比喻上来。就红色小拉车的演变来说，试想象一种情况，

如果小拉车的性能，更多地取决于生产线上的装配误差，而不是装配说明书中的拼写错误，你是否看到质量控制员把大部分时间和资源都浪费在选择“非遗传性”变异上？对豌豆公主而言，噪音问题就如她的床垫子里面有数不清的豌豆大小的疙瘩。这会不会使她的问题更困难了呢？在我们的生化教科书比喻里，噪音问题就如学生在期中考试后互换了教科书。这将会使课本中的拼写错误和学生成绩之间本来就勉强的微弱相关性变得更低。在所有这些例子中，当我们加入任何合理水平的噪音时，这些原本已经很荒诞的选择方案就变得更不可能实现。

和木村一样，已故的进化生物学家史蒂芬·翟·葛尔德 (Stephen Jay Gould) 也反对用严格的选择主义理论来解释进化。对于整个物种的生存，葛尔德认识到自然灾害、“幸运者生存”和噪音的重要性。但葛尔德和木村似乎都没有认识到，如果噪音常规地压制选择，长期的进化就是不可能的，而是必定出现基因的退化以至最后的灭绝。

2014年补充：自本书初版之后，我和同事们更细致地研究了噪音对于自然选择效率的限制作用。我们使用电脑程序《孟德尔的会计师》做了大量的数值模拟。结果表明，现实水平的生物学噪音极大地扩展了木村的非选择区。不管对有害突变还是有利突变，非选择区的宽度都要增加大约1000倍。如果群体遗传学家不考虑生物学噪音的深远影响，便是自欺欺人。见Gibson et al. (2013)，及 Sanford et al. (2013)。

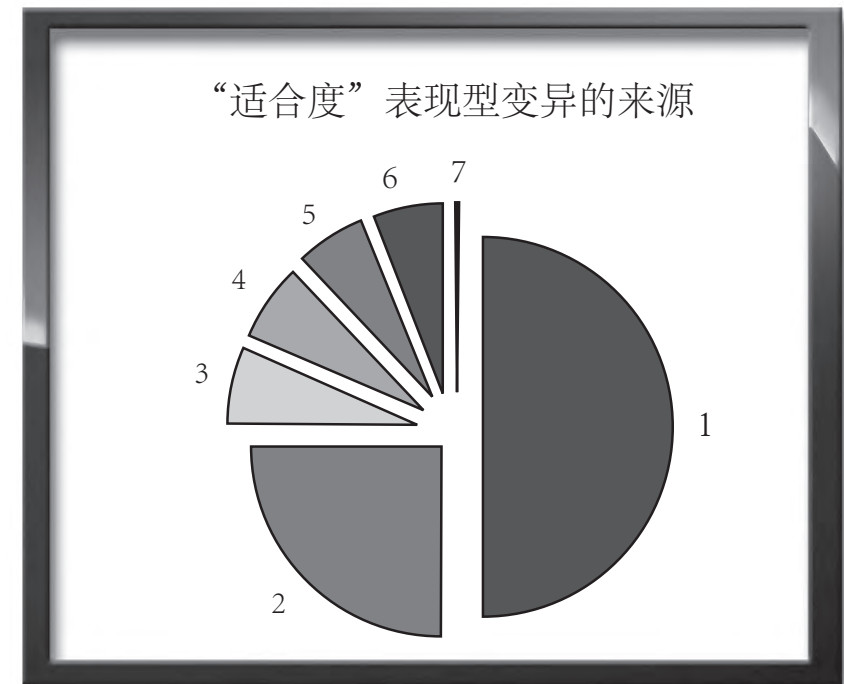


图 8a: 表现型变异的来源

表现型变异的主要来源就是环境变异（区域1）。这种变异是不可遗传的，并会对选择造成干扰。位居其次的变异源是环境与基因型的相互作用（2）。这种变异也不可遗传，并且干扰选择。在变异的基因因素里，有些变异是由于：表观遗传（3）、上位效应（4）和显性效应（5）。在这些基因因素当中，没有一个是可遗传的，而且它们都会对真正的长期选择造成干扰。最后，还有一些其它的基因因素也许是可选择的，但被内环境稳定或循环选择等过程“中和”掉了（6）。所有这些不可遗传的因素是绝大多数表现型变异的来源。这使得相加性基因变异成为表现型变异的一个无关紧要的部份（7）。对于像繁殖适合度这样一个很不特异的表现型性状来说，相加性基因因素的效果占全部表现型变异的不到1% (Kimura, 1983, 第30-31页)。换言之，基于表现型优势的选择性淘汰，其中99%以上是被浪费掉的。实际上，相加性基因变异以外的所有变异都是“噪音”，都强力干扰着有效选择。

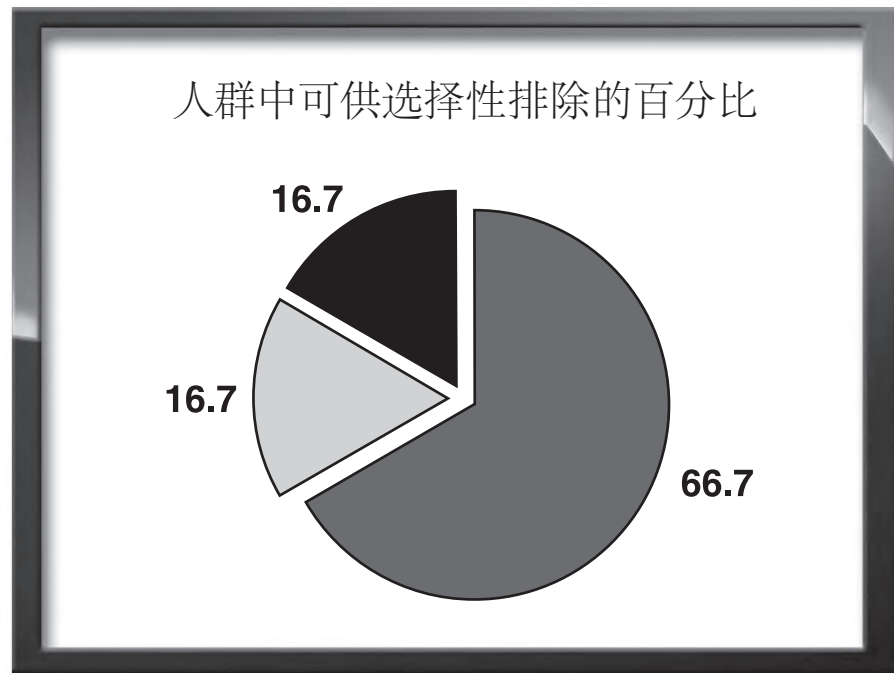


图8b 有多少人可供选择淘汰?

任何群体中只有一定的比例（群体盈余）可供选择性地排除。考虑到目前的全球人口生殖率（每对夫妇三个孩子），必须有三分之二的孩子继续繁殖以维持人口。这最多留下三分之一的孩子可以选择性地不繁殖。然而，有相当一部分（也许一半）的过剩人口不会生育，其原因完全是偶然的，如战争、事故和个人选择，与基因表现型无关。所以实际上只有约六分之一（16.7%）的人口有可能被用于任何形式的选择排除。这印证了霍尔丹的估计，他认为只有约10%的人口可真正用于选择。

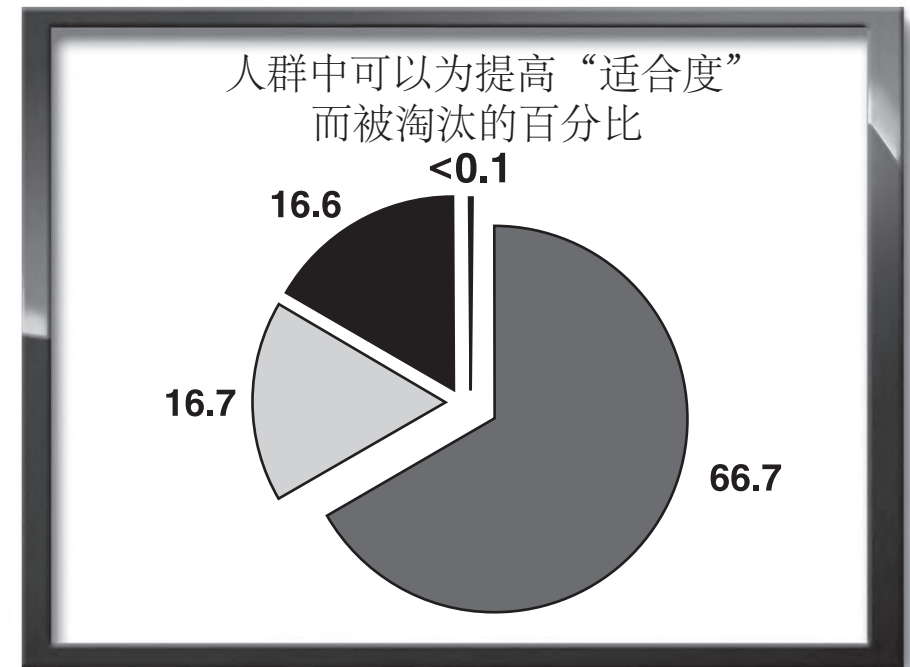


图8c 选择淘汰掉的人中有多少人的基因并不差?

人群中基于表型适合度（一个很不特异的性状）的选择淘汰，其有效性接近于零，将图8a和图8b合并就可看出这一点。人群中从理论上可以因为表现型低下而被选择排除的16.7%之中，只有0.4%的选择会产生可以代代遗传的效果。这意味着全部人口中少于0.1%（ $16.7\% \times 0.4\% = 0.07\%$ ）被用于有效选择。这个百分比低到在这种图上无法精确地显示。换句话说，一千人中不到一个可真正用于有效的选择性的繁殖排除。即使有一个像希特勒这样的人在每一代人中尽可能地“杀掉”表现型“低下”的人员，也不能明显地改进像“适合度”这样一个不特异的性状。这一结论是人类的低生育率和适合度这个复杂性状的低遗传力这两个事实的逻辑必然（Kimura, 1983, 第30-31页）。

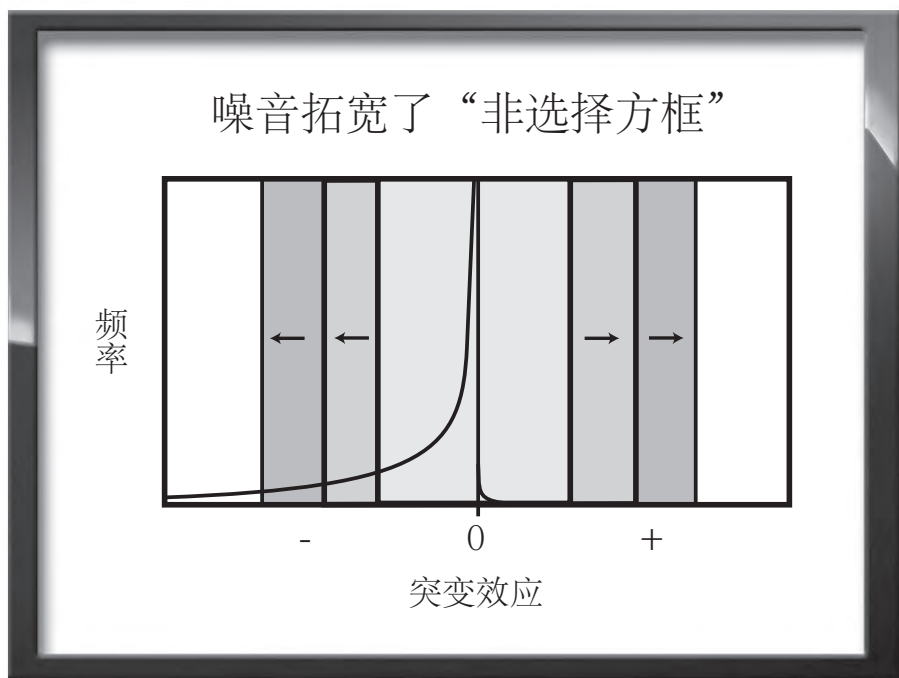


图9 现实水平的生物学噪音极大地扩展了木村的非选择区

木村的非选择方框，如图3d（第29页）所示，是基于对噪音的最低估计（仅包括配子取样的噪音）。这是经典的近中性模型，但这种观点并没有考虑到噪音的全部来源。所以木村的经典非选择方框是太小了。我们需要拓宽这个非选择方框，第一是因为低遗传力，这是源于基因型和表现型的不完全相关（99%以上的表现型变异是不可遗传的，见图8a）。然后，我们需要进一步拓宽这个非选择方框，以容纳因繁殖的随机因素而造成的表型优势与繁殖成功率之间的不完全相关（图8b）。这些都是噪音的主要来源。当我们考虑了噪音的全部来源之后，我们意识到真正的“非选择方框”很大，而且不可以仅靠想像中的大群体就把它消除。

atcgtttagc

第七章

假想的挽救机制

简讯：假想的挽救机制经不住现实考验

基因组中的净信息量究竟是在增加还是在减少？要解释自然选择如何能制造出新信息，人们最多只能说出几个莫须有的机制。这在本质上就是猜测。在某种意义上，这成了一个哲学命题，很难进行科学的分析。你可以做出强有力的论证，说明选择加突变不能产生新信息，但理论家们总是能设想出相反的情况（要证明某件事从来不会发生是很难的）。然而，我相信基因组的“下行”趋势实际上是可以科学地分析的。正是由于这个原因，我一直着眼于信息的破坏，我相信这一点可以严格地分析。这样的分析有力地证明净信息一定在减少。如果是这样，即使有些特别的例子表明通过突变加选择可能合成新信息，那还是毫无意义，因为新信息马上又开始退化了。大方向仍然是向下的，复杂的基因组从来不会自发地形成。

如果基因组确实在退化，这对人类的远期未来是个坏消息，对于进化理论也是坏消息。如果突变加选择不能维持已经存在于基因组内的信息，便很难想象它如何在最初创造了这些信息！如果每一代都有信息的净损失，我们有何理由讲基因组的建造！在可以理性地讨论信息建造这个更困难的问题之前，制止退化只是一个小小的前提（参看第九章）。

在过去几十年中，随着人们渐渐认识到有害突变率远远超过每代每人一个突变，遗传学界对突变率的担忧日甚一日 (Neel et al., 1986)。理论家们为了回避这一问题，一种做法就是宣称大多数DNA实际上不含信息，所以大多数突变的效应是完全中性的。据此逻辑，如果核苷酸置换的实际频率是每人10个，而98%的基因组被定义为垃圾DNA，有效的突变率必然是每人0.2。所以通常在媒体和教科书中引用的数目是0.2，而非10。然而，随着已知的突变率明显超过了每代每人10个，而且已知的功能性DNA百分比也大大超过了2%，再也不能用这种解释来抹掉基因退化的问题了。届此，有人为了挽救新达尔文主义理论，又发明了两大“逃避机制”。第一种机制可称为“突变计数机制”，第二种可称为“协同上位机制”。

突变计数机制

这种论证是穆勒 (Muller) 首先提出来的，为克劳 (James Crow) 博士所推广。克劳博士承认，在任何群体中，当有害突变率达到每个个体一个突变时，突变必然开始累积，群体适合度必然下降。然而，当每人累积的突变总数很大时，他意识到有些个体携带的突变数要比其他个体高许多（由于偶然因素）。他提出若选择过程能重点淘汰这些个体，就可以扫除大部分的突变。结果是花费较小的群体“代价”就能排除较多的突变（见附录1）。最终，每人携带的突变数量也许能稳定下来，适合度降低的问题也会逐渐得到控制。克劳提出的这种机制可称为“突变计数机制”，因为它要求大自然为所有的个体进行突

变计数，然后针对突变数高的个体着力选择。

如前所述，我们都是多重突变体，所以选择必然是在“重度突变体”与“轻度突变体”之间。有两种方法可以判断谁的突变严重。第一，不太紧要的问题：“谁的突变最多？”第二，要紧的问题：“谁的突变最糟？”克劳所描述的模型只考虑了前者，而忽略了后者。克劳的模型是一个脱离现实的想像架构，其目的就是要淡化一个重大理论难题。

因为我们每一个人的基因多多少少代表了人类“基因库”的一个随机样本，所以我们每个人都应该有差不多同样数目的累积突变。两个人之间在突变总数上的实际差异是很小的，仅代表抽样误差（就是与群体平均水平的微小差异）。这是因为突变已经经过了许多世代的积累和在“基因库”中的混合。突变的数量已经变得很大，突变计数也应该是相当均一的。所以就个体之间的适合度差异而言，“突变计数”是一个很次要的变量。

我们所见到的人与人之间的适合度差异，并不是源于每个人的突变数目差异。其实，明显的适合度改变都是由于少数强影响基因突变的特别效应。只需要一个主要突变就可以遮蔽千万个近中性突变的效应。一个人可能比另一个人少几千个突变，然而仅因为一个特别的突变就能够使那个人的适合度比别人差许多。所以，通过衡量每个人的突变总数来淘汰突变较多的个体的想法，不是排除突变的合理机制，也不符合现实。这个概念似乎是一个数学把戏，其发明是为了合理解释如何以较少的选择代价排除较多的突变。虽然这个过程可能在自然界中有着极有限的运作，但很清楚这不是普遍情况。

数值模拟实验已经有力地证明，突变计数机制是无济于事的(Brewer et al., 2013a)。只要参数设置略微有点靠近生物学现实，每个个体中的突变计数便随着时间直线增加，适合度也持续下降(图10a)。

协同上位机制

第二种挽救机制是“协同上位机制”。该机制与突变计数机制类似，那就是人为地让自然选择聚焦于突变计数高的个体上。然而这一机制又多了一层复杂性，它假定随着有害突变的积累，各个突变的有害效应会相互强化(这就是“协同上位效应”的含义)。从逻辑上讲，这应该加速基因的退化，但是有人想像，随着突变对基因组的损害加剧，选择或许会变得更加有效，更能针对性地排除突变计数高的个体。协同上位机制从各个方面看都像故意地混淆视听，花哨的术语经常被用来掩盖内在的问题。协同上位效应指的是突变之间相互作用，导致几个突变的集体破坏力比根据单个突变的效应计算出的总效应更严重。这样的突变相互作用不是不会发生，但这是例外——所有的群体遗传学家都同意，突变相互作用通常不是相加就是相乘。把协同上位效应认作基因组水平上突变相互作用的主要模式，这是毫无根据的。至少有一篇论文提出了实验证据，反对协同上位效应广泛存在的说法(Elena and Lenski, 1997)。如果真有大量的协同上位效应存在，基因的情况会更糟，而非更好。我们一直知道基因单位会相互作用，而且我们知道此类上位效应对于有效选择是个巨大的障碍。这一事实一直为多数遗

传学家所忽略，因为如果不把这类相互作用很方便地旁置，种种选择方案就会复杂到令人绝望，变得一概不可行。但到如今，当基因间相互作用可以用来掩盖误差灾难问题的时候，这个概念又被方便地换一个版本拿出来，以一种极散乱模糊的方式使用，就像一道烟幕。但让我们穿过烟雾看事实，如果多重突变果然以非线性和迅速升级的方式造成破坏，误差灾难发生的时间就会更早，群体随即失控，更快地进入突变熔消。我们早就灭绝了！我们看到的是一个玩弄概念的“魔术手法”，明显地不可运用于现实世界中，其目的仅仅是为了支撑基本公理。

协同上位机制已经被数值模拟实验正式地证伪(Baumgardner et al., 2013)。实验显示，即使存在着很强的协同上位效应，它也不能阻止有害突变的积累。相反，它会加速基因退化和物种灭绝——正如逻辑分析所提示的那样(见图10b)。

理论家们提出这两种主要的逃避机制以挽救基本公理，二者都是人为的、假想的，而且在生物学上是极不现实的。如今二者都已经被有力地否定了(Brewer et al., 2013a; Baumgardner et al., 2013)。逻辑分析和细致的数值模拟都显示大多数有害突变必然会逃过净化选择。这就导致突变计数随时间呈线性累积，适合度也持续地下降。这一点已经通过符合生物学现实的数值模拟得到了反复的验证(如 Gibson et al., 2013)。类似地，大多数有利突变，除了极个别的强影响突变之外，也会逃过正选择(Sanford et al., 2013)。虽然某些孤立的强影响有利突变会带来对环境的阶段性适应，但在这些个别事件发生的同

时，其余的基因组正在系统性地退化。

至此，我们应该清楚地看到基本公理并非“内在正确”，也并非对于所有明白事理的人“显然如此”，所以我们不应再把它看作公理。剩下的不过是“基本假说”（突变加选择能够创造和维持基因组），而“基本假说”也已经证明是站不住脚的。事实上，多重证据表明“基本假说”显然是错误的，也必须抛弃。

2014年补充：最近有三篇论文宣称已经解决了高突变率的问题 (Lesecque et al., 2012; Charlesworth, 2013; Qui et al., 2014)。以下我要一一显示，这些宣称都是不足信的。不仅如此，这三篇文章无一涉及近中性突变这一更深层的问题。这类突变因适合度效应低于选择门槛而发生无限制的累积。即使这些作者真能解决高突变率的问题，近中性突变的问题仍会造成基因熵现象。

Qui等的文章使用的是数值模拟，与我们的方法很相似。然而，若不设置符合现实的参数，数值模拟就不会得出符合现实的结果。他们的分析假定了人工斩截性选择、100%的遗传力及零噪音，这种设置高度歪曲了生物学的现实。根据我们自己的分析，这套参数得出的选择门槛比天然的概率选择低了大约1000倍。除了这些不现实的参数之外，他们还假定了极不现实的有利突变率（10%）。结果是，在严格的斩截性选择排除了大部分有害突变之后，剩下的有害突变虽然在积累，但是有利突变经选择放大之后足以与之抗衡。如果他们回过头去检查一下他们实验中的突变计数，我相信有害突变仍然大大地多于有利突变。如果我为了把突变的问题抹杀而去做扭曲生物学现实的数值模拟，我也会用他们这一套参数，但是这些参数根本不反映现实。坦率地讲，这些作者需要回去用符合现实的参数（比如概率选择和低得多的有利突变率）重做这一套实验。那样就会得出与我们的模拟近似的结果，证明基因熵的真实性。

Charlesworth和Lesecque等的两篇文章在开篇都承认群体遗传学家确实面对着一个重大的理论问题，即有害突变的积累。然后他们着手解决这

个问题。不幸的是，他们都把问题限定在传统的与现实相去甚远的“遗传负荷”运算背景中，都需要一长串脱离现实的简单化的假设。两篇文章都遵循一个长期的传统，即把复杂的群体动力学硬套进几个过分简单化的几何公式中。他们的论文实际上并不解释选择淘汰的速度如何跟得上突变蓄积的速度。相反，他们绕过首要问题（基因破坏的累积），而故意模糊适合度、突变效应和选择的定义。如此一来，基因熵的问题似乎消失在一片模棱两可的云雾中了。

两篇文章都使用一个脱离生物学现实的概念，叫做“软选择”（称之为“软突变”更贴切），这个词是大约25年前Wallace首先提出来的。当时Wallace想解决“遗传负荷”的问题（这个概念与基因熵相似，只是比较狭隘），于是设计出一个极不现实的模型。在他虚构的世界里，群体的适合度是一定的（只受可利用的生态资源的影响），有害突变并不造成客观上的功能丧失，而只影响在目前群体中的相对竞争力（比如性别选择）。这两个基本假定（适合度是恒定的，突变效应只是相对的）是根本错误的。Wallace 25年前在这个幻想的基础上构建了他的理论，今天这些作者们又在同样的基础上开始了他们的运算。

为理解其结论的不现实性，“软选择”假定：1. 群体适合度是恒定的，不受突变蓄积的影响；2. 所有的突变都是“相对性”的（对整体生物功能来说是中性的），并不带来任何客观上的基因破坏；3. 所有的突变只影响社会地位（比如性别选择）。Lesecque等进一步阐释了他们的假设：A. 个体的突变数是静止的，可预测的（U/s）；B. 等位基因的平均频率是静止的，可预测的（u/s）；C. 所有的突变都是完全可选择的，其适合度效应是同等的（s）。他们还假定没有连锁，突变之间的相互作用只有相乘关系，这一类模型与生物界的现实没有任何关联。

软选择模型在Wallace一发明的时候就是一道烟幕，今天的这些作者仍然把它当作一道烟幕。“软选择”把人们的注意力从日积月累的基因破坏这一活生生的现实中转移开来。在个别情况下，这一类软选择在某些孤立的性状上确实会起作用。比如在喜欢金发的文化中，金发女人可能比乌发女人留下更多的后代。但在这个例子里，影响发色的突变累积显然不会导致物种的灭绝。个别的软选择事例与全基因组退化的问题无关，与净化选择强度不足

的问题无关。

这些作者必须明白，要制止有害突变的累积，一定要有“硬”选择（真选择）。他们使用抽象的数学创造一个虚拟世界，在那里根本不会有基因退化。他们在玩弄概念魔术，不是要解决问题，而是要掩盖问题。他们似乎已经下定决心，要不惜一切代价来挽救基本公理。从最近这些罔顾现实的模型可以看出，这个领域在空抓救命的稻草。

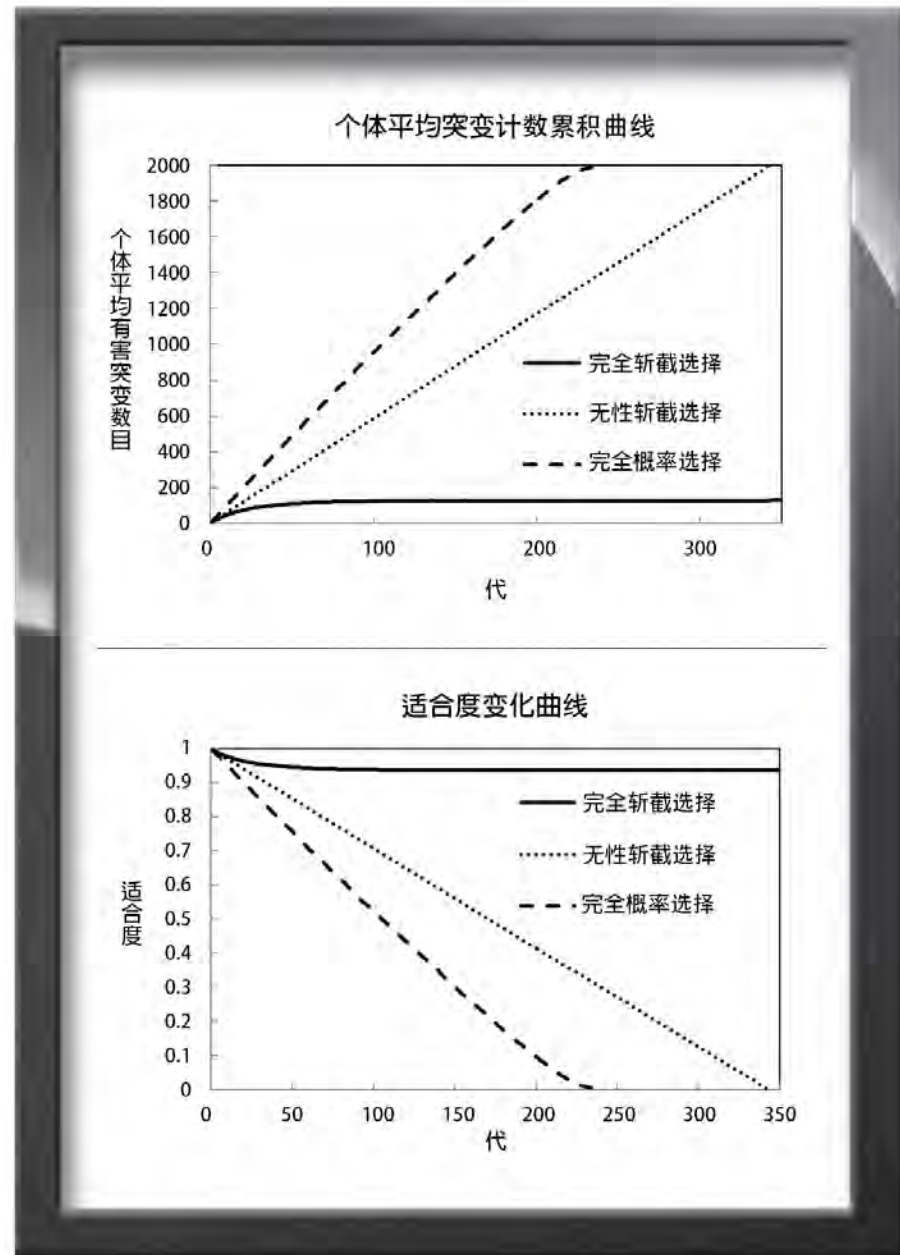


图10a 数值模拟实验检验“突变计数机制”的真实性

引自Brewer et al., 2013a, 显示三种不同的突变计数机制中个体平均突变计数的积累(上部)和适合度的变化(下部)。表型适合度仅取决于突变计数,所有的突变都有同样的效应(-0.001),无环境噪音。选择模式为:A.完全斩截性选择;B.完全概率选择;C.无两性重组的斩截性选择。如果采用符合现实的概率选择(间断线),突变会恒速积累,适合度持续下降,直至灭绝。在无性繁殖的群体中使用人工斩截性选择也会出现相同的情况。但是在有性繁殖中使用斩截性选择,各个突变的适合度效应又完全一致,突变计数和适合度会通过突变计数机制很快地稳定下来。突变计数机制会发生效力,但其前提条件极不现实。任何程度的生物学现实性都让突变计数机制失效,持续的基因退化在所难免。以下情况都会破坏这一高度假想的机制:A.天然的突变效应分布(包括近中性突变);B.概率选择;C.无性繁殖。概率选择发生于自然界,而斩截性选择发生于人工培育过程中。这一切都证明突变计数机制在现实情况下不发挥作用,而只可能发生于严格的人工控制之下(Brewer et al., 2013a)。

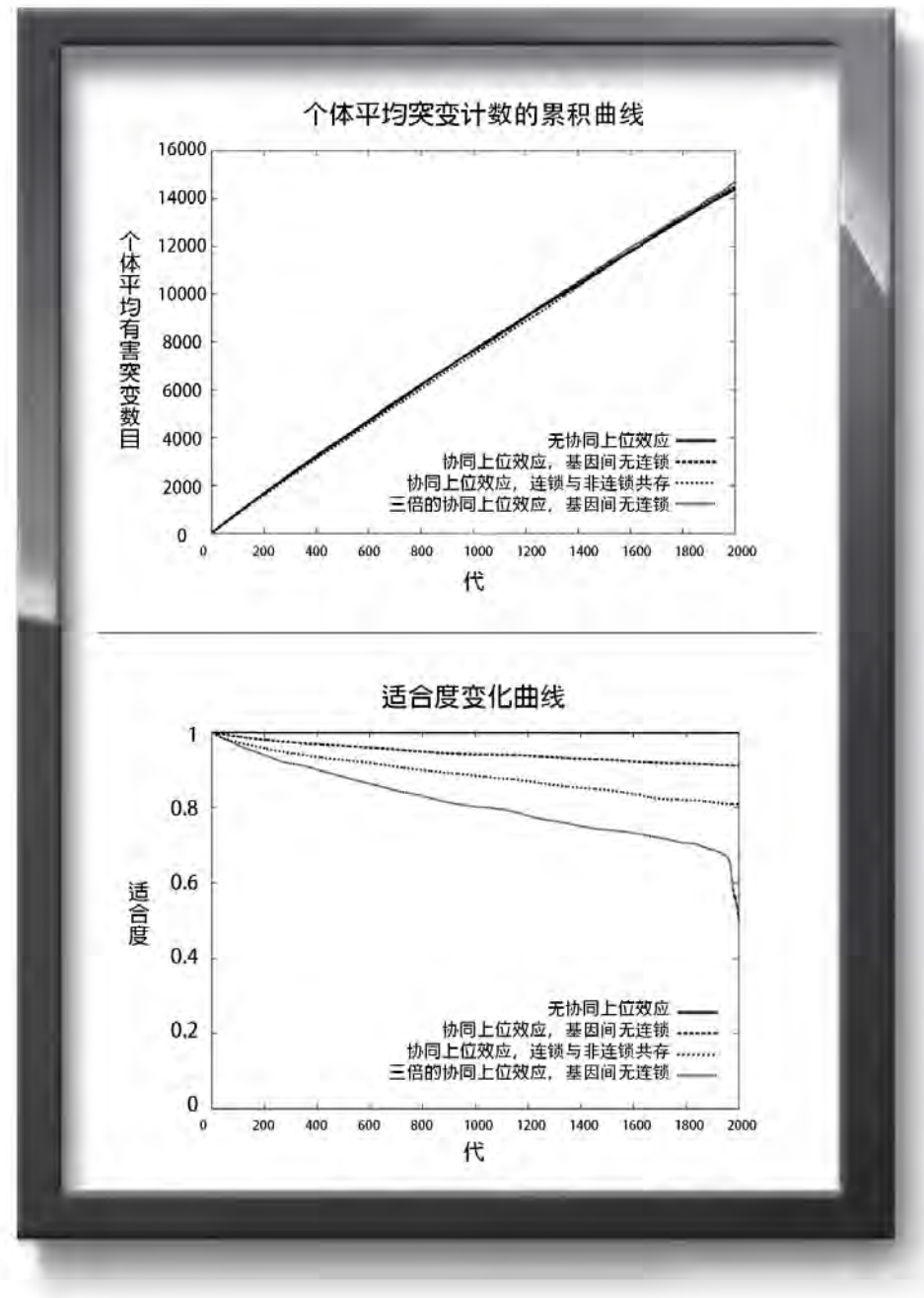


图10b 数值模拟实验检验“协同上位效应”的真实性

引自Baumgardner et al., 2013, 显示四个协同上位效应实验中个体平均突变计数的积累(上部)和适合度的变化。实验中采用了极端夸大了的选择压力和极端夸大了的协同上位效应。四个实验是: 1. 无协同上位效应; 2. 协同上位效应, 假定基因间无连锁; 3. 协同上位效应, 有连锁基因, 也有非连锁基因; 4. 协同上位效应, 假定基因间无连锁, 将效应加强三倍。选择皆为斩截性, 遗传力为100%, 生育力为2.0 (即每代中50%的后代可供淘汰)。即使在如此完美的条件下, 协同上位效应仍无助于阻止突变累积, 而只会加速适合度的下降。如果协同上位效应有效, 个体突变计数会停止上升, 适合度会停止下降(Baumgardner et al., 2013)。



人类可以自救吗？

简讯：优生学不能阻止退化，而克隆使突变问题变得更糟。

大多数遗传学内行都会承认，由于选择压力的松弛，人类存在基因退化的问题。他们或许也承认这个问题还延及其他的物种，至少是那些濒临灭绝的物种。他们甚至可能承认，要建立一个有效的选择方案来阻止人群中不断累积的基因破坏，还有一些理论难题。如果任其发展下去，基因破坏必然日甚一日，最终可能出现失控（熔断）。对于那些活着只为及时行乐的人来说，这与我何干？但对于那些寄一切希望于人类之生存和发展的理想主义者来说，这个问题当然令人揪心。他们可能将希望的目光转向科学和技术：人类的干预能够战胜这一威胁吗？基因退化实质上是不断积累的破坏，深入扩散在整个基因组中，要一个基因一个基因地修补是不可能的，所以像激光束般精准的基因工程显然也于事无补。然而，人们或许会问，人工选择或者新出现的克隆技术能否解决这个问题呢？

优生学能挽救吗？

有关人类退化的普遍认识在古今文献中有广泛的记载，所有的文化都有关于“上古之人”的传说，他们聪明、强悍、长

寿。达尔文的著作，《经由自然选择的物种起源》或称作《优良种族在竞争中的生存》，介绍了一种新观点，认为强力持续的选择过程可以制止这个退化趋势。他以人工选择在动植物育种方面的成就作为论据。在《人的发源》一书中，达尔文更进一步，辩称有必要让“优等”民族（即白人）取代“劣等”民族。该理论为现代种族主义开辟了道路，并在希特勒的德国登峰造极（Weikart, 2004）。第二次世界大战以前，许多国家，包括美国，都有政府领导的优生计划。这些计划包括对“不适者”施行强制绝育和对下层阶级大力推行堕胎和生育控制。从达尔文的时代开始，几乎所有他的追随者都在内心里是优生主义者，且鼓吹对人类进行基因改良。那些为进化论开创了现代“合成理论”的哲学家和科学家，无一例外都是优生主义者。然而，经历了二战的恐怖以后，几乎所有的对优生学的公开讨论都被悄悄地搁置了。

鉴于人类基因组江河日下，是否应该重新审视优生学？不幸的是，已经有人在这样做，但这无论在道义上还是在科学上都是站不住脚的。本书的主旨非但在逻辑上不可用来支持优生学，而且坚决反对之。优生主义者的愿望是个狡猾的骗局。没有一种选择形式能够阻止基因组的退化。试图在全球范围内人为地增加选择的强度从科学、社会、和政治的角度都是不可能的。只有一个无情的、全球性的、独裁的“权力精英”才会有能力制定选择的标准，决定谁可以生育，谁不可以。即使这样一个邪恶的计划能够有效地实施，它也不会阻止基因组的退化。任何可能的基因“进步”都将是杯水车薪，不足以对抗基因组退化的大趋势。“劣等”和“高等”都是模糊的、遗传力极低的特征，主要受非遗传（环境）因素所影响（见第六章）。对非遗传性状进

行选择，对基因改进的效果为零。鉴于前述的遗传学讨论，即使选择遗传力较高的性状也效果不佳。“适合度”是个非常笼统的术语，遗传力甚低，受无数基因的影响，这些基因大多是次要的或近中性的。正如我们刚刚学过的，有效的选择只能作用于少数“重要”的基因。我们的确可以人工选择任何单一的人体性状，使人变高或变矮，变白或变黑，变胖或变瘦。但我们不能有效地选择“高等”，因为这本身就是主观的，更牵扯到千万个基因和几百万个核苷酸。

在人类，强化选择或人工选择可能带来的任何益处不过是对极少数、极个别性状的轻微改进。但是基因组在大体上仍然会以同样的速度“锈掉”。优生的代价将是社会上和道德上灾难性的动乱。优生学从一开始就是一个种族主义的概念，并一直为基本公理所驱动。优生学缺乏稳固的遗传学基础。而且，它是与威权主义政府、精英主义哲学、压制人权和侵犯人的尊严紧密结合的。优生学不能把我们从基因组退化中拯救出来。如果优生主义者愚蠢地试图用这种方法来拯救基因组，那么谁来从优生主义者的手中拯救我们呢？

克隆来拯救吗？

克隆人就是从一个（应该是“高等”的）成年人身上取一个细胞，然后用这个细胞来制造出那个人的一个或多个拷贝。克隆会带来深远的遗传学后果。当我还是个进化论者的时候，对人类基因退化的问题也很担忧，但我天真地相信克隆也许就是答案。我当时希望克隆可以阻止基因组的退化，甚至可能快速地改善人类群体。

在植物遗传学里，当一个物种很容易繁殖时，克隆选择是改善群体最稳妥而且最快速的方式。既然如此，也许你会觉得奇怪对人的克隆选择为什么一直没有得到积极推广。然而，即使撇开克隆所带来的道德和社会问题，克隆的最佳效果也只能是短期改善，而从长期来讲一定会造成基因退化。

本书一个隐含的主题是对于（无性）克隆群体的知识也适用于有性群体中。一切克隆群体必然发生基因退化，这是穆勒首先证明的，被称为“穆勒的棘轮”（Muller, 1964; Felsenstein, 1974）。在任何克隆谱系中，突变将会随时间而积累。即使不断地在一个克隆谱系中选择最好的亚克隆，也还是不能制止退化，因为每一个亚克隆都比母克隆带有更多的突变。“棘轮”只能单向运转，所有的变化只能走下坡路。每一次细胞分裂都增加突变，而且没有消除突变的机制。即使可能出现有利突变，每一个有利突变也总是与众多有害突变相连接。在克隆群体中，没有什么机制能够切断罕见的有利突变与大量有害突变之间的联接。有害突变的数目总是比有利突变增长得快，而且终将拖累有利突变一起沉沦。所以，在克隆群体之内，净信息必将持续减少。再强调一下，这个机制也运作于最好的、最强力选择的亚克隆。遗传学家们很清楚在克隆群体中基因组的退化是必然的。至于为什么现在还存活许多克隆植物（和一些动物）群体，显然是因为这些群体的历史并不是很长，它们退化到绝种的时间还没有到。

动物克隆的初步实验表明，克隆动物甚至不会带来短期的基因进步，克隆的动物常规地显示即刻的和严重的基因缺陷。为什么会这样？克隆动物常表现出遭受突变破坏的迹象，似乎

它们“未老先衰”（Hochedlinger et al., 2004）。造成此现象的原因可能是多重的（遗传性的和表观遗传性的），然而一个主要的原因是突变在体细胞中的不断积累（克隆动物来源于体细胞而非生殖细胞）。正常生殖细胞的“设计”是尽量减少代与代之间的细胞分裂的次数，以便将每代突变积累降到最低限度。当我们估计每代每人点突变最低“只有”100个时，针对的就是生殖细胞。对体细胞（至少对不断分裂的干细胞）来说，细胞分裂的次数比性细胞要多得多，所以体细胞突变率也应该高得多（至少高10倍），而且随着年龄增长而急剧上升。随着我们长大并开始衰老，我们的体细胞中的突变越来越多，每个细胞也越来越“独特”。要保持一个庞大的基因组一成不变是不可能的，即便在一个克隆之内也不行。每个细胞中都发生着新突变，大约每次细胞分裂增加1-3个突变。所以，基本上我们身体的每个细胞都是独特的。鉴于这些原因，每个克隆人群总会比被克隆的成年“来源个体”要差。这样的一个克隆人体从某种意义上讲就是未老先衰，因为他继承了来源个体先天的突变负荷，再加上他自己成长和衰老过程中的突变累积。由于在身体发育和干细胞维持过程中有多次细胞分裂，一个从体细胞来的克隆人（就退化而言）相当于有性繁殖许多代之后的个体。在这个意义上，一个克隆体是该物种之未来的缩影。这是沉沦，而非提升（图11）。

伦理、社会和遗传学界都存在着反对克隆的强力争论。克隆可视为优生学的一种高科技的体现形式，在技术上和伦理上存在着许多的问题。一般意义上的优生学和克隆这种特别形式，都肯定不是解决基因组退化问题的途径。

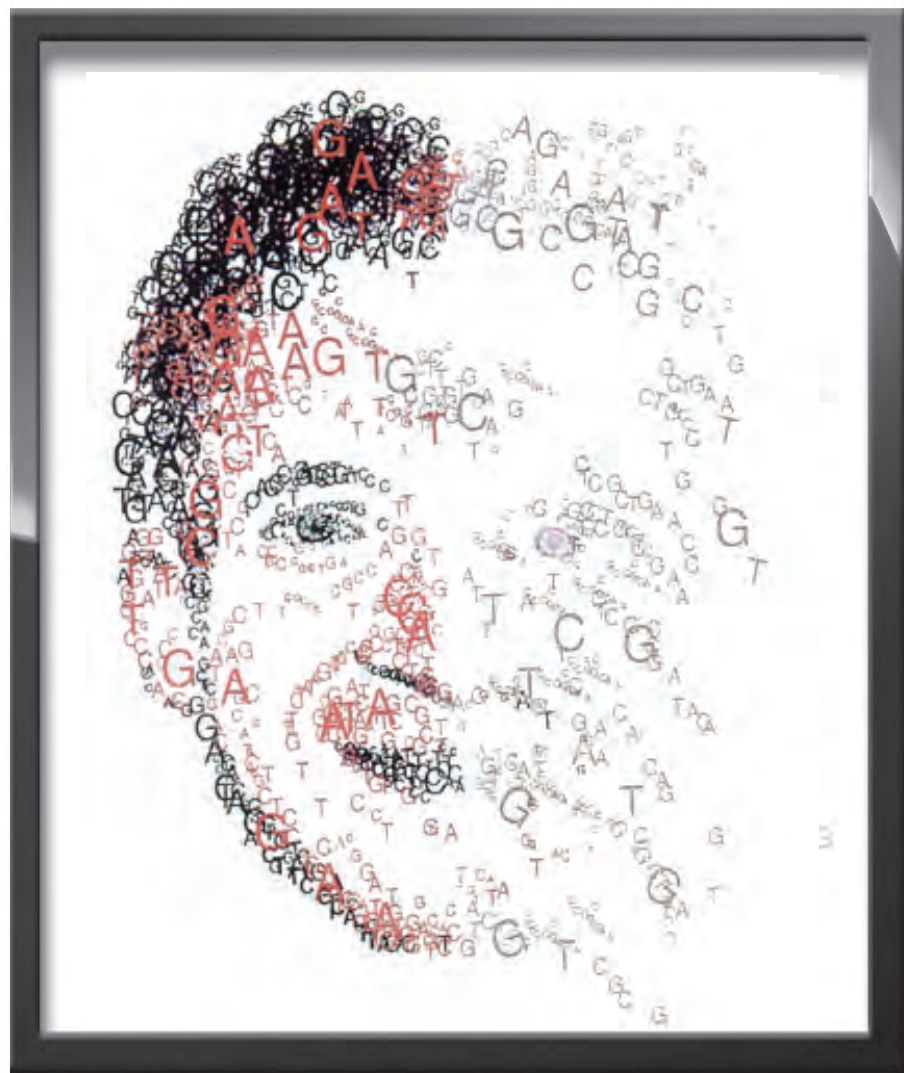


图11 基因组的退化、人的退化和人类的退化

我们亲身经历，更耳闻目睹。这就是“基因熵”，人对它无能为力，这是生物学的必然。这是物种灭绝的原因之一，也是为什么我们每个人都处在死亡的过程之中。

atcgtttaactgagtagc

第九章

自然选择能创造吗？

简讯：突变加选择连一个基因也创造不出来

我们已经审查了基因组退化的问题，发现有害突变的发生率很高。自然选择只能淘汰最糟的突变，其余的突变逐渐累积，就像汽车生锈一样。这种持续性的、系统性的信息腐蚀能否被基因组其他位点上的有利突变所补偿呢？回答是有利突变太过罕见，效应太弱，赶不上如此无情的、系统性的信息腐蚀。对此已有细致的分析 (Sanford et al., 2013; Montañez et al., 2013)。系统性地破坏信息很容易，然而除非有智能参与，要创造信息是很难的（许多人认为是不可能的）。

这个问题使全基因组正向进化的希望荡然无存。然而，某些个别性状还是可能经过突变/选择而有所改进。这种前进性（“创造性”）突变/选择究竟能走多远？至此应该已经很清楚，靠说明书中的随机拼写错误永远不会发明出一个飞机零件（比如说一个铝合金部件），让飞机整体性能显著改善。即使允许无限次的试飞、无限次的空难和无限度的资金，也无助于事。所以我们当然有理由问一个类似的生物学问题，“突变加选择能够从无到有地创造一个有功能的基因吗”？

一个基因就像一本书、书中的一章、或一个可执行的程序——最少也有1000个字符串。突变加选择连一个基因也创造

不出来，这是因为大量的有害突变占据了绝对优势，即使在一个基因的内部也是如此。一个基因或一个连锁块里的净信息量永远呈下降的趋势。即使已经造出了一个基因的50%，有害突变很快就会把它降解，这比有利突变创造另一半的速度要快得多。然而，为了更好地理解正向选择的局限，让我们现在忽略所有的有害突变而只考虑有利突变。这样的话突变/选择能不能产生一个有功能的新基因呢？

1. 定义第一个理想的突变

要通过突变/选择创造一个新基因，我们遇到的第一个难题是如何定义第一个有利突变。就其本身而言，没有一个核苷酸（A、T、C、或G）比其他核苷酸更有价值，正如字母表中的任何一个字母，如果没有其他字母的配合，就没有特别的意义。所以如果没有周围其他核苷酸（实际上是整个基因组）的配合，就不可能对任何一个核苷酸进行选择。在一个词或一章书中改变一个字母，其效果只能在它周围的文理背景中评估。这就是“不可简化之复杂性”原则的一个极佳的例子。事实上，这是不可简化之复杂性的最基本层次。我们立即发现这是一个困境：要创造一个新功能，我们需要选择第一个有利突变，但我们只能从比邻关系中定义一个新核苷酸的价值，而我们很快又必须改变大多数比邻的核苷酸！我们在这样地兜圈子，在不断地破坏着要赖以作为基础的“上下文”。这个核苷酸之间的基本关系称为上位效应。真正的上位效应具有无限的复杂性，实际上无法分析，这就是为什么遗传学家们总是干脆忽略它。

这个让人不知所措的复杂性正说明了为什么语言（包括基因语言和基因信息）从来不是机遇的产物，而是需要智慧设计。基因组其实就是一本书，是用真正的语言写成的，短序列其实就是句子。要让散乱的字母自己随机拼成一个有意义的句子，所需要的试验次数和时间，在地球历史内根本容纳不下（“我觉得那像一只黄鼠狼”这句话，莎氏原文是：Methinks it is like a weasel. 这是用26个英文字母加空格键制造出的28个字符，共需试验 27^{28} 次，即1后边40个零），任何有功能的核苷酸序列也是一样。如果一个有功能的序列中有超过几十个核苷酸，我们现实地知道这决不仅仅是突然“巧合”而成，这已经用数学方法反复验证过了。我们很快就会看到，这样的一个序列也不会一个核苷酸一个核苷酸地逐渐形成。一定得有一个预先存在的“概念”作为一句话或一个功能性序列得以建立的框架，而这个概念只能预先存在于创作者的头脑中。从第一个突变伊始，我们就遇到了如何定义第一个理想的有利突变的基本问题！

2. 等待第一个突变

假定我们知道第一个突变应该是什么，要等多长时间才会出现呢？人类的进化通常被认为是在约一万人的小群体中发生的。就任何一个特定的核苷酸来说，每代每人的突变率极低（大约在3000万人中才会在某个特定核苷酸位点上发生一次突变）。所以在一个一万人的群体中，要等3000代（至少6万年）才能指望某一个特定的核苷酸发生突变。但它突变为“错误

的”核苷酸的可能性是三分之二，所以要在某个特定位点上发生一个特定的理想突变就需要三倍的时间，至少18万年。突变在一个人中发生以后，必须固定下来（就是说要让群体中的所有个体都带有双重拷贝）。因为突变刚出现时在群体中只有一个拷贝，所以它一出现就濒临灭绝。绝大多数新突变很快就会在群体中漂失，即使有利突变也不例外。所以任何理想的突变必须发生多次才可能在群体中“站稳”。唯有当突变为显性而且很明显地有益时，选择过程才有一定的机会，将其从随机泯灭的命运中拯救出来。根据群体遗传学家的说法，除非存在有效的选择，在一个一万人的群体中，一个新突变不随漂变而流失的机会只有二万（群体中该位点上非突变核苷酸的总数）分之一。即使有一定程度的选择，随机流失的可能性还是很大，尤其是隐性突变和表现不明显的突变（我们实际上知道几乎所有的有利突变都是隐性和近中性的）。所以，在多数情况下，即使是一个有利突变也将随基因漂变而随机流失。我们的数值模拟提示，弱效应有利突变的流失率为99%。

所以一个典型的轻微有利的突变必须发生100次才可能在群体中“站稳”。这样，平均来讲，我们要等 $18万 \times 100 = 1800$ 万年，才会稳定地得到我们所想要的第一个有利突变，以建立我们假想的新基因。假如我们是600万年前从类人猿进化而来的，那么我们还没有足够的时间，在基因组中要建立新基因的位点上，得到第一个被固定下来的理想突变。1800万年间会发生大量的突变，但一个特定核苷酸突变为另一个特定核苷酸并免于基因漂变而固定下来的可能性只有一次。

3. 等待其他突变

在第一个突变出现并固定之后，我们需要重复这个过程以得到我们所想要的基因的所有其他核苷酸。一个基因最少也有1000个核苷酸。更现实地讲，如果算上调节元件和内含子，一个人类基因的平均长度大约有50000个核苷酸。退一万步，我们就算一个基因只有1000核苷酸（而且假定每个核苷酸可以被独立选择）。如果这个过程是直接的、线性的和依次，就需要花 $1800万 \times 1000 = 180$ 亿年时间来创造一个最小的基因。从所谓的宇宙大爆炸至今还没有这么多时间！所以说理想突变的稀少限制了进化速率，这绝非夸大其辞。再者，单个的核苷酸本身并不携带任何信息，因而根本无法选择。特定信息需要很多字符（至少需要一个句子或类似的字符串）。与任何信息一样，编码特定生物功能的遗传信息也必须含有许多核苷酸以达到其“功能性门槛”。功能性门槛就是要传递有意义的信息的最低字符（或核苷酸）数。若低于这个门槛，单个的字母或核苷酸就没有利处可言，因而不可选择。这就意味着在现实中，等待的时间要长得多——因为在最小核苷酸序列偶然出现之前并没有选择的作用。假如选择的功能性门槛是12个核苷酸（在所有12个字母都到位之前并无选择），那么我们假想的人群就要等待数万亿年。

4. 等待“霍尔丹之困境”

当第一个突变终于在群体中发生后，在它固定之前，它需

要时间被自然选择所放大。在一个一万人的群体中，一个全新的突变只是2万个相互竞争的核苷酸中的一个（整个群体在那个位点上一共有2万个核苷酸）。突变的核苷酸必须凭借漂变或者自然选择在群体中渐渐增多。也许很快就有2个突变拷贝了，然后4个，然后100个，最终达到2万个（每个个体2个拷贝）。这个过程需要多长时间？对于显性突变，假设有非常强的单向选择，这个突变的核苷酸有可能以每代10%的速度在群体中增长。以此高速度，大概还需要105代（2100年）才能从一个拷贝增加到2万个拷贝（ $1.1105^{105} = 20000$ ）。然而，突变固定需要的时间通常要长得多，因为选择一般很弱，而且大多数突变都是隐性和轻微的。当突变是隐性，或者当选择不是持续单向或不强时，这个计算就复杂了，但固定过程显然会慢许多。例如一个完全隐性的有利突变，即使它能增加1%的适合度，也要至少100000代（200万年）才能固定下来（Patterson, 1999）。

霍尔丹（Haldane, 1957）的计算表明，如果显性和隐性突变按着一个他认为“合理”的比例发生，要使一个新突变经选择而固定下来，（平均来说）需要300代（>6000年）。这一选择速度是如此缓慢，以至于跟不选择没有什么两样。这就是被称为“霍尔丹之困境”的经典问题。依此选择速度，人类如果真的是从猿猴分化出来的，在此后的6百万年里，整个人类基因组中只能固定下1000个有利的核苷酸突变。这个简单的事实已被克劳和木村（Crow and Kimura, 1970）、赖敏（ReMine, 1993, 2005）及路普和圣弗德（Rupe and Sanford, 2013）分别验证。再者，选择的本质决定了选择一个核苷酸的时候会降低对其他核苷酸的选择能力（选择干扰），所以要对多个弱效应有利突变同

时进行选择，基本上没有什么效果。

表面上看，以上计算似乎暗示着自我们从黑猩猩分化出来以后的时间里，至少可能经过选择固定一个小基因（长达1000个核苷酸）。但这并不正确，原因有二。首先，霍尔丹的计算仅针对于独立的、互不连接的突变。选择1000个特定的比邻的突变（以创造出1000个字母构成的序列）在6百万年内不会发生，因为邻接的突变构成的那个特定顺序永远不会出现，即使几万亿年也不行（据上述运算），你不能选择还没有形成的字符串。其次，一个基因中的核苷酸大部分是近中性的，根本不可选择，不管时间多长。霍尔丹之困境说到底就是，新的有利突变经选择而固定下来是个极缓慢的过程，选择的核苷酸越多，进展就越慢。这就严重地限制了正向选择。在合理的进化时间内，我们只能选择数量极有限而且互不连接的核苷酸。在过去6百万年中，选择最多只能固定1000个互不相连的有利突变，所产生的新信息比一页文字还少。这么少的信息没法把一只猿猴变成一个人。

5. 等候重组？

因为有性物种（例如人类）每一代都可以在一定程度上交换突变，也许有人会认为，建立一个新基因所需的全部突变，可以在群体中的不同个体身上同时发生，然后把所有的理想突变通过重组“连接起来”。这就意味着突变不一定非要一个一个地依次发生，如此就缩短了创造希望中的基因所需的时间（所以我们也许不需要几十亿年）。这里有两个问题。首

先，当我们审查人类基因组时，我们总是发现基因组以大连锁块（20000 - 40000个核苷酸）的形式存在，自人类起源以来连锁块内还没有发生过重组(Gabriel et al., 2002; Tishkoff and Verrelli, 2003)。这就是说，在局部核苷酸序列的水平上其实不会发生有意义的重组。只有像基因一般大的DNA片断才会互换。我再说明一下，实际上没有核苷酸水平上的互换！其二，即便存在有效的核苷酸互换，要使群体中所有的突变连接在一起构成我们希望中的1000个核苷酸的序列，其可能性实在微乎其微，而且这样的互换拆毁字符串的速度要比建立字符串快。字符串大量互换所需的时间比依次突变更长。如果真的存在一种核苷酸水平上的互换机制可以用来建筑一个新基因，那么在新基因巧合形成之后的下一代，它又会立即被同样的核苷酸交换机制所撕裂。在扑克牌中，你摸到一连串王牌的可能性不大。如果你真走了一回运，那么在洗牌之后，同样的手气再来一次的机会又有多大呢？重组无助于我们创造一个由1000个核苷酸构成的序列。

6. 无尽的适合度低谷

进化论者承认要创造一个新基因需要经过许多“尝试”。在一个新基因的建构阶段，我们必须预料到在一段时间之内，这种尝试会使一个物种的适合度降低。这是一个适合度低谷。一个未完工的基因既不是有利的，也不是中性的，而将是有害的。从某种意义上讲，物种在变得更好之前，必须先变得更差。如果适合度低谷不多而且短暂，就不难想象一个物种能够从中存活下来。

深度的适合度低谷可能导致灭绝。有益突变的稀少再加上霍尔丹的困境将使适合度低谷深而又长，无边无岸。持续的进化创新必将使物种的适合度不断下降。生命的适合度低谷将是一个接着一个。进化的超级高速公路将永远在施工之中，总适合度将总是减少而不是增加。某个物种成功地穿越适合度低谷的说法，只有在考虑个别性状的时候才有进化上的意义。当考虑整个基因组时，无限多、无限长的适合度低谷将使进化方案无法实施。

7. 多重限制的DNA

多数DNA序列具备多重功能，因而也必须受到多重限制。这就是说DNA序列在不同层次上有不同的含义（多重功能），而且每一层含义都限制了未来可能的改变（多重限制）。例如，试想一句话按正常格式阅读时有一个很具体的信息，但倒过来读时还是一句有条理的话。现在假设每间隔一个字母读下去它又有第三个信息，而且用一个简单的解码程序来翻译后还可以发现第四个信息。这样的信息就是多重功能和多重限制的。我们知道一个正常句子中的拼写错误通常不会改进原来的信息，但至少有这个可能。然而，一个多重限制的信息是不可思议的，它不可以被改进，而只能被破坏（参看图12，第134页）。任何可能改进正常句子格式的拼写错误将会打乱其他层次的信息，任何改变将无疑会减少总信息量。我和我的同事们最近在一篇论文中对此进行过数学演示(Montañez et al., 2013)。

大量证据表明多数DNA序列都有多重功能，因而也受多重限制。垂弗诺夫(Trifonov, 1989)曾深入探讨过这个问题。例如，人类基因组中大多数蛋白编码序列，在不同的方向上编码两个不同的RNA分子（就是说，DNA的双股都被转录，参见Yelin et al., 2003）。有些序列使用多个翻译起始位点和多重读码框架编码不同的蛋白质。有些序列靠mRNA的交替剪切编码不同的蛋白质分子。有些序列同时发挥多种功能（例如，既被用作蛋白编码序列又被用作内在转录启动子）。有些序列既是蛋白编码区，也是蛋白结合区。启动子和外显子内部还可见到Alu元件和DNA复制起始位点。基本上所有的DNA序列都受限于等值区（局部GC含量）、“词”含量（种属特异性的双核苷酸、三核苷酸、或四核苷酸序列的出现频率）和核小体接合区（因为所有DNA都必须压缩）。选择性的压缩明显地与基因调节有关，而选择性核小体接合是由贯穿于全基因组的特定DNA序列模式来控制的。最后，可能所有序列都会影响总体性的间距和DNA折叠/构型，因为这些都明显地受控于序列。为了解释基因组所包含的难以置信的信息量（考虑到生命的极端复杂），我们确实必须猜想基因组内还隐藏着更高级的组织结构和信息。例如，我们知道在表观遗传的水平还有另一层的组织(Gibbs, 2003)。在整个细胞核水平和染色体水平上似乎还存在广泛的、序列依赖性的三维组织(Manuelidis, 1990; Gardiner, 1995; Flam, 1994)。垂弗诺夫(Trifonov, 1989)已经表明，也许基因组内所有的DNA序列都锁定了多层“密码”（可高达12层）。在计算机科学中，这种数据压缩只能来自于最高水平的信息设计，以求最大密度的信息整合。这些高层次上的组织结构和信息内涵使

得DNA的多重限制问题更为突出。每一个核苷酸都与许多其他核苷酸相互作用，基因组中的一切都似乎取决于背景因素。这个无处不在的、贯穿整个基因组的、多重限制DNA的问题对于进化理论来说似乎是绝对压倒性的。改变任何一点都有可能改变一切！突变/选择虽然在细微的水平上起一定作用，但DNA的多重限制性，有力地证明了高级基因组不可能经过突变/选择进化而来。从逻辑上讲，一切多重限制的DNA只能来自设计。

8. 不可简化之复杂性

关于不可简化之复杂性的问题，贝希(Behe, 1996)作出了非常精彩的讲解。他举出各式各样的多元件系统来讲述不可简化之复杂性的概念，比如一个捕鼠夹子需要5个独立的部件，一个细菌鞭毛需要10到20个部件。他指出每一个部件只有在整个功能性单元中才有价值，所以不可简化之系统的所有部件都必须同时出现，而不可一一地产生。在捕鼠夹的例子中，所有部件也许曾同时排列在发明者的工作台上，但它们不可能靠机遇或者任何符合现实的进化过程自己组装起来。它们是在发明者的头脑中作为一个综合概念同时出现的。高度的复杂性都是在意识领域以整合功能的形式首先形成的。

在我们的运输工具进化的例子中，我们所能想象的最简单的第一个改进也许是拼写错误把我们的红色小拉车转变为一辆蓝色的三轮自行车。确实容易想象一个拼写错误有可能导致颜色编码的改变（虽然肯定需要有蓝色油漆，而且已经编了码）。同样地，一个拼写错误肯定能让一个轮子掉下来。然

而，三个轮子的拉车不是三轮自行车，而是一辆坏掉的拉车。要将一辆小拉车变为一辆三轮自行车，就需要对说明书进行大量的改写，并要彻底地更换大多数的部件。在实现这些复杂的改变之前，并没有什么有功能的中间形式，所以辅助这个过程的质量控制员根本没有什么目标可以选择。事实上，他会试图淘汰我们所想要的一切拼写错误和改变。因此，理想的拼写错误的组合必须偶然一下子产生，这是永远也不会发生的事件。显然，从一辆小拉车到一辆三轮自行车的转变只能通过智慧来进行大量的重新设计和彻底地改写说明书（参看图13，第135页）。

一辆小拉车或三轮自行车或许有几十个零部件，但最简单的蛋白质也要比它们复杂得多。每一个蛋白质分子都有几百个零件，达到了比比喻中的小拉车更深层次的不可简化之复杂性。随着一个设计的部件数目呈线性递增，部件之间的相互作用（由此带来的复杂性）将呈指数增长。

蛋白质已经是复杂的了，在每个蛋白质背后是一个具有更高度不可简化之复杂性的遗传系统。实现蛋白质之编码、转录和翻译的分子机器是不同凡响的。撇开所有其他的辅助性蛋白质分子不谈，仅仅是DNA/RNA序列的设计就足以让人叹为观止。虽然一个简单的蛋白质是由几百个零件组成，但编码它的基因却有几千个零件。所有这些零件都相互影响并相互定义。每一个核苷酸只有在其他核苷酸的配合下才有意义。每一个核苷酸都是多重功能的，并与许多其他核苷酸相互作用。DNA序列决定着局部染色质的三维结构、区域性蛋白质结合、解螺旋、转录、并同时编码着一个或多个RNA序列。RNA序列限定了RNA的稳

定性、RNA的交替剪切、RNA加工、RNA运输、转录效率和蛋白质序列。

当我们考虑一个基因的方方面面，包括其调节元件和结构元件，这大约有5万个部件。我猜想这个数目比一辆现代化汽车的部件还要多。汽车零件不会沿着一个简单的直线过程自动地组装成一辆有功能的汽车。然而，一个基因的不可简化之复杂性与一个细胞的不可简化之复杂性相比，充其量只是宇宙中的一点微尘。生命本身就是不可简化之复杂性的精华，这就是为什么我们根本不能想象人可以从头创造生命。生命是一层又一层的不可简化之复杂性。我们最引以为豪的生物化学流程图，也只不过是真实生物复杂性的儿童卡通而已，因为有些东西我们根本无法理解。靠着人类理性的成就，我们已经开始理解单个基因的运作，并有能力设计和构建一些很小的人造基因了。但是我们还不能设计一个新基因来编码一个全新的未曾有过的蛋白质，并将其精确地融合于一个高级生命形式的细胞之中。如果我们不能做到，为什么我们认为随机突变，加上非常有限的繁殖筛的作用，就能做到呢？如果读者有兴趣，我已试图用整体复杂性的概念来拓展不可简化之复杂性（参见附录3）。

9. 几乎所有的有利突变都是近中性的

我们已经充分地讨论了选择淘汰近中性有害突变是如何困难，这个问题也为多数遗传学家所承认，尽管承认得不甘心。然而，这个问题还有另一面，这一面也很重要，但据我所知，从未有人承认过。正如我们已经在图3d（第29页）中讨论

过的，近中性的问题对于有利突变来说比对有害突变要严重得多。基本上每一个有利突变都必然落在木村的“非选择区域”。所有这样的突变从来不能被选择。这个问题加重了我在上文中已列举的一切问题。我们希望中的新基因当然会有少数强效应的核苷酸，比如那些编码一个酶的活性中心的核苷酸。但是这样的核苷酸，只有在整个蛋白质和整个基因序列的背景中才会发挥主要效应。构成一个蛋白质/基因的大部分组分，对于整个分子的影响都很小，对于整个生物体适合度的影响更是微不足道。但这些核苷酸综合起来包含了基因中的大部分信息，没有了它们的配合，那些“重要的核苷酸”就毫无意义。然而多数核苷酸是不可以单个地选择的。在基因构建过程中我们怎样才能建立它们并维持它们的位置呢？答案显然是我们不能。而离开了这些“无关紧要的群众”核苷酸，那些“重要的核苷酸精英”也不能被选择。因为近中性的问题，我们在建造理想基因的过程中根本走不出第一步。新基因的整体框架是由近中性核苷酸决定的，但自然选择没有办法设置或维系这些核苷酸。有益突变的近中性本质，有力地证明了每个基因必须是被设计过的，根本无法通过选择一个核苷酸的方式来建造一个基因。

10. 再来看看坏突变

我们已经从不同的侧面简要地考虑了为什么前进性突变/选择的作用范围必然是有限的。这些论证都暂时搁置了一切有害突变。然而，前进性选择必须发生在真实世界中。在真实世界

中，有害突变远远压倒了有利突变。为公平起见，我们现在必须重新引进有害突变。

A. 穆勒的棘轮

正如我前面提过的，当我们研究人类基因组时，我们看到巨大的DNA连锁块内基本上没有重组的历史证据(Gabriel et al., 2002; Tishkoff and Verrelli, 2003)。重组主要发生在基因之间而不是核苷酸之间。所以在任何一个有限的基因序列之内基本上没有重组。任何像这样没有重组的DNA连锁块必受到“穆勒的棘轮”的影响(Muller, 1964; Loewe, 2006)。这意味着好的突变和坏的突变不能分离。既然我们已经知道坏的突变远远多于好的突变，所以可以肯定这样一段DNA必然退化。成群的坏突变总是会裹挟着极少数的好突变一起沉沦。当我们还在等待着一个罕见的有利突变时，大量有害突变就已经在附近到处堆积。即使我们能够成功地在一个区域积累100个“好”突变，正等着下一个出现的时候，我们会看到许多好突变已经开始逆转变坏了。在这样的情况下，时间是我们的大敌。时间越长，剩下的信息越少。穆勒的棘轮在一个新基因还没有发挥功能之前早已把它杀死了。

B. 选择的代价太高

在前面的几章中我们讨论了选择的代价。霍尔丹的困境只考虑了前进性选择。但我们只有在支付了其他所有的繁殖代价

包括排除有害突变的代价之后，才有能力“资助”对有利突变的前进性选择。正如我们已经知道的，有害突变的数量太多，以至于我们根本没有足够的过剩群体来排除他们。既然我们无力阻止退化，显然更没有盈余群体用于前进性选择。只有一个办法。短期来讲，我们可以把对抗远期有害突变的“选择金额”借来用于对有限数目的性状进行前进性选择。然而需要明白，这意味着因个别有利突变而带来的短期适应性进步，是以加快基因组其余部分的退化为代价的。

C. 非随机突变

其实，突变并不是完全随机的，但这能帮助创造新基因吗？不能，这使问题变得更糟了！例如，我们现在知道有些核苷酸位点比其他的位点更易于突变（“热点”），而有些核苷酸在置换后出现率高。突变“热点”可能很快地给出我们想要的基因型，但当我们等待“冷点”内产生相应的突变时，热点又开始逆转了。所以在等待冷点内出现第一个有利突变的同时，我们被迫一次次地重新选择热点内的有利突变。这使事情变坏而不是变好。当突变成为某种核苷酸（比如说T）的可能性较大时，对于需要这种核苷酸(T)的位点有利，但这会减慢需要其他核苷酸(G、C、和A)的突变。所以，75%的情况下，对某种核苷酸的偏好会使正向突变的速度放慢。“非随机突变”从建立信息的角度听上去不错，但不幸的是，这不是讲的设计上的非随机性，而是在讲一种与信息建立相反的非随机性。

D. 人类和黑猩猩都应该已经灭绝

虽然我们可以暂时不考虑有害突变，但是应该指出的是，自从假想中的人类与黑猩猩分化以来，遗传学家相信两个谱系中都应该已经通过基因漂变而固定下了成千上万的有害突变(Kondrashov, 1995; Crow, 1997; Eyre-Walker and Keightley, 1999; Higgins and Lynch, 2001)。从逻辑上讲，我们已经比我们的类人猿祖先显著地退化了。这种逻辑是无可辩驳的。事实上，我们知道人和黑猩猩至少有1.5亿个核苷酸差异(Britten, 2002)，据说这是由于超过4000万个被固定下来的突变。所以，如果我们假定人是从类似黑猩猩的祖先进化而来的，那么在此过程中，人类与黑猩猩两个谱系中必然都固定下了大约2000万个核苷酸突变。然而其中只有大约1000个是自然选择的结果，其余的只能来自随机漂变，也就是说基因漂变留下了数千万个近中性的有害突变。结果呢？有利突变最多只有1000个，有害突变有上千万个。这不仅使我们比类似黑猩猩的祖先低下，而且在6百万年内显然会把我们杀绝了。

我们已经考察了一些有力的证据，即使忽略有害突变，突变加选择也不能在人类进化的时间范围内创造出一个基因。当把有害突变考虑在内时，我们看到突变加选择永远也不会创造出一个基因。这对于否定基本公理来说是压倒性的证据。我的观点是，这从逻辑上证明了基本公理的谬误性。

根据以上论证，基因组和其中的每一个基因只能来自设

计，而非进化。然而我们都知道微观进化（适应性选择）确实发生。何以如此？这是由于大多数适应都是靠着微调，而非新信息的创造。进一步说，大多数适应并不是新出现的随机突变所致，而是来自预先存在的基因多样性。用我们前面提到的说法，突变是生命中的撞痕、刮伤和损毁的部件。所以我们有理由认为，多数有益的变异是设计中的多样化。当我们看到适应性选择发生的时候，一般地说，我们所见到的是有用的基因型和基因部件的分离和重组，而这些基因和基因部件在最初被设计时就是可分离、可重组的。我们所见的一般不是随机突变的结果，因为随机突变总是有害的。在适应过程中，大自然会选择经设计而来的多样基因的理想组合体和分离体。举例来说，一对夫妇只要在极少部分核苷酸上是杂合体，如果他们携带的都是设计中的功能性基因型，他们就可以通过重组和分离而产生几乎无限的、有意义的多样化后代。这种经设计而来的多样化才是自然选择最有效的作用对象。所有这些设计中的变异类型都可以在有用的连锁基因群中形成，而且其出现的频率会很高。譬如最初只有一对夫妇，那就只有4套染色体，所有核苷酸变异类型的初始频率至少是25%。功能性连锁群和如此高的等位基因频率大大地促进了选择的效率，使人类快速适应所居住的地区。这种情况就像一叠有序排列的扑克牌，最初的净信息量最大，但是要等多轮洗牌以后才会出现最大的多样化。新的功能性信息只在最初的时候需要，以后就不再需要合成新信息了。

2008年补充：2007年迈克尔·贝希发表了第二本著作《进化之界限》。他的第一本书大体上是从理论出发来讨论问题的，而这本新作则综述了从三种医学上很重要的微生物（爱滋病毒、疟疾病原体和大肠杆菌）中得出的大量分子生物学实验数据。这些微生物都被公认具有高度的可变性，并且已经经历了大量的选择周期，被当作“进化系统”的有力例证。贝希表明虽然这些生物能快速地适应新的外部环境，但它们并不能创造出新的内部功能。即使在那些“理想的进化系统”里面，所记录到的变化也只是“微调”，而不是真正的创新。虽然科学家已经追踪了这些模型系统的大量繁殖周期（比高等生物在天文时间内所能经历的繁殖周期还要多得多），他们所观察到的改变却全是临时性的“补漏子措施”。贝希有力地显示了“基本公理”即使在最简单的层次上也不能产生不可简化之复杂性。贝希的新书受到了激烈的攻击，然而最近的新资料有力地验证了他的分析（见 http://www.evolutionnews.org/2014/07/so_michael_behe087901.html）。

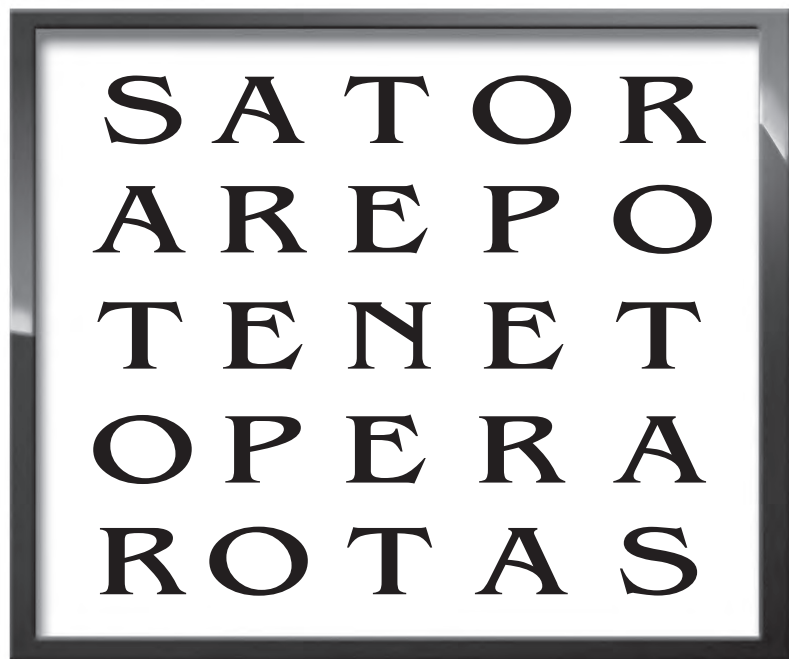


图12 多重限制的信息和多重限制的DNA

像双关语、回文和其他文字游戏一样，DNA包含了多功能的字母、单词和短语。这类序列只能来自很认真的设计。一旦形成之后，它们不会通过“突变”得到改进。一个绝好的例子就是以上所示被巧妙构思出来的多功能拉丁短语（参看 Ohno 和 Yomo, 1991）。这个古老的拼字游戏（可追溯到公元79年）翻译出来的意思是：“一个叫阿雷波的撒种者抓住了运转中的轮子。”这个短语从四个方向来读都是一样的：从左上角开始：从左到右，或从上到下；从右下角开始：从下到上，或从右到左。改变这里面任何一个字母都会同时破坏这四个信息（这个例子当中，四个信息恰好都是一样的）。同样地，一个简单的回文句子是：ABLE WAS I ERE I SAW ELBA（我在看见艾尔巴以前还真行）。正反读起来都是一样的。改变任何字母都会同时破坏两个信息。一个简单的多功能单词是LIVE（活），反过来是EVIL（恶）。把LIVE变成HIVE（蜂房）也许还可以，但反过来读，就使EVIL变为没有意义的EVILH。所以这个双重意思的单词，像上面两个例子一样，也是多重限制的，因为它是多重功能的。



图13: 不可简化之复杂性

在我们红色小拉车的例子里，人所能想象的最简单的改进就是拼写错误能够将我们的红色小拉车转变为一辆蓝色三轮车。这个看起来不大的进步不可能随机产生，因为它要求创造出许多新的部件，每一个都有“不可简化之复杂性”。不难想象，一个拼写错误能够改变颜色的编码，或使一个轮子掉出来。然而，要制造一辆能用的三轮车实际上需要对大部分的零件进行改造。一个新部件（脚踏装置）就能说明这一点。一个有功能的脚踏装置不会从几个拼写错误而来。要制造和组装脚踏装置的不同部件，需要在我们的说明书中增加几章全新的内容。但是光有脚踏装置而没有座子和放腿的位置也不行。非常明显，小拉车的组装说明书因拼写错误而败坏之后（即使加上我们的质量控制员帮忙）永远也不会导致一辆暂新的蓝色三轮车。它只会导致不可用的破拉车。



果然是向下沉沦？

简讯：一切证据都指向基因退化

把基因组比作说明书，把突变比作文字处理错误，这个贴切的比喻帮助我们看到基因组一定会退化，信息的本质也告诉了我们基因组退化的道理。这一常识性的结论得到了信息理论的支持(Gitt, 1997; Gitt et al., 2013)。人类的高突变率显示人类正在退化。同时选择大量突变所需要的超高代价也表明人类正在退化。近中性突变的问题、选择阈值、和选择干扰都提示人类一定在退化。最先进的数值模拟无可辩驳地证明我们一定在退化(Sanford et al., 2007b; Gibson et al., 2013)。

从图4（60页）我们可以看出基因熵应该是什么样子——这是一个经典的生物学衰减曲线。该图是根据克劳(Crow, 1997)所称的人类适合度衰减速度（每代1%），显示了数千年内人类基因衰退的情况。然而根据林奇最近的研究(Lynch, 2010)，实际的衰减速度可能要高得多——可认为每代5%。本书中有多个类似的适合度衰减曲线（图4, 10a, 10b, 14, 15, 16）。有些图仅仅是基于遗传学理论，有些是基于数值模拟，有些是基于生物学研究数据，有些是基于历史资料。然而它们都显示同样的基本曲线——一条清晰的生物学衰减曲线。这些图互为印证。

因为人们心中有一个根深蒂固的设想，认为到处都有上升性进化，所以如此快速的基因退化似乎让人难以置信。然而在过去的60多年里，群体遗传学界的领袖们对此已经反复表达了严肃的关切，基本上他们就是在问：“遗传学理论怎么能坦然地排除这种逆向进化？”（见附录1）。我在他们工作的基础上已经花了十多年的时间，以求增进我们在这个领域的理解。可以说，我在基因熵这个问题上的专门研究比任何科学家做的工作都多。结合着简单的逻辑和数学的描述，我仔细地分析了突变和选择过程实际上是如何运作的。最重要的是，在过去十年中，我和同事们开发了最先进的数值模拟程序，现实地模拟了突变/选择过程。我们的数值模拟程序同时、全面地考虑了所有影响突变/选择的相关生物学因素。因为我们现在能够进行符合生物学现实的数值模拟，所以我们实际上可以看到突变蓄积的过程是如何发展的（见第11章）。这项工作清楚地揭示，基因熵是一个非常真实的问题，而且比以前的群体生物学家所承认的要严重得多。符合现实的数值模拟无一例外地得出一条生物学衰减曲线，与图4（60页）中适合度的理论性衰减几乎完全同样。适合度的轨迹总是下行的，而非上行。

除了这些证据之外，现在又在RNA病毒（如流感病毒）中发现了基因熵的有力证据（图15，第148页）。曾在20世纪引起大流行的三株流感病毒，都表现出与突变蓄积相关的毒力减弱，而且减弱的速度相当快。

几十年来，进化生物学家一直在哲学的水平上坚持认为，在生物系统中，自然选择是熵的反作用力，它一定能逆转生物界的退化。然而所有的最可靠的研究结果都与这个哲学假设相

抵触。尤其是在大基因组中，突变熵的力量是如此强大，绝非选择过程所能扭转。因此大基因组最终难逃灭亡的命运（除非在选择之外，还存在某种尚未被发现的反作用力）。我称这个根本的问题为基因熵。基因熵不是一个公理性的预设立场，而是一个逻辑的结论，是基于对突变和选择的实际运作过程的仔细分析。

既然突变/选择不能阻止基因组的退化，那么基本公理就是错误的。不只是难以置信的，不只是不大可能的，而是大谬特谬。它不只是一个错误的公理，而是一个毫无根据而且已经被否定的假设，所以应当被抛弃。现在已经很清楚，突变加选择不能阻止基因组信息的丢失，显然更不能创造基因组。为什么？因为选择作用于整个生物体的水平，它不能阻止在分子水平上突变所带来的复杂信息的损失。这就好比用一把锤子来修理电子计算机。计算机芯片的微观复杂性使得锤子毫无用处。同样地，基因组突变的微观复杂性也使得整个生物体水平上的选择过程用处不大。

在第一章我们考虑了一个类似的过程，目的是改进运输工具。我们提议使用一个近乎瞎眼的机器人“记录员”，在汽车制造厂中负责在前期指导说明书中引入拼写错误。我们接着又加上一个近乎瞎眼的机器人“法官”在组装生产线的另一端负责质量控制。从来没有人进入这个工厂。我们要问，“从组装线上出来的汽车会越来越好还是越来越糟？”汽车会进化成宇宙飞船吗？我们所讨论过的一切问题应该使答案很明白了。汽车不会演变成宇宙飞船。事实上，汽车的质量会越来越低下。质量控制机器人最多只能延缓这个工厂不可避免的倒闭。在这

个比喻里延长时间会有所帮助吗？不会，无限的时间只能为系统崩溃带来无限的确定性。但我们的工厂即没有无限的时间也没有无限的资源。生命的现实就像是做生意，要么持续“负担一切成本”，要么倒闭（死亡）。

虽然有我们的知识精英对大众思想多年的不良影响，我还是相信多数人可以凭直觉明白，组装说明书中源源不断的随机拼写错误并不能把一辆汽车转变成一艘宇宙飞船。不管质量控制员等待多久，也不论说明书中发生多少拼写错误，他永远不会看到一辆带有火箭引擎的变异汽车。这是由于概率，也是由于贝希(Behe, 1996)所描述的“不可简化之复杂性”问题。就算质控法官看到一辆变异汽车上有火箭引擎，他（它）也不得不弃绝它，因为一辆带着火箭引擎的汽车还不是一种更好的交通工具。我们的法官（自然选择）的智商为零，没有任何远见，更不知宇宙飞船为何物。因而他（它）不懂得需要什么样的偏差会使一辆汽车变得更像宇宙飞船，他（它）也决不会选择这些偏差。朝宇宙飞船方向选择的唯一可能的途径是选择更好的汽车，这显然是自相矛盾的。就算我们的法官能够开始选择更像“宇宙飞船”的汽车（他绝对不能），要造一艘有功能的宇宙飞船需要如此天文数字的拼写错误，他需要选择到永永远远。但要记住，我们的汽车工厂没有永恒的时间，它今天就得付账。事实上，那些更“像宇宙飞船”的汽车就更不“像汽车”。换言之，它们会是功能不全的产品，类似于生物学中的怪物。这样的企业离破产不远了。在许多层面上的仔细分析一致揭示基本公理的绝对错误。

你们当中有谁愿意倾一生的积蓄投资于一家决定采用突变

加选择这样一种生产方式的新公司？它的发起人说所有的操作都是自动化。工厂不需要人力，而且他们保证工厂将制造出越来越优良的产品。切记，我们不是在讲一个另有资助的研发项目，我们讲的是一个旨在盈利的、或兴隆、或垮台的生产线！有人要投资吗？你我肯定都会说这是个骗局。尽管有耀眼的宣传册，又有重要人物的支持，我还是知道这样的生产线只能制造出越来越低劣的汽车，而永远不会生产出一艘划时代的新型宇宙飞船。

如果基因组是在退化，那我们的种族也是在退化。一个物种的衰老和一个个体的衰老很相似。两者好像都涉及突变的逐渐积累。突变发生于体细胞系，也发生于生殖细胞系。不管在哪一种情况，拼写错误积累到一个临界点，系统就开始迅速崩溃。结果导致个体的寿命有明确的上限。目前人类的预期寿命平均为70岁，最高将近120岁。然而，当亲表兄妹婚配时，其子女的预期寿命会严重缩短。为什么会这样？因为近亲繁殖暴露出基因组内还没来得及“浮出水面”的基因错误（隐性突变）。近亲繁殖就好像管窥我们作为一个物种的基因走向，近亲繁殖所生子女缩短了的预期寿命反映了基因组的整体性老化，并显示出一直隐藏着的长期积累的大量基因破坏（隐性突变）。如果所有这些基因破坏都突然显露（如果我们成为完全纯系近亲，在所有位点上都是纯合子），那肯定是致命的。我们必死无疑。我们的物种会立即灭绝。**基因破坏导致衰老，而衰老缩短寿命。这对于个人和群体来说都是一样的。**

如果这是真的，从逻辑上讲，过去必有一个时期基因组中有较少的基因破坏，因而寿命也长，而且近亲繁殖带来的害处

较小。这一点有无事实证明？

圣经记载，曾有一个时期，人们的寿命极长，近亲婚配完全无害。事实上，《创世记》中记载的预期寿命似乎难以置信。按照圣经的说法，最初的时候人们常规地活到900岁以上。从今天的角度理解，这似乎荒诞无稽。然而我们的视角和理解力都是有限的。我们仍然不知道为什么大多数哺乳动物的最高寿命都不超过20岁，而人类却可达120岁。即使是黑猩猩的最长寿命也不足人类的1/2。然而我们至少知道一点：突变显然与衰老有关。所以如果最初没有突变，难道你不觉得人类的最高寿命要长得多吗？在没有突变的时期，人活几百岁又何足为奇？这完全符合逻辑，为何一定要早死？

一篇由数学家和神学家合著的论文提供了一些有意思的数据 (Holladay and Watt, 2001)。文章考察了诺亚之后圣经中人物寿命的变化。数千年前圣经所记载的资料竟然清楚地揭示一个指数衰减曲线——一个典型的生物学曲线。我和我的同事们又进行了更完整的分析，其结果更令人振奋（图16，第150页）。

圣经资料中这一意想不到的模式令人惊叹。我们不得不承认，《创世记》、《出埃及记》、《约书亚记》及其他经卷的作者们，他们要么是忠实地记载了人类寿命的指数衰减，要么是共同伪造了需要高等数学知识的数据，而且存心要显示指数衰减。然而他们即不懂遗传学（19世纪的发现），也不知突变（20世纪的发现）为何物，为什么他们要显示生物学衰减曲线呢？把这些数据归因于数千年前的刻意欺骗似乎没有道理。最合理的结论就是这些资料是真实的，人类的预期寿命确曾有几百年之长，后来逐渐衰减到今天的寿数。鉴于以上的讨论，关于寿命衰减的最显

而易见的解释就是因突变蓄积而造成的基因退化。诺亚之后紧接着的几代人寿命下降得特别快，提示当时的突变率可能一下子升高了很多。

我和我的同事们为了深入理解人类群体中的基因退化，曾经做过成千上万次的数值模拟。使用已知的突变率和在生物学上合理的其他参数设置，我们一次又一次地看到：1. 大多数有害突变不能为选择所排除，所以会无限度地直线积累；2. 有利突变太过罕见，效应一般太弱，无力抗衡有害突变的蓄积；3. 适合度持续下降，呈现生物学衰减曲线，其轨迹与图16所显示的圣经资料非常近似。

多重证据提示，这个下行的曲线是很真实的。我们每个人都在走向死亡，我们作为一个物种也在走向灭亡。基本公理是个谬论。除非有智能的干预，信息总是会退化。除非有比自然选择更强的反作用力，基因组一定会退化。生命不是在向上提升，而是在向下沉沦。选择不会创造信息，最多它只能延缓衰退。

信息理论强烈地提示信息和信息系统只能通过智能的方式而产生，而且只能以智能来保存 (Gitt, 1997; Gitt et al., 2013)。计算机和计算机程序并不是自己形成的，而是经过艰辛的脑力设计出来的。甚至计算机病毒也不是像人们所想的一样无中生有，而是有人煞费苦心恶毒地设计出来的。互联网提供了一个大规模的实验室，让我们看信息能不能自行组织。显然不能。网上的一切，甚至于不好的东西，都是被设计出来的。

退化是信息的基本特性，这是随处可见的事实，从充满了细声细语的屋子，到层层传达的指令，到计算机的经常性死

机。我们的信息系统之所以没有退化得太快是由于精心的策划和智慧的设计，以求最大限度地稳定和保存信息。然而即便是设计得最好的信息系统，如果没有智能的维护和持续的智能干预，最终总是要崩溃，计算机一般来说用了5-10年后就成废品了。

我们可以把生物遗传系统看作智能设计的信息系统，把自然选择看作智能设计的稳定机制。尽管这些系统的设计都是异常精致，除非有那位设计者的智能干预，其衰败也是难以挽救的。

基因组的奥秘是什么？就是它的存在本身。在一个比不能见的微尘还小的空间里蕴藏着超乎想象的信息和复杂性。突变加选择绝不能解释这一切。很显然我们的基因组不可能是自然而然地出现的。如果它不是自发形成的，唯一合理的解释就是设计的结果。这难道不是一个令人敬畏、发人深省的奥秘吗？

2008年补充：“熵”这个术语有几种用法。我所用的是最常见的一种，即在没有任何智能干预的情况下，一切物体都要衰败的普遍趋势。基因熵特指基因组的衰败。它反映了在没有任何智能干预的情况下，基因组随时间流逝而退化的内在趋势。基因熵直接来自工程师和物理学家所正式使用的物理熵的概念。突变是物理熵在分子水平上的体现。它来源于随机原子力，也来源于DNA复制和修复的“纳米机器”运作失误。自然选择本身可视为通过过滤某些突变来降低突变熵的一种机械装置。正如所有的机器一样，影响DNA复制、DNA修复、和选择排除的生物机器都不能以100%的效率工作（机械失效也是熵的一种衡量方法）。所以，最正式意义上的传统熵是“突变熵”和“选择失效”的根源，后两者又是基因熵的根源。

“沙农熵”这个术语可能被有些人用来混淆基因熵的问题。不幸的是，沙农熵是带有误导性的一个说法，用于描述潜在信息的某些统计特性。它衡量一串字符内的一个字符的“意外性”。任何简单重复的模式都降低一串字符的沙农熵。一组随机编排的字母或者一首严谨的诗都可以有一个很高的沙农信息值！换言之，高沙农熵值既可能反映智能设计，也可能反映随机混乱状态。它在本书所讨论的问题中并没有什么用处。高沙农熵只是在说“这里没有一个简单重复的模式”。这个术语仅在非常有限和特别的意义上使用，而不能有效地适用于生物信息系统。它只可以用来故意地模糊我们对于基因组及其衰退的理解。沙农本人也提醒不要将他的这个术语用于生物系统。

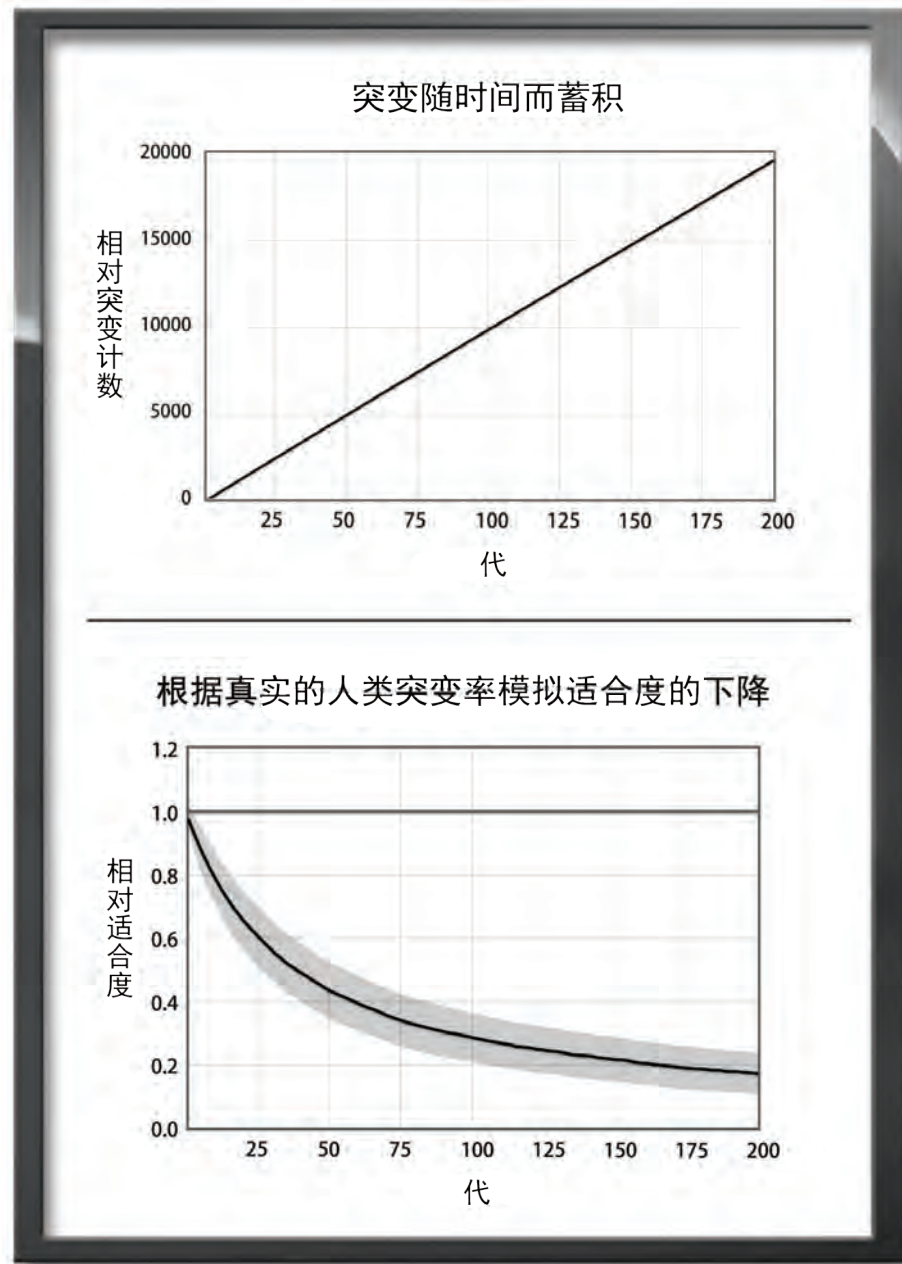


图14 使用符合生物学现实的参数对人群中的突变/选择过程进行的数值模拟

该数值模拟实验中将人类突变率设为每代每人100个，群体大小为10000人，假定“垃圾”DNA极少，并假设每1000个突变中有一个是有利的。

上：显示每人携带的有害突变计数随时间的变化（有利突变蓄积得太少，该图显示不出）。即使使用强力选择，也只是一小部分有害突变被排除。基因损害匀速增加。

下：适合度的下降（相对于起始群体）。纵轴为适合度，横轴为时间（200代大约为4000到6000年）。将突变率减半或将群体增大10倍并没有从根本上改变下行的趋势。该数值模拟实验得出的轨迹与图4、10a、10b、15、16十分相似。阴影反映群体适合度的标准差。

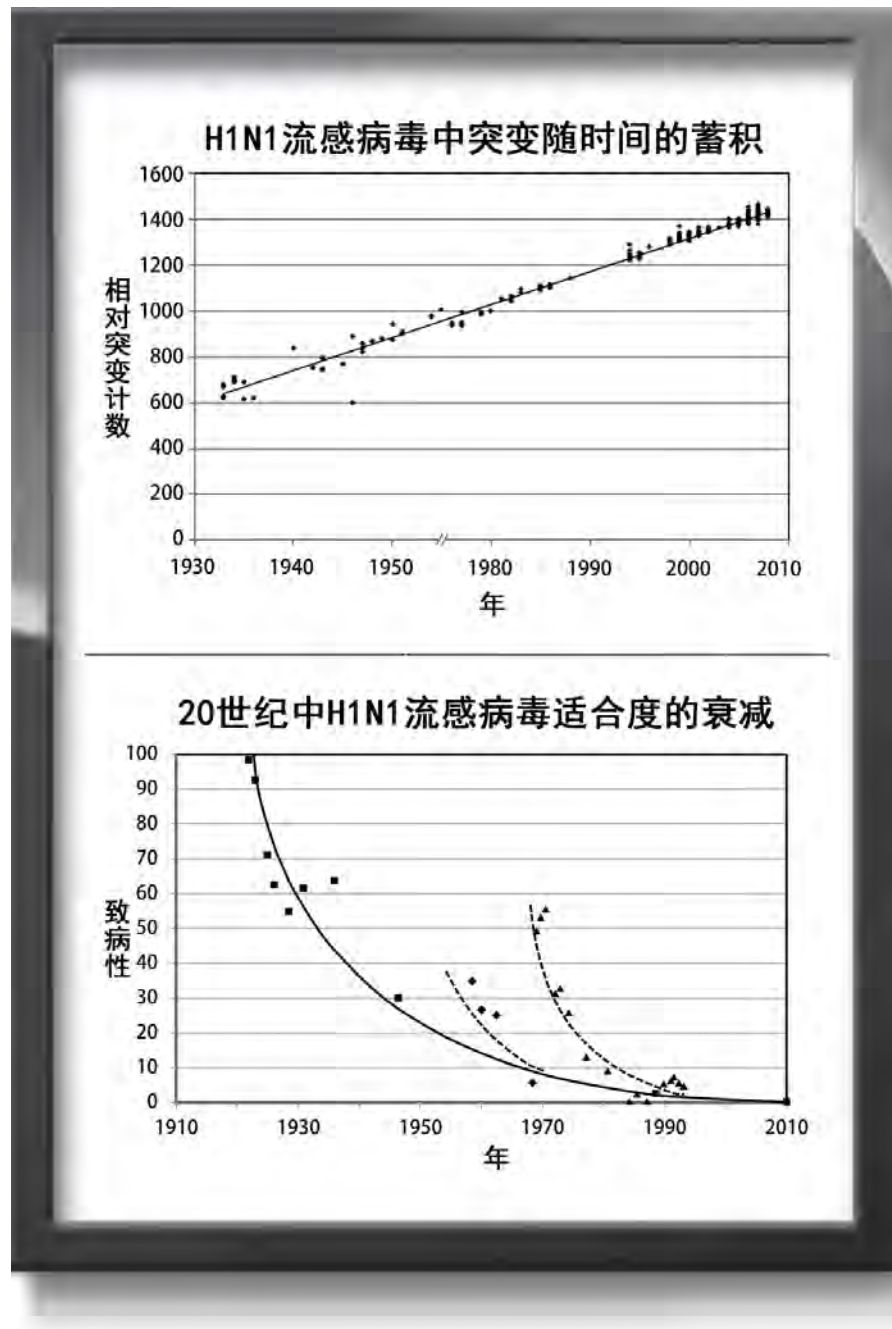


图15 生物学记录人流感病毒中突变的蓄积和和适合度的下降

上：该图显示自1918年大流行以来H1N1病毒株中突变的蓄积（病毒基因组的改变超过了10%）。突变蓄积的速度惊人地恒定，正如逻辑分析和我们的数值模拟所预测的一样（见Carter and Sanford, 2012）。

下：实线显示H1N1病毒株的致病性（即适合度），虚线显示另外两株曾在上个世纪引起大流行的流感病毒株的致病性（资料来自Simonsen, et al., 1998）。三株病毒（H1N1, H2N2, 和H3N2）均表现指数衰减。至2009年，最初的H1N1人流感病毒株似乎已经绝迹。本图为鲍勃·卡特博士(Dr. Rob Carter)所绘。

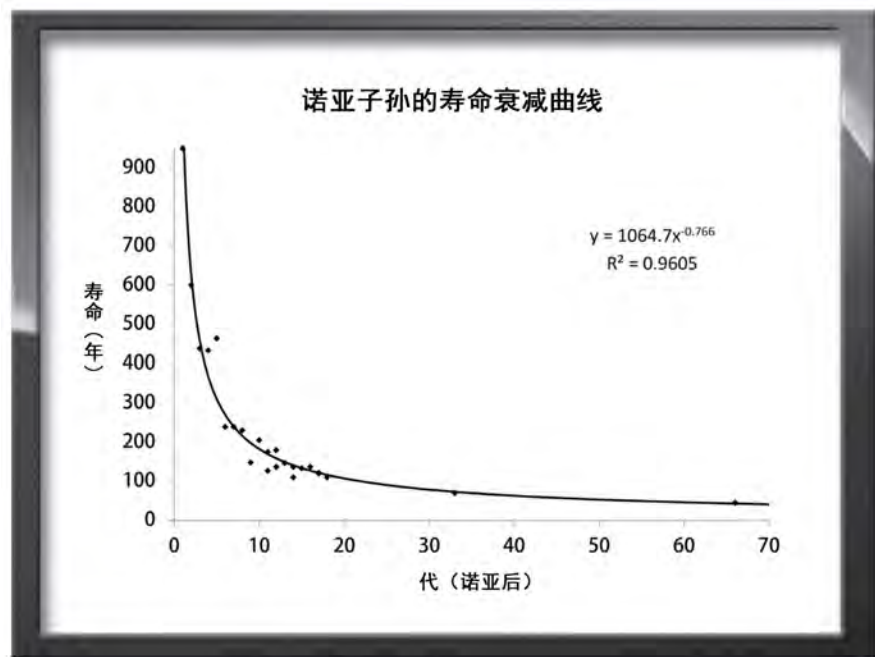


图16 圣经所载早期人类的寿命

圣经所载人类的寿命在诺亚之后发生急剧衰减。其模式表现为清晰的生物学衰减曲线。从资料中得出的最佳拟合线符合公式 $y = 1064.7x^{-0.766}$ 。该曲线与资料高度拟合，确定性系数为0.96(极限为1)。要说圣经资料为古人捏造似乎不可能。该曲线与突变蓄积导致基因组衰退的概念不谋而合，与图4、10a、10b、14所示的理论曲线及图15所示的生物学资料也十分相似。更多的对于圣经中犹太人的祖先及其年龄的分析，可见于LogosRA.org(文章标题为《Genetic Entropy Recorded in the Bible?》)。

atcgtttagc 第十一章 atcgtttagc

综述近期主要科学发现

简讯：基因熵的证据日益增强

自本书2005年初版以来，又有许多新的科学发现强烈支持本书的主题。这些新进展提示基因熵问题十分严重，基本公理已经难以立足。有些证据来自其他的实验室，大部分证据来自我和与我合作的科学家网络所进行的研究。我已在先前的章节中提及这些新进展，这里我希望将所有的新发现加以综述（重复的内容请见谅）。

我和合作伙伴得到的实验证据

1. 使用数值模拟验证基因熵

在过去的10年间，我和同事们研发了当今世界最先进的研究突变/选择过程的数值模拟程序(Sanford et al., 2007a, 2007b; Sanford and Nelson, 2012)。这个符合生物学现实的计算机模拟程序叫做《孟德尔的会计师》，我们用它进行了一系列仿真实验，有力地证实了本书中所有的逻辑结论。该程序忠实地模拟了一个符合生物学现实的群体，在每一代的个体中制造一系列符合生物学现实的突变。然后该程序模拟现实中的基因重组、基因复

制并繁衍后代。在后代中，根据其适合度排除一定的百分比，然后将这个过程一代一代地重复。影响选择效率和突变积累的所有相关的参数均可由使用者随意设置。《孟德尔的会计师》并不是为了一定要得出什么样的结果而设计的，它只是根据生物学家的理解而模拟突变/选择在自然界的运作方式。每一次特定的模拟实验的结果取决于使用者所设定的具体生物学情况。只要设置了一套生物学上合理的参数，基因熵就一定呈现（即有害突变的持续累积和随之而来的适合度衰减）。只有在使用了生物学上极不现实的设置时，才会见到上升性的进化。

1A. 大多数有害突变会无限度累积

如果使用生物学上合理的设置进行模拟实验，《孟德尔的会计师》会反复揭示以下事实：1. 选择只能有效地排除最糟糕的突变，其余的突变（大约90%）会以十分恒定的速度蓄积(Gibson et al., 2013)；2. 对于低影响力的有害突变，有一个可以测量的“选择门槛”，低于此门槛的选择压力实际上无效(Gibson et al., 2013)；3. 任何类型的生物学和遗传噪音都会提高测得的选择门槛，加重有害突变的积累(Gibson et al., 2013)。

1B. 大多数有利突变得不到放大

关于有利突变，《孟德尔的会计师》的数值模拟一贯地揭示：1. 有利突变也有可测量的“选择门槛”，选择对低影响力突变无效(Sanford et al., 2013)；2. 任何类型的生物学和遗传噪音都会提高有利突变的选择门槛，导致更多的有利突变不能被选择性放大(Sanford et al., 2013)；3. 在符合现实的设置下，

超过98%的有利突变得不到放大(Sanford et al., 2013)；4. 稀有的高影响的有利突变会被选择作用强力地放大，并得到迅速的固定，然而与此同时它会强烈地干扰对弱效应有利突变的选择。这就导致选择门槛的提高，得以固定的有利突变总数因而减少。除非在有利突变率很高、而且一些有利突变的效应很强的情况下，适合度总会衰减。即使有利突变源源不断，再加上强力的选择，最后积累起来的有利突变和有害突变的比例也不会反转。这导致功能性基因组的持续萎缩（有害突变累积的数目总是大于有利突变）。随着时间的推移，灭绝在所难免(Sanford et al., 2013)。

1C. 两个主要的“挽救机制”均被推翻

为了应对有害突变蓄积这个心照不宣的问题，人们提出了两个挽救机制（见第七章）——离开了有效的挽救机制，基本公理就很难保。

第一个挽救机制就是突变计数机制。这个概念就是要引导自然选择，让它的作用方式是基于个体有害突变的数目（而不是基于适合度）。这需要人为地将强力选择聚焦于有害突变数目较多的个体。该挽救机制的关键假设很简单——人为地假定一切突变的有害效应基本上相同。这倒是很方便，有了这个凭空的构想，近中性突变就不存在了，近中性突变的问题就消失了。该理论与现实相去太远，它要求大自然把所有的个体排队，衡量每个个体的突变数目，然后高度精确地将突变计数高的个体选择排除，这根本就不会发生。我们用数值模拟实验细致地研究了这种机制，即使让所有的突变都产生相同的效果（显然这不是自然界的实情），我们的模拟仍然清楚地显示，当运用符合现实的概率选择

时，突变计数机制会彻底失败。这方面的研究发表在最近的一篇文章中（Brewer et al., 2013a, 及本书104页图10a）。至此突变计数机制已经被彻底推翻，不应该再拿出来使用。

最近宣称解决了人类高突变率问题的文章似乎都在无意中再次搭建了条件，让人为的突变计数机制死灰复燃（见第七章）。这种效应是许多过度简化的模型所带来的假像，在现实条件下都会立即幻灭。

第二种挽救机制是协同上位效应。该挽救机制与第一种机制类似，但是使用了更多的迷惑手法（见第七章）。它也是让选择聚焦于突变计数比较高的个体，但它并不假设所有的突变都有相同的效应，而是认为当个体内的突变计数增加的时候，每一个突变的有害效应都增强了，如此就放大了针对高突变数之个体的选择力度。突变之间的这种负面作用被称为“协同上位效应”。大家都知道突变之间一般地并不是这样相互作用的。群体遗传学家发明这一花哨术语的唯一目的，就是试图编造出另一种貌似可信的挽救机制。我们做了大量的数值模拟实验，以验证这一人为的机制是否切实可行。即使使用了最慷慨的设置，协同上位挽救机制也无力阻止有害突变的积累。该机制非但没有延缓适合度的下降，反而加快了退化和灭绝的进程（这是逻辑的必然，本书早先的版本也是这样预测的）。这些发现见鲍姆加德纳的论文（Baumgardner et al., 2013）和本书第106页的图10b。这第二种挽救机制至今也被彻底推翻，不应再拿出来使用。此外再也没有人提出什么可信的挽救机制，

基本公理已经完全站不住脚了。

1D. 否定Avida

Avida是一电脑程序，当初研发这一程序的目的就是模拟进化。据称，该程序已经证明突变/选择系统能够在仿真条件下合成新信息。Avida并非真正的数值模拟，而是更像一个模拟电脑游戏，能够随机产生并选择各种各样的计算机指令字符串。这与进化有明显的相似之处，据说它已经验证了基本公理。从表面上看，用Avida所做的实验与《孟德尔的会计师》的模拟结果相对立。鉴此，我们将这两个计算机程序进行了比较。

《孟德尔的会计师》是迄今研发出的最符合生物学现实的对突变/选择过程的模拟，相比之下Avida与生物学现实南辕北辙。Avida中最重要的人工特色就是其选择性奖惩系统的运作方式。一个“突变”只要被Avida认定是有益的，其繁殖率（Avida的“适合度”）立即就人为地加倍。反之，如果一个突变被定义为有害的，其繁殖率立即就人为地减半。当出现多重突变时，Avida就将奖惩效应相乘。使用如此极力夸张了的奖惩效应（比实际情况高出许多个数量级），Avida中的选择变得极其有效。只需要几个有利突变就能将适合度放大一百万倍。有意思的是，当类似的选择性奖惩机制使用在《孟德尔的会计师》里时，也可以见到同样的爆炸式进化，适合度也会放大一百万倍以上。用了这样极不现实的设置，两个程序都会显示爆炸式进化。但由于这种设置与现实不匹配，这种爆炸式的适合度上升在生物界从来没有见到过。当我们把Avida的设置加以修

改，以反映符合生物学现实的选择性奖惩机制时，前进性进化就突然变成零了(见Nelson and Sanford, 2011)，连一个有利突变也选择不出来，连一个字节的的信息也创造不出来。如果开始时使用Avida内在的默认设置(极高的奖惩效应)，后来再调整到符合生物学现实的设置，群体会发生什么变化呢？在使用了不现实的设置时，我们确实能够将所有可能的有利突变放大，并淘汰所有可能的有害突变。结果适合度很快就会提高一百万倍以上。然后，当把设置调整到现实水平上时，之前进化出的功能又迅速消失了。这基本上等同于灭绝——新创造出的“信息”重归零点。只要使用了符合生物学现实的设置，《孟德尔的会计师》与Avida都证明基因熵的真实性，这着实令人惊奇。以上结果发表于两篇文章(Nelson and Sanford, 2011; Nelson and Sanford, 2013)。在使用了符合生物学现实的设置时，Avida验证了基因熵的真实性，并有效地否定了基本公理。

1E. 验证“霍尔丹困境”

在1950年代，DNA立体结构刚刚发现不久，一位群体遗传学领袖(也是基本公理的创始人之一)就意识到一个问题。霍尔丹意识到即使有源源不断的有利突变，自然选择将其放大并在群体中固定的能力也必然是十分有限的。根据他的计算，在像人类这样的哺乳动物群体中，假如进化群体的大小为10000，在六百万年内(目前估计从猿进化到人的时间)大约只能选择性地固定1000个有利突变。从基因组研究的资料中，我们现在知道人类基因组与猿类基因组有着深层次的差别。在这一段进化历史中需要发生并固定数千万甚至上亿的有利突变，这就需要每一代固定大

约1000个相互独立的有利突变。霍尔丹认为这是不可能的；他估计每300代最多只应该固定1个突变。这个问题叫做“霍尔丹的困境”。沃特·赖敏(Walter ReMine)曾对此进行了详细的研究，他使用了与霍尔丹完全不同的数学运算，但得出的结论却完全一致(ReMine, 2005)。我们现在有了第三种研究这个问题的方法，即数值模拟实验。我们的结果有力地验证了霍尔丹和赖敏的工作。实验发现，在三万代(大约相当于600万年)之内，人类实际上只能选择固定数百个至数千个突变(具体数字取决于参数设置，见Rupe and Sanford, 2013)。我们相信，数值模拟实验是理解这个问题的最佳方法。从霍尔丹、赖敏到我们的工作，结论已经很明显：从猿到人的故事毫无可行性。

1F. 正在进行的研究

数值模拟还在进行，其结果将继续验证基因熵的真实性，并昭示基本公理的坍塌。尚在进行中的工作会理清几个重要问题。其中一个就是在创造具体的核苷酸系列的时候，有一个不可逾越的“等候期问题”(信息不仅仅需要字母，而是需要具体的像句子一样的字符串)。另一项工作显示，要同时选择许多有利突变，也有明显的、可测量的“选择干扰”。我们的工作很快还会显示，正在基因组中数百万个位点上蓄积着的隐性有害突变，最终必将导致极为严重的适合度危机。当这样的突变漂移到足够高的频率时，纯合体便开始出现，其毁灭性的基因负荷终将浮出水面。最后，我们很快还会显示，群体选择不能解释利他行为。

所有这些经验性的模拟已经得出(而且正在得出)本书早先的版本所预言的结果。这说明本书的逻辑是坚实的，并带

有预测性。数值模拟实验一贯地证实本书中所论述的逻辑推理 (Baumgardner et al., 2008; Baumgardner et al., 2013; Brewer et al., 2013a; Brewer et al., 2013b; Gibson et al., 2013; Nelson and Sanford, 2013; Rupe and Sanford, 2013; Sanford et al., 2007a; Sanford et al., 2007b; Sanford et al., 2008; Sanford and Nelson, 2012; Sanford et al., 2013)。

对突变/选择过程的完整的数值模拟必将带来群体遗传学上的革命。我们终于能够实际地看到突变/选择过程在任何具体状况下是如何运作的了。我们再也不局限于抽象的数学运算了。这些抽象运算都是建立在一层又一层不合理的、简单化的假设上的。随着人们逐渐认识到这一点，群体遗传学界必将发生翻天覆地的改变。另有一个研究团队也在对突变/选择过程进行全面的数值模拟，他们写道：

“为遗传过程建立计算机模型可以看作是细胞自动化研究的一个前沿分支，Wolfram称之为……‘另类的新科学’……Wolfram曾经显示，任何由相互作用的元件所组成的系统都会排列成一些模式，对这些模式很难进行数学推测，但用计算机来复制和研究却是轻而易举的。”(Qui et al., 2014)

这正是我们在自己的数值模拟工作中所看到的。如果我们总是试图将复杂的、动态的生物群体硬塞进一些可笑的、过度简单化的数学公式里，那么群体遗传学里关于“整体图像”的问题就永远无法解答。只有通过完整的数值模拟，我们才能理清大问题。所以，整个学科需要重新从头开始。

2. 在真实的群体中验证基因熵

数值模拟实验必须与现实世界中的生物学观察相吻合。寻

找这种资料很难，因为基因熵只有在很多世代过去以后才能清楚地表现出来。这对于长寿的生物，如哺乳动物来说，显然是个问题——一个基因熵实验至少需要持续几百年。即使在细菌中也是如此，因为细菌虽然繁殖期短，但是每代的突变率也很低。要在生物学研究中观察基因熵的证据，我们需要一种繁殖期短而且突变率相对高的生物。RNA病毒特别适合这一目的。

2A. 关于H1N1流感病毒株基因退化的记录

我们追踪了曾经在1918年引起大流行的人流感病毒株基因退化的资料 (Carter and Sanford, 2012)。在查验病毒基因退化的历史证据之前，我们曾做过初步的数值模拟实验。我们使用《孟德尔的会计师》，模拟了一个(与流感病毒相似的)假想RNA病毒中突变积累的情况(Brewer et al., 2013b)。初步结果强烈地提示，RNA病毒会以恒定的速度很快地积累有害突变，适合度下降，并在相当短的时间内走向灭绝。由此我们推测人流感病毒H1N1株会呈现类似的模式。H1N1大约在100年前侵入人类，在第一次大流行中造成5千万到1亿人死亡。在过去几十年中人们采集了H1N1的样品，并在过去10年中完成了对样品中RNA的测序工作。我们使用重塑的1918病毒基因组，将原始毒株与所有后来的基因型进行比较，这样我们就可以分析该病毒在过去一个世纪中突变积累的情况，并借此来检验我们的预测(Carter and Sanford, 2012)。分析结果发现了突变积累不仅迅速，而且积累速度惊人地恒定。突变的增加与“适合度”(致病能力)的持续衰减紧密相关——直至2009年该毒株最终灭绝(届时病毒基因组已经改变了10%以上)。这些结果总结在图15中(第148-149

页)。这些结果不仅验证了基因熵在现实的生物群体中的真实性,也验证了《孟德尔的会计师》预测生物群体中基因退化时的准确可靠性。

2B. 重新审视著名的林斯基大肠杆菌实验

从1988年以来,林斯基(Lenski)和他的同事们一直在从事一项长期的研究,旨在观察大肠杆菌在实验室条件下的前向进化(Barrick et al., 2009; Lenski, 2011)。几十年的时间实在不算长,尽管细胞分裂了5万多次。宣称本实验在某种程度上涉及“深度时间”或反映宏观进化是不准确、不合理的。在几十年的时间内,一株大肠杆菌对某种人工培养基的适应性略微有所改善。这并不出人意料,也没有多少意义,不过是反映了微调。细菌还是同一种——大肠杆菌。由于林斯基的实验时间只有H1N1历史的几分之一,又由于细菌中每个核苷酸位点的平均突变率至少比流感病毒低1000倍,流感病毒群体比林斯基的细菌群体更反映了“深度时间”。在林斯基群体中,每个细胞累积了数百个突变,而在流感病毒群体中,每个病毒颗粒蓄积了将近2000个突变。病毒基因组突变了10%以上,而细菌基因组的改变微不足道(一个由400多万碱基对组成的基因组只发生了数百个突变)。要观察到像H1N1病毒那样的极度基因熵,林斯基的实验需要进行几个世纪。

然而,如果仔细地观察,林斯基自己的数据仍然揭示出明显的基因退化的证据(Rupe and Sanford, 待发表)。当我们考察了使得大肠杆菌适应人工环境的具体的基因改变,我们发现,适应性的建立无一例外地是缘于突变导致的功能丧失或调节失

效(即信息损失)。由于一半以上的大肠杆菌基因在林斯基实验所用的静态人工培养基中毫无作用,细菌不断地将这些临时无用的基因关闭,把无关生存的功能丢弃,直到细菌基因组萎缩到最低的功能状态,这个应该需要几个世纪。如此以来细菌就会达到最节能的高效状态——由于功能性基因组缩小了许多,细菌能够在固定的人工环境中以最大的速度生长。但这不是前进性进化,而是适应性退化,或称为收缩性进化——实际上是基因熵的一种形式。“进化中”的菌株最后会退化到只能在特殊的、极其有限的、完全人工的环境中生长。事实上,这样的菌株正一步一步地变成残废,在真实世界里的适合度越来越差。虽然几个造成功能丧失的突变有助于提高在人工环境中的适应性,数目更多的其他突变却也在无情地累积(因为没有自然选择),而这些突变的效应都是观察不到的。我们很有理由假定大多数突变是接近中性但轻度有害的。几百年之后,林斯基“进化中”的菌株一定会灭绝(如果能够在人工环境中保存那么长时间)。近中性有害突变将使细菌基因组不可避免地“锈坏”(Gibson et al., 2013; Sanford et al., 2013)。

3. 我和协作者从研究中得出的支持本书主题的其他证据

3A. 重叠密码

目前已经有强力的证据,表明在基因组中的任何一段DNA中,都有多个相互重叠的信息。这好比收到一封电子邮件,从中读取有用的信息,然后发现其中有一部分,在反过来读时还有另一个有用的信息。其中还有一段文字,每隔一个字读一个字

时，又可以读出第三个有用的信息。我们现在知道人类基因组的一半以上都是从正反两个方向上阅读的。一个基因使用不同的起始位点和终止位点平均大约产生七个转录产物（RNA）。任何一个转录产物，还可以经过不同方式的剪接而编码多种蛋白质。我们已经指出过，这样多种重叠的密码会严重影响真正有利的突变出现的可能性。大多数“有利”突变虽然可能改善一个信息，却会打乱一个或多个与之重叠的信息。这样的突变并非真正有利的，至少其利处是模棱两可的。鉴于这一点，生物学界必须大幅度地下调对真正有利突变率的估计（Montañez et al., 2013）。明确地有益而且影响力强得可以选择的突变一定少之又少。

3B. 推翻“费歇尔”定理

我们正在撰写一篇论文，意在正式地否定费歇尔(Fisher)“关于自然选择的基本定理”。费歇尔和他的定理是新达尔文主义赖以建立的关键。费歇尔最大的失误在于他关于所有突变的总效应为中性的信念。他认为有利突变与有害突变一样地常见，并有一样的影响力。当时人们并不明白基因中的信息就像说明书一样，而突变就是说明书中的随机拼写错误。费歇尔不明白绝大多数新突变必然是有害的，而可能出现的有益错误一般地只有很微弱的益处。由于突变的总效应是负值，所以费歇尔的中心前提是错误的，他的理论已经从整体上被否定了。费歇尔的定理并不适用于真实的生物界。这意味着新达尔文主义理论的首要的最根本

的柱石之一已经证明是个谬误。

其他科学家得出的新证据

1. 《生物信息——新视角》是最近预示新达尔文主义理论覆灭的最重要出版物

这本书是最近在康奈尔大学一次学术会议的论文汇编，刊载了29位知名科学家的研究成果。这些作者来自许多不同的科学领域，他们一致同意基本公理不能解释基因组的起源，更不能解释活细胞中任何其他类型的生物信息系统的起源（见Marks et al., 2013，其摘要可见 Sanford, 2014）。

2. ENCODE项目的巨幅报告(2007, 2012)是主流生物学研究最高水平的成果

生物学上的ENCODE项目堪比美国航空航天局(NASA)的太空研究计划，汇总了数百位主流科学家的研究发现。ENCODE记载着“垃圾DNA”概念的覆灭。大量的证据显示人体基因组中大部分是有功能的。这意味着大多数人体突变不是中性的，而是有害的。这是敲响基本公理丧钟的又一得力证据。ENCODE还显示了重叠密码的大量存在，这意味着明确的有利突变必然少之又少。这些都说明还有大量的信息目前尚不能解释，而基本公理的解释潜

能显得更加不足。

3. 又一篇有影响力的论文证实人类基因组在迅速退化

最近林奇的一篇文章(Lynch, 2010)详述了人类突变的性质, 以及突变造成的基因熵在人体基因组中的表现。作者尽管争辩说大部分基因组是垃圾DNA, 他还是承认问题的严重性(见附录1)。

最后的总结: 在我写本书第一版的时候, 想不到在不到十年的时间里, 会出现这么多的证据支持本书主题的方方面面。这既令人惊叹, 又发人深省。

atcg

后记

希望在何方?

简讯: 有希望

既然您已经辛苦地跟着我的思路读完了这本书, 现在请容我偏离科学的论述, 转而触及哲学的层面。论到希望之所在, 我愿意将自己的结论谦恭地提交给您。

我还年少时, 就接受了自己要死去, 并且我所爱的人都要死去这一事实。我接受了死亡的概念, 但对死亡的预期至少夺走了我的喜乐。老师教导我说还有一线希望之光: 那就是世界会变得越来越美好。科学在发展, 文明在进步, 甚至人类也在向上提升。通过我们的努力, 我们能让世界更美好。通过进化, 我们的后代会演变成更高级的生命形式。由于进步, 也许我们最终能够战胜死亡。或许有一天我们能逆转宇宙之退变! 我个人的希望就是, 也许我能为人类的进步贡献微薄之力。我相信我们这一代人大都存有这一基本的希望。*

*木村, 1976: “难道我们仅满足于在这一宇宙之微尘上保存自身, 当一例最高级的活化石吗? 或者我们是否应该尽一切努力自我改进成超人, 甚至更高的生命形式, 扩展到更广的宇宙空间, 借以表明生命不仅仅是一个毫无意义的插曲?”

而今我相信这只是个虚假的盼望。我仍相信我们应努力使世界变得更美好，要做我们被赋予之世界的负责任的管家。但我看到我们努力的结果最多是维持现状。虽然科学有望延长寿命，但不能战胜死亡。衰退是必然的趋势，我们的身体、我们的种族、甚至我们的世界都将死去，要停止这一根本过程绝不是你我力所能及的。环顾四周，这岂不是是有目共睹的吗？既然如此，希望在哪里呢？如果人类的基因组在不可逆转地退化，我们必须在进化之外寻找对未来的希望。

有一位审稿人对我说这本书的主题既可怕又可悲，他认为我有点像泰坦尼克号上的一位施虐狂管家，四处兴奋地宣布着轮船要沉没的消息。但事实并非如此。我恨恶熵的效果（就是退化）。我不愿看到自己的退变，不愿看到亲属恶化的健康状况，也不愿看到新生儿的畸形。这些绝对地是目不忍视，但又绝对地是不可否认。泰坦尼克号上的真正管家当然有责任通知大家轮船在下沉，即使有人会因此恨他。我以为自己就在他的位置上。有责任心者听到凶信会表示感激，进而建设性地应对。假若我们寄一切希望于一艘正在沉没的船上，是否应当及时醒悟而明智地放弃虚幻的希望？只有如此，我们才能理解凶信的用意。只有理解了凶信我们才能真正地听从喜讯：这里有救生艇。

正如我们不会创造生命，我们也不可能战胜死亡。然而有一位创造生命者，他也是基因组的设计师。我不知他是如何创造的，但他既然制造了硬件，就必然写出了最初的软件。他被称为“生命的主”（《使徒行传》3章15节）。生命的主必有能力战胜退化和死亡，这就是喜讯。

我个人深信耶稣是我们的希望。此外，我们没有别的盼望。他最初赋予了我们生命，所以今天能赐我们新生命。他最初创造了天地，所以将来能创造出新天新地。因他从死里复活，我们也会从死里复活，尽管死亡已经在吞噬着我们。真实的希望就在这亦深亦浅的道理中。因为这希望是根植于那永恒者，所以是不可动摇的。这希望经历了时间的考验，克服了宗教的腐败。凡愿意接受的，都可以白白地得到。在进化论破灭之时，我谨提此见解，供您考虑：耶稣是唯一真正的希望。



1. 基因熵网站

GeneticEntropy.org 低价出售本书英文版。上面也有补充、辅助材料、相关链接和“评论”栏。也会有各种语言的译本。本书英文版在Amazon.com 上有印刷版和电子版。

2. 《生物信息——新视角》

凡是对批判新达尔文主义的新证据有兴趣的人，都应读新书《生物信息——新视角》(Marks et al., 2013)。Amazon.com 有售，在BINP.org 也减价出售。其中各章可在worldscientific.com 网站免费下载。《生物信息——新视角，总结和评论》是总结性的小册子，在BINP.org 免费以pdf 形式下载，也在Amazon.com 上以Kindle 版销售。

3. 《孟德尔的会计师》

本书所描述的简单逻辑有力地驳斥了新达尔文主义理论。然而，基本公理如今在人们的头脑中已根深蒂固，有些人会不满足于逻辑分析，而索要其他的证据。但除了逻辑，我们还有没有客观的经验性的方法来检验新达尔文主义理论，以做到让有推理

能力的人满意呢？

要客观地分析新达尔文主义理论，只有一个有说服力的经验性的方法，那就是“数值模拟”。在真实的群体中，亿万突变同时出现并在后代中分离。这使得突变/选择过程复杂得难以置信。由于这种复杂性，要理解这个过程的唯一方法就是系统地追踪人群中发生的每一个突变（如同一个会计师用表格追踪一个大公司的每一项财务交易一样）。这就是所谓“数值模拟”的实质。数值模拟用于遗传系统中可称作“基因会计”。

“孟德尔的会计师”这个程序就是为此目的而设计的。这是第一个为群体遗传学设计的符合生物学现实的前瞻性数值模拟程序。该程序集研究和教学工具于一体。当使用一组合理的生物学参数时，“孟德尔的会计师”会给出强有力的实证数据，证明基因组随时间的退化，和进化遗传学理论中一切缺陷的真实性。诚实的数值模拟会有效地否定基本公理，其明确性让任何开明的人都信服。

“孟德尔的会计师”可用苹果机或PC机运行。读者可免费下载(MendelsAccountant.info)，网上配有详细介绍和使用说明。

atcg

附录 1

一个根深蒂固的观念

“显然，无所不能的自然选择可以成就一切，解释一切……”

以上陈述出自一位早期的达尔文主义者，其具体来源已无从查考。然而这句话背后的哲学思想无处不在，所以具体来源并不重要。任何一个达尔文主义者都可能说出这样的话。事实上，15年以前，我自己就可能这样讲。在以上“信仰表白”发表了一百多年之后，基本公理依然控制着大多数科学家的心智，让他们忠贞不二，尤其遗传学家们更是如此。然而，当我们回顾一些最有名的群体遗传学家的著述时，可以看出他们对基本公理的忠诚似乎主要出自意识形态上的信仰，而不是基于证据。一次又一次地，他们自己的详细分析最终与公理相悖。他们对公理的信念似乎不是基于他们的证据，而是无论证据如何，“此心矢志不移。”因此，他们需要不断地解释，为什么公理尽管与他们的数据相悖，更与常识相悖，却仍然是“不证自明的”。

以下所有的语录都出自杰出的进化遗传学家（霍伊尔除外，他是著名物理学家）。我认为这些科学家都是很有才华的，但同时我也认为他们的工作（或许还有他们的生活？）是建立在一个虚假的公理之上的。我已从他们的文章中抽取了一些具体的言论，这些言论并不反映他们自己的哲学思想，但都指向我所提到的问题。

今天，很少有哪个大学里的遗传学家愿意公开讨论突变/选择理论的弱点。事实上，我怀疑大多数遗传学家甚至从未认真考虑过这些弱点（虽然我知道有许多怀疑论者选择了沉默）。大多数遗传学家从未认真地质疑过基本公理。这是因为，他们一直凭信心把它当作不证

自明的公理来接受，就像我自己从前一样。因此，他们还未曾以开放的头脑和心态对其全部问题进行审查。每一项研究、每一篇论文都似乎是为了使这一理论行得通而设计。在遗传学界，对基本公理的信仰力量是巨大的。然而，其中有一些人（我认为是那些对基本公理最有信心的人）已经公开承认了一些具体的问题。当这些已被承认的个别问题加起来时，就形成了对基本公理的强力挑战。以下语录说明了这一点（黑体字是我认为需要注意的）。

穆勒的恐惧

Muller, H.J. 1950. Our load of mutations. *Amer. J. Human Genetics* 2:111-176.

“这将意味着突变基因不断堆积……退化完全无法辨认的形式，群体中每一个体和其他个体的基因都不相同，而且毫无规律……最后不如选择适当的原材料重新造人，倒比用祖先留下来的那堆可怜的遗传物质打造出人形更容易，也更合理。因为这些遗传物质彼此都非常不同，每一个基因组都带来一系列最复杂的研究课题……如果被淘汰的20%是被迫放弃生存……其余的80%尽管已经成功地繁衍，但整体上讲与那些灭亡的废物大同小异……实际上在原始条件下，它们注定都要灭亡……很难评估基因退化过程的速度……”（第146-147页）。

“目前因选择放松所导致的基因退化比通常想象的其实要快得多……很显然，人类的自发突变率是如此之高，而其生育率又如此之低，以至于根本没留下多大的选择余地……如果 u 值取最小的0.1……每人就平均需要生育2.4个孩子……这还没考虑非遗传因素造成的死亡和不育……那么很显然，目前每对夫妇的生育数量不足以让选择跟上0.1的突变率……如果情况更糟糕一点， u 接近0.5……，我们目前的生育水平将与人类的需求相去甚远。”（第149-150页）

穆勒博士是迄今为止唯一获诺贝尔奖的群体遗传学家。穆勒计算出当时（1950年）人类的生育率无法应对0.1的突变率。后来我们了解到实际突变率比他所估计的至少高1000倍。而且，生育率已经急剧下降。

“……现在的遗传负荷是严峻的……若再增加25%……后果将是严重的，而若将其加倍，后果则是灾难性的……如果 u 值超过0.5，所需的选择淘汰数量……就如我们已经看到的，比原始人类所容许的有效繁殖率还高……基因组成将不断退化，人口数量将不断减少，直至灭

绝。”

1950年穆勒就得出结论，如果突变率高达0.5，早期人类不可能进化，而是将退化并灭绝。但现在已知的突变率却比这要高200倍！

“在许多文明国家，出生率下降到了平均每个家庭不超过两个，即使存在强力的选择作用，突变率的上限（超过它则无法保持平衡）也必须大大低于0.5，正如我们已经看到的，也许要低于0.1。”（第155页）

“如果我们现在假设有高突变率、低出生率的情况……这一下降趋势最终能否被遏止将很成问题。”（第156页）

穆勒的意思是，在突变率高达0.5时，即使我们制定最严格的选择方案，现代生育水平已不足以制止基因退化。自1950年以后，我们已经了解到，我们的实际突变率大约是100。

霍尔丹的困境

Haldane, J.B.S. 1957. The cost of natural selection. *J. Genetics* 55: 511-524.

“众所周知，育种者很难在一群牲畜中同时选择所有想得到的品质……在本文中，我将尝试定量描述这一相当明显的道理，即不可同时对多个性状进行高强度的自然选择……”

“我怀疑这种高强度的选择在进化过程中是否常见。我认为 $n = 300$ （300代）是个更可能的数字，而这个数字将使 $I = 0.1$ （群体中选择淘汰的个体总数占10%）。”

“如果两个物种在1000个位点上不同，而基因置换的平均速率是每300代发生一个，那就需要至少300000代（6百万年）……”

“使用这个过程对数千个位点进行选择即便是地质时间也嫌太短……同时选择多个基因能避免这样的慢速度吗？我对此表示怀疑……”

“……要用自然选择的方式确保一个基因的置换，所需的死亡数量大约是一代中全部生物体数量的30倍……每个基因置换需要花费的平均时间约为300代。”

“我深信，像这里提出来的这种定量论证，在今后所有关于进化的讨论中都应该使用。”

霍尔丹是第一个认识到选择有代价的人，而这个代价限制了对选择的实际期望。他并没有完全意识到他的观点对进化理论造成的巨大难题。他计算出，对于人类，要固定仅仅1000个突变就需要6百万年时间（假设每一代为20年）。他那个时候不可能知道基因单位的实际数目是30亿，而且任何可能进化成人类的种群中每一代都发生至少一百万新的基因突变，其中大部分都需要筛选。人与黑猩猩之间至少有1.5亿个核苷酸的不同，这代表至少4000万假想的突变 (Britten, 2002)。如果人是从类似黑猩猩的物种进化而来，人类谱系

中至少固定了2000万个基因突变（4000万除以2），但自然选择只能筛选了其中的1000个。其余的必定是靠随机漂变固定下来的，造成了数以百万计的近中性的有害突变。这不仅意味着我们比类似黑猩猩的祖先更低等，而且我们早该死绝了。霍尔丹的论文发表后，学术界曾有多次努力要把霍尔丹的困境扫到台面以下，然而问题至今还是一样。赖敏 (ReMine, 1993, 2005) 对此问题做过全面的研究，并用一种完全不同的数学公式进行了分析，得出来的结果相同。我们用数值模拟实验进一步验证了霍尔丹的困境，得出了不容辩驳的结论 (Rupe and Sanford, 2013)。

穆勒的棘轮

Muller, H.J. 1964. The relation of recombination to mutational advance. *Mutation Research* 1:2-9.

“虽然有某种程度的优越性，但这种优越性太小以至于不能被选择有效地利用，可以说它的声音被噪音淹没了。在不同环境下（遗传的，生态的等等），其优越程度必然大不相同，但对这一问题的分析至今仍少得可怜……虽然对它进行研究应该是很有价值的。”

穆勒预料到了近中性突变的问题，但却没有看到它对进化论造成的严重困难。他的确认识到了，许多不同的环境因素（不只是群体规模）会使这一问题扩大。

“看起来一个不进行重组的物种似乎将……会发生基因退化……可能大家都这么认为……但是，最近本作者在一片文摘中提出的这一结论有可能误导读者，其实这一结论的意义仅限于人为的概念化……然而……一个无性群体含有一种棘轮机制，造成……一些谱系背负着越来越重的突变负荷。”

在承认很快就以他的名字命名（“穆勒的棘轮”）的这个问题时，穆勒是多么极端地谨慎啊！在他的整个职业生涯中，穆勒深切地关注着辐射诱发的突变和人类基因的退化，他还是优生学的主要倡导者。然而，即使他想警告公众基因退化的问题时，他也是表现得极其小心，避免做出任何可能会损害进化理论之“确定性”的言论（这显然反映了他最忠诚于什么）。他关于无性繁殖的物种中基因退化问题的观点（他正是因此而成名）措辞如此谨慎，以至于人们很难辨别他想要传递的信息。

穆勒称，他的“棘轮”适用范围有限，因为他认为突变是极为罕见的。因此，他认为每个突变都可以被看作一个独立的、可以明确选择的单位，并可以一个一个地消除。现在我们知道基因突变是大量的

和弥散性的，“一次一个”的消除机制是不可能成立的。

如果我们将穆勒的近中性（即不可选择的）突变的观点与他的突变进展的观点结合起来，我们会看到没有哪种选择系统可以阻止穆勒的棘轮，哪怕在有性繁殖物种中也不行。在无性繁殖物种中，每一个克隆都会一代一代地增加突变数目。即使在群体中选择最好的克隆用作下一步繁殖，每一个亚克隆还是要比初始克隆带有更多的突变。所以信息只能退化。这是一个单向（棘轮式的）过程。新一代个体的平均突变计数必然要高于前一代。具有讽刺意味的是，我们应该认识到这不仅发生于无性繁殖的物种中，在有性繁殖的物种中也是一样的。“棘轮”之所以运作是因为有利突变很少，而且每一个有利突变总是与大量的有害突变紧密连锁的。选择不能将少数有益突变与多数有害突变分离开来，因为它们共存于大连锁块中，无法分开。因此，按照穆勒的棘轮理论，基因组的每一个部分（每一个连锁块）都必将分别退化。

木村的疑惑

Kimura, M. 1968. Evolutionary rate at the molecular level. *Nature* 217:624-626.

“……在哺乳动物的进化史上，核苷酸的置换是如此之快，群体中平均每两年就有一个核苷酸被置换。这个数字与霍尔丹著名的估计……即每300代固定一个核苷酸置换，形成了鲜明的对比……”而且，“……按照每两年一个置换的速率……置换负荷变得如此之大，以至于没有哪种哺乳动物可以承受……”“这就让我们得出了一个相当惊人的结论……即中性突变的发生率为……每个受精卵平均有4个……”

木村估计的突变率比实际情况低25-100倍。但令人惊叹的是进化理论家们是那么容易地包纳任何新的数据，他们似乎有一个无限灵活的模型，允许对其细节进行持续的无限制的发挥和修改。

Kimura, M. 1983. *The Neutral Theory of Molecular Evolution*. Cambridge University Press. p.26.

“该公式比霍尔丹的公式得出的选择代价要大2倍……假设大多数在分子水平上的突变置换是经过正选择得来的……要在突变置换的同时保持群体数量不变，每个亲代个体必须生下……327万个后代，以保障生存和繁殖。这是我在提出分子进化的中性突变/随机漂移假说时使用的主要论据……”

木村认识到霍尔丹是对的，选择确实非常缓慢，而且能同时选择的突变数量是有限的。木村意识到他那个时代所有的进化论专家都过分依赖于选择，而且考虑同时选择的位点也太多，从而导致了荒谬的成本（例如，每个成年个体都要生下300多万个后代供选择淘汰）。他提出中性理论来应对这一重大的进化难题。悖谬的是，他的理论让他相信大多数突变是不可选择的，因而大部分的遗传信息都是

无关紧要的，大部分的进化与选择无关。由于他完全忠于基本公理，木村显然从未考虑过他的成本理论可能被合理地用来否定基本公理的正确性。

聂耳的领悟

Neel, J.V., et al. 1986. The rate with which spontaneous mutation alters the electrophoretic mobility of polypeptides. PNAS 83:389-393.

“……点突变在配子中的发生率……大约每代30个……这种规模的突变对于群体遗传学和进化理论的影响是深刻的。许多群体遗传学家的第一反应是指出突变大多发生在‘沉默’DNA中，并没有真正的生物学意义。对于这种思维来说，不幸的是……沉默DNA的数量在持续地缩减。我们人类如何适应如此高的突变率是进化论的核心问题。”

康卓寿的问题

Kondrashov, A.S. 1995. Contamination of the genome by very slightly deleterious mutations: **Why have we not died 100 times over?** J. Theor. Biol. 175:583-594.

“立田(Tachida, 1990)得出结论，VSDM（非常轻微的有害突变）只损害一个功能，即它与核小体的相互作用，这可能会导致过高的突变负荷。”

“兰德(Lande, 1994)和林奇等(Lynch et al, 1994)……得出结论……VSDM可以迅速导致种群灭绝……”

“同时选择多个突变会导致 N_e 进一步减少，加速灭绝（Li 1987; Lynch et al. 1984）。”

“我从整个基因组的角度来解释这些结果，结论与立田(Tachida, 1990)的观点相同，VSDM可能导致过高的突变负荷，即使 $N_e = 10^6 - 10^7$ ……导致令人困惑的高负荷的条件是相当现实的……”

“……即使 $U < 1$ ，负荷也会过大……我的分析表明，VSDM的污染意味着过度负荷，而这会导致随机突变负荷难题。”

“……对不同位点的选择过程会互相干扰。”

“……因为随机突变负荷难题是真实存在的，它需要一个解决方案。”

“切特维利科夫(Chetverikov, 1926)认定突变对物种的污染会与时俱进，最终可能导致其灭绝。”

“VSDM在某个谱系中的积累……就像一颗定时炸弹……脊椎动物谱系的存活……应限于 10^6-10^7 代。”

如果康卓寿博士相信自己的数据，他将得出这样的结论：基本公理是错的，基因组必须退化。相反，他最终求助于“协同上位效应”，来打消他自己精彩描述的这一问题的。

林奇等的突变熔消

Lynch, M., J. Conery, and R. Burger. 1995. Mutation accumulation and the extinction of small populations. *The American Naturalist* 146:489-518.

“由于有害突变的固定和累积，个体的平均存活能力逐渐下降……净生殖率降低……导致随机基因漂变和突变积累互相促进，我们称之为突变熔消……熔消阶段的时间通常是很短的。”

“这些结果表明，仅因为遗传的原因，有效群体规模小于100的有性繁殖种群就不可能延续超过几百代，尤其是当其繁殖能力相对较低时。”

“……我们的结果并没有证明一个群体规模阈值的存在，超过这一阈值，种群就完全可以避免突变熔消……”

“……同时分离的突变会互相干扰自然选择对它们的消除……”

林奇等人认识到基因退化对目前所有濒临灭绝的物种是一个主要因素。他们公开承认选择干扰问题，但是没有进一步阐释在大多数以往的物种灭绝中这一因素也是重要的，而且由于基因退化而灭绝的基本过程从逻辑上讲也应该适用于所有的高等基因组。

豪威尔的挑战

Howell et al. 1996. Evolution of human mtDNA. How rapid does the human mitochondrial genome evolve? *A. J. Hum. Genet.* 59: 501-509.

“我们应该进一步注意到一个更广泛的问题，即就进化而言，生物体如何（能否）容忍其遗传系统内存在如此重的突变负担。”

豪威尔对我们提出的挑战是基于他自己的数据，这些数据表明，仅在线粒体基因组中，突变率就可能达到每代每人一个突变。他是对的，只需每人0.1-1.0个线粒体突变，就能对进化理论造成不可克服的难题。而这相比于同时发生在其他染色体中的数以百计的突变还是微不足道的。

克劳的担心

Crow, J.F. 1997. The high spontaneous mutation rate: is it a health risk? PNAS 94:8380-8386.

“……突变过程的总体影响一定是有害的……典型的突变是很温和的。它通常没有明显的效应，只是活力或生殖力略微下降……每个突变最终将导致一个“基因死亡”……肯定会加重人类群体的突变负荷……所以这是一个问题。”

“……有一条出路是……通过明智的选择，牺牲一个个体剔除数个突变……凡是突变超过一定数量的个体都会被淘汰……当然……自然选择不会把所有人排起队来，把那些带有超过一定数量突变的人都除掉……这一模式的不切实际让我许多年都不曾将其视为人群应对高突变率的一种方法……虽然斩截性选择是完全不现实的，准斩截性选择是合理的。”

“我们似乎清楚地看到，在过去的几个世纪中，有害突变一直在积累……因突变积累而造成的生存能力的下降约为每代1-2%……如果战争或饥荒迫使我们的后代回到石器时代的生活，他们将不得不接受他们的石器时代祖先所面临的一切问题，还要加上在此期间积累的突变……因为环境的改善，突变对生存和生育的平均水平的损害只是轻微的……我确实认为突变积累是一个问题。它在某种意义上类似于人口爆炸，只是导火线更长些。”

克劳博士承认了由于高突变率的发现而造成的一些根本的进化问题，但他试图用一种很不现实的理论模型来消除这些问题。这个理论模型涉及一个基于突变计数的高度人为的选择系统。至于这个人人为的选择方案是用斩截还是“准斩截”都是无意义的争辩。他接着承认，现在的人类基因一定比我们石器时代的祖先要低等。这是一个对基因退化的现实所做的惊人的（但也是实际的）坦白。

克劳博士还对突变固有的有害特性做过评论：

Crow, J.F. 1958. Genetic effects of radiation. Bulletin of the Atomic Scientists 14:19-20.

“即使我们在这一点上没有很多的数据，我们仍然可以凭理论分析，相当有把握地讲，突变通常是有害的，因为突变是一个有高度组织性的、相当顺畅地运行着的活体的随机变化。在构成生命的高度整合的化学反应系统中所发生的随机变化，几乎肯定地会损害它，正如在一台电视机中随机地改变线路连接时不大可能改善其图像质量一样。”

霍伊尔的高瞻远瞩

Hoyle, F. 1999. *Mathematics of Evolution*. Acorn Enterprises, LLC, Memphis, TN. （注：与我在本附录中引用的其他作者不同，霍伊尔博士不是一位遗传学家，而是一位很杰出的理论数学家和物理学家）。

“老龄化进程说明，我们经常听到的那些认为达尔文理论就像地球围绕太阳转那么显而易见的说法，要么是一些天真得难以置信的言论，要么就是谎言……鉴于我们周围的世界已有大量种群衰老的证据，很难相信还有那么多人以为整个生物系统“显然”在朝着相反的方向发展……”

“自然选择最多能在特定环境下控制有害突变。然而当环境不稳定时就会有缓慢的基因腐蚀，这是自然选择无法阻挡的。”

“……如果有害突变很轻微，自然选择就无法消除它们，而假以时日，大量的小缺陷就会加重成为一个严重的残障。由于自然选择不能长期维持遗传物质的完整性，遗传物质的使用寿命就必然是有限的……”

艾尔沃克和凯特利的退化

Eyre-Walker, A. and P. Keightley. 1999. High genomic deleterious mutation rates in Hominids. *Nature* 397:344-347.

“根据保守的假设，我们估计在人类谱系中平均每代每个二倍体已发生4.2个氨基酸改变……”

“……接近像人类这样的物种所能耐受的上限……大量轻微有害的突变可能因此而固定在原始人的谱系中……很难解释人类种群是如何生存下来的……高有害突变率 ($U \gg 1$) 在一个低繁殖率的物种中是个难以自圆其说的问题……如果新突变的相当一部分是轻度有害的，这些突变累积起来……会导致适合度逐渐下降。”

“……在人类及其近亲中有害突变率似乎是如此之高，很难相信这些物种竟能生存下来……”

“……人类蛋白质编码序列的制约水平非常低，大约有一半新的非同义突变似乎已经被接受……如果有害的新突变如今还在积累，这会对人类的健康产生破坏性的后果……”

这两位作者依然低估了突变问题的严重程度。他们只考虑到基因组中蛋白质编码部分的突变，但全部功能性基因组比这一部分要大10-30倍。即便使用这些较低的估计，他们也承认了一个根本性的问题，并得出结论，即许多有害突变已在人类进化过程中累积，并且可能在继续累积。这难道没有清楚地表明，我们实际上是在退化而不是在进化吗？

拿迟曼和克洛维尔的难题

Nachman, M.W. and S.L. Crowell. 2000. Estimate of the mutation rate per nucleotide in humans. *Genetics* 156: 297-304.

“人类二倍体基因组……每代约发生175个新突变。人类的高有害突变率带来一个难题。如果突变之间的相互作用呈倍增关系，这种高U值所造成的遗传负荷对于一个低繁殖率的物种是不可承受的……如果 $U=3$ ，平均适合度将下降到0.05，换句话说，每个雌性个体需要生育40个后代，以保证其中2个可以成活，来保持群体规模。这还假定所有的死亡都是选择的结果……所以，保持恒定群体规模所需后代的实际数量可能更高。”

根据康卓寿的说法（第189页），U值（每人携带的新有害突变的数量）实际上比这两位作者所说的（他们假定97%的基因组是沉默的垃圾）要高10-30倍。此外，我们知道，事实上总死亡率中只有一小部分可归因于选择。尽管他们的假设不切实际，这二位仍然承认了一个根本性的问题。然而，他们最终以“协同上位效应”的说法打消了这一难题。

希金斯和林奇：更多的熔消

Higgins, K. and M. Lynch. 2001. Metapopulation extinction caused by mutation accumulation. PNAS 98: 2928-2933.

“在这里我们表明，复合种群结构、栖息地丧失或分裂、以及环境的随机变化会大大加速轻度有害突变的积累……这种积累会达到一定的程度，即便是大型复合种群也可能面临灭绝的危险。”

“……轻度有害突变可能造成更大的突变负荷……因为单个的突变几乎是自然选择所看不到的，尽管它们加起来会造成群体生存力的显著下降。”

“我们发现，新的轻度有害突变的积累从根本上改变了灭绝的时间范围……导致一些从群体统计学的角度视为安全的种群走向灭绝。”

“在同步环境波动的情况下，突变累积所造成的灭绝加速是惊人的……没有突变，灭绝的时间是2000代……有突变累积时，灭绝时间仅稍长于100代……”

“对于有完整栖息地的复合种群，轻度有害突变比高度有害突变的破坏性更大……正如高携带能力的情况，轻度有害突变的效应最具破坏性，导致灭绝所需的时间最短。”

“早期的研究表明……有害突变的积累会威胁小型孤立的群体……这里我们显示有害突变的积累对大型复合型种群也是一个重大的威胁……群体萎缩是突然的，但灭绝本身仍然需要一段时间才会发生。复合种群在中长期时间范围看可能全无生存能力，尽管短期看来还是健康的。”

希金斯和林奇有力地证明，对于哺乳动物和一切类似的动物种群来说，基因退化是个普遍问题。他们指出，现有的基因损害会导致最终灭绝，尽管其效应还需要一定的时间。而在此时，物种仍然会表现

得健康有活力。对于人类，这种情况会发生吗？人类生育率和人类精子计数目前正在急剧下降(Carlson et al., 1992)。许多国家正面临着人口负增长，这可能是由于非遗传的原因。然而，是不是可以想象，我们可能已处在突变熔消的早期阶段呢？

康卓寿的数字

Kondrashov, S. 2002. Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian diseases. Human Mutation 21:12-27.

“……每代人每个二倍体基因组中新突变的总数约为100……其中至少10%是有害的……对人类变异的分析表明，每个正常人携带着成千上万的有害等位基因……”

这篇文章发表以后，康卓寿博士曾在与我的私下交流中表示，100只是他估计的下限，而300则是上限。他还对我说，他现在认为高达30%的基因突变可能是有害的。这意味着，从他的角度来看，“U”（每代每人有害突变的数目）将在30-90之间。这比先前的估计要高100倍。但到最后，他用“协同上位效应”和“斩截性选择”把整个问题消除了。

楼伟的时限

Loewe, L. 2006. Quantifying the genomic decay paradox due to Muller's ratchet in human mitochondrial DNA. Genet. Res. 87:133-159.

“数量惊人的符合生物学现实的各种参数集合应该使得正在向人类进化的谱系在2千万年以内就灭绝了……”

楼伟的灭绝时限仅是基于线粒体基因组的破坏，而整个基因组的退化速度比这要快20万倍！

林奇的最新分析

Lynch, M. 2010. Rate, molecular spectrum, and consequences of human mutation. PNAS 107 (3): 961-968.

“根据以上结果，我们推测平均每个人类配子中发生了大约38个新的碱基置换突变，在其复杂序列中约有3个小插入/缺失，还有大约1个剪接突变……所以说，在双倍体的水平上，平均每个新生儿可能要带来50-100个新突变，其中的一小部分必定是有害的。”

林奇博士的研究证实了其他科学家的分析，认为人类的突变率大约是每人50-100个（即每个配子25-50个）。如果考虑到整个基因组中所有的突变类型，50-100个仍然太低。更重要的是，把大多数突变假定为完全中性已经不是合理的了。多数突变必定是轻度有害的（即近中性），见下文。

“绝大多数点突变发生在编码区之外（每个配子约40个），其中大多数可能会有很轻微的适合度效应，其平均值几乎肯定 $\ll 10^{-2}$ 。”

如果每个配子中有40个不影响蛋白编码的新突变，那么每个新生儿就有80个这样的新突变。林奇博士在这里承认这些突变大多数不是中性，而是轻度有害。

“尽管以上数字有很大的不确定性，但是结论已经不容回避：人类由于反复发生突变而造成的适合度下降不会小于每代1%，很可能高达每代5%……在两三百年（大约6代）之内，其后果可能变得很严重，再加上人类活动造成突变率的进一步升高（环境中致突变物质的水平增加），情况堪忧。”

突变率也许已经高于林奇博士的估计（即使不考虑致突变物质可能的增加），这是因为：1. 他的分析不包括基因组中最容易突变的部分；2. 他假设绝大多数突变发生于“垃圾DNA”中，所以无害。

“如果不设法减少经性细胞遗传下去的有害突变，在两三百年之

内，发达国家居民的平均表现型就会有明显的改变，其外形、生理功能和神经生物学功能会出现显著的残障。”

一如本书中所指出的，林奇博士所记载的基因退化不能靠提高选择强度来阻止。这是因为突变率太高，自然选择又太低效，而且多数突变是近中性的，自然选择无能为力。

林奇博士不仅认识到了生殖细胞中突变的重要性，他还指出在我们体内的其他细胞中突变也在迅速地累积，从而造成我们的衰老。人到了60岁，每个细胞中就带有数万个突变。这就是限制着我们的预期寿命的根本原因，没有什么医学上的突破会中止我们体内每一个细胞中持续的突变积累。

大科学揭示人类基因组具有难以置信的信息密度

The ENCODE Consortium, 2012. An integrated encyclopedia of DNA elements in the human genome. Nature 489: 57-74.

ENCODE项目历时十余载，耗资4亿美元，招揽了世界各地442位科学家，其结果在多家杂志上发表成多篇论文。其关键性总结发表在Nature杂志，其中有以下陈述：

“这些数据使我们能够为80%的基因组（尤其是以前知之较少的、蛋白编码区以外的序列）鉴定了功能。”

“在ENCODE所注解的功能性区域中，不同个体的基因组之间在非编码序列上发现变异；这些变异的数目至少不亚于编码区的变异。”

“这些新发现的元件跟那些与人类疾病相关的序列变异有统计学上的关联，所以有助于对这些变异的解释。”

“根据GWAS发现的与疾病相关的单核苷酸多样性(SNP)集中于非编码功能性元件中，其中大多数发生在ENCODE所认定的、蛋白编码区之外的区域或附近。在多数情况下，病态表现型可以与特定的细胞类

型或特定的转录因子联系起来。”

Nature杂志配有一段录像，其中Nature的编辑采访了NCODE项目的负责人（录像的标题叫做《ENCOCE之声》，见<http://nature.com/ENCODE>）。

伊纨·博内博士（Dr. Ewan Birney）：

“80%有可能成为100%……垃圾的比喻不是很有用……很难理解如此的信息密度。我想人们从前以为基因组是个井井有条的地方，基因各就各位，各自摆出相当沉稳的舞姿。然而数据表明并非如此，数据揭示出一个密密麻麻的丛林，以前我们以为明白的事情原来要复杂得多。基因组中我们原以为完全沉默的地方原来充满了生命，充满了故事……”

Nature的编辑：

“至于基因组中哪一个区域与人类疾病有关，人体遗传学已经发现了不少。令人沮丧的是那些区域大都属于基因组中的非编码区。我们对基因组中的这些部分一直知之甚少，而这正是ENCODE项目的目标。”

atcag

附录 2

选择能同时作用于多少核苷酸？

在一个特定的繁殖群体中，可以同时选择多少性状、多少核苷酸？这是一个很重要的问题，但是一直没有得到足够的重视。它与人工育种、群体生物学和进化理论都有关系。从数学上讲，这个问题是可以解答的。

C和c的定义

一个群体的总选择代价（C）是指为了达到全部选择目标而被禁止繁殖的那一部分群体。在最简单的情况下，我们设想被淘汰的那部分群体为C，这一部分产生的后代数为零。剩余的（1-C）就是被保留的那部分群体，允许按正常速率繁殖。不同物种所能支付的选择代价各不相同。假如，在某种植物，每株可以产生100粒种子。这个物种的C值就可高达0.99。也就是说99%的幼苗可以被淘汰，而种群仍然可以充分繁殖。就人类而言，现今的人口生育率大约是每两个成人3个孩子。所以要保持群体的规模，人群中每三个孩子中只有一个可以被淘汰。在人类，选择的代价必须低于人口的1/3，C值不可超过0.33。在现实中，这个代价还是太高，因为许多人由于非遗传的原因（意外死亡，个人选择等等）不生育。我们无法知道有多少不育是由于非遗传因素造成的，但肯定不少。一个比较现实的估计是人类所能承受的

选择代价肯定不高于25%，也许更接近10%。大方一点，我们可以假设C值高达0.25。在决定人类选择代价的理论上限时，我们将假设一个不现实性的高生育率群体，让C为0.50（每代基于遗传因素从繁殖群体中淘汰掉一半数目的孩子）。

每个性状的选择代价（c）是指由于一个特定性状（或核苷酸）的存在而被淘汰的那部分群体。要淘汰某个特定性状（或核苷酸），我们需要决定选择的强度。换句话说，为改善某一个性状，我们愿意淘汰掉总群体中多大一部分？为了某个性状而被淘汰掉的那部分群体就是那个性状的选择代价（c），代表着那个性状的“选择压力”。比如，为了影响一个性状而淘汰10%的群体，那么那个性状的 $c = 0.10$ 。如果 $c = 0.01$ ，那么为了影响那个性状，1%的群体被禁止繁殖。

相加模型

最简单的模型是相加模型。在此模型中，我们假设选择是相加性的，而且对所有性状的选择是同时进行的。假如我们有条件在一个100人的群体中淘汰掉25人，我们可以在每一个性状上淘汰1人，这样就可以同时选择25个不同的性状（或25个核苷酸）。通用公式就是，群体总代价（C）等于每个性状（或核苷酸）的代价（c）之和，即 $C = c_1 + c_2 + c_3 + \dots + c_n$ ，n为性状的数目。如果选择压力对于每个性状都是一样的，那么 $C = n \times c$ 。例如，每个性状的选择压力均为0.001（为影响一个特定的性状，1千个个体中有一个被淘汰），而总选择代价限于群体数量的25%， $0.25 = n \times 0.001$ 。在这个例子中，最多就可以选择250个性状。虽然250个性状可以被同时选择，但每个性状的选择压力却是微乎其微，导致了选择进展缓慢或毫无进展。随着选择作用于越来越多的性状，选择的进程迅速趋近于零（参看图6a，第72页）。

相乘模型

相乘模型比相加模型更现实，但稍微复杂一些。在此模型中，先针对一个性状（或一个核苷酸）进行选择，然后在剩余的群体中再选择下一个性状（或核苷酸），选择是先后进行而非同时进行的。经过一次选择后，剩余的群体就是 $1 - c$ 。如果想选择两个性状，就把两个余数相乘， $(1 - c) \times (1 - c)$ ，得到二次选择后的剩余群体百分比，然后用1减去这个数就是总选择代价。例如，我们为选择一个性状而淘汰群体中的10%，然后又为另一个性状淘汰剩余群体中的10%，那么我们的总代价就是： $1 - [(1 - 0.1) \times (1 - 0.1)] = 0.19$ 。换句话说，在选择了这两个性状之后，还有81%的群体可以繁殖。对许多（n个）性状的选择来说，假设每个性状的选择强度大致相同，通用公式就是 $C = 1 - (1 - c)^n$ 。

我也曾依照相乘模型来研究被选择性状的数量和每个性状的最大选择强度之间的关系。假设 $C = 0.25$ ，即为了全部选择目标，可以淘汰25%的人口。得到的曲线的形状与相加模型基本雷同。随着被选择性状数量的增加，每个性状所允许的选择压力呈指数下降，迅速趋近于零。即使是一个生育力极强的群体，这个基本趋势也并不改变。假设一个生育力不凡的人群（ $C = 0.5$ ），随着n值的增多，每个性状所允许的选择作用也是急剧下降。即使考虑一个生育力极强的物种，如C值高达0.99的种子植物，当被选择的性状（核苷酸）超过1000时，所允许的最大选择压力也会变得很小。见图6b，第73页。

这个微乎其微的选择压力意味着什么？当作用于一个性状的选择压力趋近于零时，选择进程也趋近于零，而要靠选择来改变性状所需的时间将趋于无穷大。当选择进程趋于无穷小和无穷慢时，我们就知道有问题了。这是因为新突变持续地、高速地涌进群体中。我们没有“深度时间”来挽救退化的问题。我们需要淘汰率跟得上突变率，不

然突变就会在群体中经漂变而固定。更重要的是，当允许的选择压力低到一定程度时，有效选择会真正地停止，这是由于“噪音”现象和基因漂变的结果。取决于群体的大小和与繁殖相关的生物“噪音”的总量，选择总会在需要选择的性状达到一定数目时完全停止。

选择完全失效的具体的临界点受许多因素的影响，所以很难精确地核算。但是常识可以帮助我们对选择失效的阈值做出大致的估计。针对某一特定性状的选择系统，如果不能从1000繁殖个体中淘汰掉1个，这个系统肯定值得怀疑。这相当于对该性状的选择代价为0.001。鉴于人类群体中存在高度的“噪音”，当选择代价低于0.001时，对该性状的有效选择可能会完全停止。再通俗一点说，在一个1000人的群体中，如果我们不能允许为影响一个特定的性状而淘汰掉1个人，对该性状的选择实际上就停止了，随机漂变可能已经在发挥作用。用这个0.001的截断点和相加模型，可以算出最多能够同时选择多少性状。在一个现实的人群中，最多只能选择500个性状（核苷酸），而在一个理想化的（生育力极强的）群体中，也只有990个性状可以被选择*（参见表2）。然而已知的人类突变率告诉我们，我们每代需要选择好几百万个核苷酸位点才能阻止基因组退化。

*木村资生也提到了同样的问题。虽然他没有说明他的计算方法，但他提出当 $C = 0.50$, $s = 0.01$ 时，只能同时选择138个位点（Kimura, 1983, 第30页）。

可以同时选择多少基因单位?
(假设每个性状的选择代价之最低值为 $c = .001$)

C	n^a	n^m
.25	250	300
.50	500	700
.99	990	4,600

n^a =按相加模型，最多可被选择的性状。

n^m =按相乘模型，最多可被选择的性状。

表2 在典型群体中可以同时有效地选择多少突变?

能够同时有效地选择的突变（核苷酸变异）数目取决于多种因素，其中一个关键因素就是每一代允许淘汰的群体比例（选择代价），即C值。这里列举三个假想的群体。第一个C为0.25，即每一代允许淘汰群体的四分之一（可能见于一个繁殖力和选择压力都很强的哺乳动物群体），这种情况下只能有效地选择大约300个突变（见正文）。第二个C为0.5，即每代容许淘汰群体的一半，这样也只能有效地选择700个突变。这代表了选择压力更高的情况，比如繁殖率高的鱼类。当C为0.99时，每代允许淘汰99%的群体，这样也只能有效地选择4600个突变。这代表极强的选择力度，比如人工育种中的植物群体。

atcg

附录 3

合一现象和整合复杂性的概念

如何识别智能设计这一难题已经渐渐地得到破解。从前人们一直靠直觉识别自然界中的设计。这种看法符合逻辑，也符合常理。但因为有人拒不承认这显而易见的道理，圣经从《创世记》到《启示录》都明确地指出了设计的问题。后来，科学界几乎所有的“开山鼻祖”都主张设计，包括哥白尼、培根、牛顿、巴士德、麦克斯韦、法拉第和开尔文。佩利(Paley)率先(1802)用复杂性当作设计的论据。贝希(Behe)最近(1996)将这一概念进一步精炼，提出“不可简化之复杂性”的论证。复杂性论证后来又发展成两个相关的分支，信息理论(Gitt, 1997; Gitt et al., 2013)和特定复杂性(Demski, 1998)。然而，我认为至少还有一种更有用的诊断性标识，就是合一现象，该现象是出于“整合复杂性”。

设计的一个诊断性标志就是大量组分的全方位整合，我称之为**整合复杂性**。自然界随处可见的合一现象即以此为基础。合一是个客观事实，任何有理性的人都很容易看到，而绝非出于主观。因此对合一进行科学分析是有理由的。合一是通过众多部件的全面整合而形成的。拼图游戏有合一性，一堆沙子就没有合一性。

一架战斗机是由数以千计的部件和无数的原子组成的，但其功能和形体都有合一性。这就是为什么一眼就能看出它是一个被设计的产品。它以一个整合单位的形式存在，超越了它的所有部件。在没有

老化的原始状态下，每一个部件都有其用途和位置，而且都与其他部件完美地整合。尽管有数不清的部件，但只有一架飞机。这就是“合一”（完整性）的内涵所在。我们不说，“哦，看看那些金属和塑料零件吧。”我们说，“哦，看看那架飞机。”谁也不会否认一架飞机不仅仅是它所有部件的总和。飞机是一个新物体，与任何零件处于完全不同的档次。它能飞，零件不能。同样地，不用说一艘宇宙飞船不止是大量的金属，不用说一本书不止是一堆字母，也不用说生命不止是它所有部件的总和。这些说法都是浅陋的，如同说大海里不止有一滴水。如果不是对着一个神志不清的人，我们岂可大声地说出这样俗不可耐的话来？

人体内有超过一百万亿个细胞，但我们**不是一百万亿个细胞**。我重复一次：我们不是细胞。我们每个人都是在形体上、功能上、存在上真正合一的整体。我们是无数部件几近完美的整合，因此我们是一个新层次上的实体。我们作为人，是与我们的分子截然不同的事物，这是一个既奇妙深奥又单纯明显的事实。只有灵性沉睡的人才会对此视而不见。我们急需醒悟！当我们看到合一的现实时，我们也就看到了何谓美。我们就会发现我们所称的“美”其实就是被设计的物体的全面合一性。由此看来，美并不仅仅是主观的。美，如合一一样，可以看作是真正客观而具体的现实。*

当科学家有点诗意的时候，他们会把合一的美称作优雅。优雅指的是一种精彩奇妙的设计，每个细节、每个侧面都完美地结合而形成一个新事物，一个**全方位整合的统一体**。合一性就是没有一处松结或毛边。例如，在人体中，每个细胞各就各位，各司其职，这样就定义

*作为作者的旁白，美的反面就是丑。我要提议丑也是个客观的现实。丑是设计的朽坏，是对合一的损毁。这就是为什么一个疣子客观上可看作是丑的。这就是为什么衰老是一个丑化的过程。这就是为什么生锈的汽车、生物畸形、战争和谎言全是丑的。

了一个整体。一个人的形体，如一架完美打造的喷气式飞机一样，展现的是形体的优雅和功能的一致。我愿说，合一性是我们所谓美的具体而客观的基础。我相信合一性也是高级设计的明白无误的诊断性标识。

人体（我们的表型组）的奇妙合一性对于任何一个有思想的人来说应该是显而易见的，跟懂多少生物学没有关系。当我们看到一个人时，我们不会想，“看看那些细胞和组织！”我们看到的是一个个体，一个人。

这与人类基因组有何关系？基因组应该是表型组合一性的基础。然而，让人不解的是，多数现代遗传学家却把基因组大致看作一堆散乱的核苷酸。按照他们的说法，所有我们的基因组加起来仅仅组成一个“基因库”。这正是合一性的对立面。基因组被看作是大量分子的陈列，而这种陈列大致上是偶然的，并且几乎完全是随机的。每个核苷酸被认为是独立出现的，而且正在独立地“进化”（或漂变）。整个这种思维模式（即人只是一包分子）叫做还原主义。典型的现代遗传学家认为基因组大部分是“垃圾DNA”，其中寄生着超过一百万个“自私基因”（他们也认识到当中有些真实的信息，就是几十万有功能的基因）。普遍的看法是每个自私基因都有其自私的策略，以整个基因组为代价来进行自我繁殖。

怎么会这样呢？依照熵的概念和热力学第二定律，表形组奇妙的合一和秩序怎么可能全然是从零碎和混乱的基因组中产生的呢？从理性上思考，如果表型组的秩序和合一是由基因组而来，那么基因组岂不应该比表型组更复杂更统一吗？

试想走进一个称为S. S. 表型组的星际宇宙飞船。你经过标着“超光速引擎间”和“模拟空间”的门，然后看见一扇标着“总设计师和总工程师办公室”的门。你打开门，看到一个杂乱无章的办公室，到处是纸片，有发臭的食物的气味和已损坏的电脑屏幕。在一张

桌子上站着两只黑猩猩，它们正在为一个香蕉而打架。你会天真地认为你看到的确实是S. S. 表型组的总设计师和总工程师吗？你真的会认为S. S. 表型组是在这个退化和混乱的环境中被创造和保存的吗？然而这正是现代科学对基因组的观点！这就是有关基因组本质的主导模式，它描述了该模式中基因组和表型组之间白痴主人和天才奴隶的关系。鉴此，难道我们不应该批判性地重新审视我们对基因组的认识吗？是否到了模式转变的时候了？

假若整体复杂性就是生命的特征，假若基因组确实是经设计而来的，我们可以预见基因组应该表现出更广泛的整合和统一。我们应该能够发现基因组内有许多层次的结构和功能的合一性。我相信已经开始有这样的发现了。随着基因组内精致的多维结构被解开，我预测未来几年我们会看到越来越多的整体调控。我预测，当我们对基因组了解得更多的时候，我们会看到所有不同层次的整合和统一。我预想，既然突变就是退化，而且选择不能遏制突变带来的退化，我们也会看到越来越多的退化和衰败的证据。基因组无疑正在经受着由于高突变率而导致的巨变。在我看来，我们所见到的完全是“下坡”的改变。这样的随机改变不可能是整体复杂性的来源。简单来说，合一性（全方位的整合复杂性）是不能通过一次一个错误来形成的（正如这本书的主要部分所论述的）。

生命的高度合一性揭示了还原主义的真实面目：就是一种盲目的精神。还原主义其实就是对我们周围不言自明的合一性的彻底漠视。更具体地说，基本公理，包括“基因库”的概念和每个核苷酸的独立进化的理论，不过是生物学上的极端还原主义。所以，这些理论实质上是不能成立的。从这种意义上讲，本书的一切论述都是没有必要的。我个人确信即使没有本书对遗传学的论述，仅凭随处可见的合一现象就可将基本公理推翻。



基因复制和多倍体能增加遗传信息吗？

也许有人会反对本书的主要论点，辩称复制是理解基因信息自发增加的关键。复制确实在所有的遗传系统中都会自然发生，是突变的一种形式。一次复制的规模可大（一条染色体或所有染色体一起被复制）也可小（一个或几个核苷酸）。当细胞中增加一条染色体时称为非整倍体，当所有染色体倍增时，称为多倍体。正如文字处理的错误，一个字母、一个单词、一整章书、一整本书、或整个图书馆都可能被复制。问题是：这样的复制会创造新的信息吗？

如果我重复一个字，这会改改改善我们的句子吗？如果我重复一个句子，我有没有给你更多的信息？如果我重复一个句子，我有没有给你更多的信息？如果我重复一个句子，我有没有给你更多的信息？如果这一页在书中其他地方再次出现，这本书会写得更好吗？如果本书的每一页都重复一次，你会学到双倍的知识吗？显然，不管复制的规模大小，所有这些复制都是有害的。它们非但没有促进沟通，反而明显地带来不便。人们怎么会以为这种复制是自发地扩增有用信息的一个现实的途径呢？答案当然是这些人想象将突变复制与全能的选择结合就行了。但是我们在本书的许多章节中已经表明，虽然选择可以减慢突变带来的信息流失，却无法阻止这个过程。选择绝对更不能逆转信息的流失。如果您已经读过这本书，现在就不难明白，像其他所有类型的突变一样，几乎所有的复制都是有害的而且是近乎中性的。

这就意味着选择只能淘汰那些最糟糕的复制。其余的复制会无情地积累，逐渐毁掉基因组。

生物学研究的发现是否支持这个关于复制的常识性认识呢？断乎如此！让我们考虑人类的情况吧。有没有多倍体的人呢？当然没有！整个人类基因组的复制绝对是致命的。有没有非整倍体的人呢？有，多一条染色体的人不少。这些个体有更多的信息吗？当然没有。虽然非整倍体对于大的染色体来说是彻底致命的，多一条最小的人类染色体却不总是致死的。不幸的是，有这种“额外信息”的个体都表现严重的遗传障碍。最常见的例子是唐氏（先天愚型）综合症，患者细胞中多了一条21号染色体。还有无数小的复制和插入也可引起遗传疾病。现在读者该明白，像其他所有类型的突变一样，大多数复制都是有害且是近乎中性的。

大家普遍认为，不论是在书面文字里还是在活体基因组中，复制都会破坏信息。也许有一些个别的例外，复制从某些小的方面来说是有利的（可能导致“微调”），但这并不能改变绝大多数随机复制会破坏信息的事实。就这点而言，复制就像其他类型的突变一样。

当一个特定的基因经过了有害突变的长期积累之后，就到了一个部分退化的状态。如果那个基因被复制，其中的有害突变也随之被复制。这样的复制怎么会使退化的过程减慢呢？显然不会！仔细地考虑一下，我们可以看到一个基因被复制以后，两个拷贝都会比以前退化得快。为什么呢？因为一旦有了一个备用拷贝，选择对两个拷贝的作用就都放松了。经常有人声称，基因复制以后，一个拷贝可能保持不变，而另一个可能自由地进化出一个新的功能。然而，这两件实际上都不太可能。就像我们已经知道的，由于近中性突变的积累，两个拷贝都将以大致相同的速度退化。两者没有一个会保持不变。再者，基因转换现象从理论上应该不断地在所谓“不改变的拷贝”和“进化的拷贝”之间造成交叉污染。从理论上讲，基因转换也应该允许每个拷

贝内的突变跳进另一个拷贝，结果使两个拷贝的突变率都增加了。这无疑也会加快退化的速度。总的来说，由于选择的放松、突变积累的加速和基因转换带来的突变混合，复制后的基因拷贝显然会互相促进退化的过程。如果这还不够，第九章明确指出，当两个拷贝都不可逆转地退化时，猜想一个基因拷贝可能会进化出一个新功能是多么地不合理。

那么多倍体植物呢？有人称，既然有些植物是多倍体的（染色体数目加倍），这就证明了复制肯定是有利的而且会增加信息。我的博士论文研究课题就是多倍体。有趣的是，多倍体根据其产生的方式而表现迥异。如果体细胞经一种称为秋水仙碱的化学试剂处理，细胞分裂就会中断，导致染色体数目加倍但没有新的信息产生。培育出的植株几乎总是发育不良、畸形、不结种子。原因应该很明显。这些植株浪费了两倍的能量制造出两倍的DNA，却没有产生新的遗传信息！细胞核几乎是原来的两倍大，破坏了正常的细胞形态和体积。事实上，这些植株的信息比以前减少了，因为大量的基因调控信息依赖于基因的拷贝数。失去调节控制就是失去信息，这和多余的染色体导致唐氏综合症是同样的道理，多余的染色体妨碍了数以千计的基因的正常调节。

如果体细胞的多倍体化向来是有害的，为什么还有多倍体植物（如马铃薯）呢？原因是多倍体还可以通过一个不同的过程，称为“性细胞多倍体化”而产生。这发生在一个未减数分裂的精子和一个未减数分裂的卵子结合时。在这种特殊情况下，所有在双亲里的信息都集合在后代身上，在后代个体中就会有信息净增。然而在群体中，总信息并没有增加，双亲的信息只是简单地郁积起来了。在这种情况下我们看到的是信息的郁积，而不是新信息。

在二倍体生物中，每个基因都有两个版本。在杂合二倍体中，如果一个基因版本因突变而失去了功能，而另一个是有功能性的非突

变基因，后者可作为前者的备用拷贝。因此二倍体可看作是一个被设计的有后备功能的系统，是针对将要出现的突变问题而预先设计的。从另一方面看，进化不能预测未来的事，所以我们可以很合理地得出结论说，进化永远不会产生后备系统。性细胞多倍体化使得杂合潜能加倍，因而同一个体内的一个基因可能有四种版本。这样的系统有双重的备用机制。正如宇宙飞船上有四台重复设置的电脑，某个特定位点上可能有三个突变的等位基因，但只要第四个还有功能，这株植物还能存活。因此，多倍体化并不是增加新信息的一个途径，而只说明了冗余基因作为后备系统的重要性，这是为有效地延缓退化而设计的。这种后备系统的代价是让选择不能有效地淘汰突变，所以从长远看来，退化更加不可避免。在某些特殊的情况下，多倍体内的多重后备机制掩盖了基因调节障碍和繁殖力降低的问题，因而带来一种净增。然而这样的“净增”更准确地讲是退化速度的净减。

那么重复基因和基因家族呢？如果多重基因版本可以解释二倍体和多倍体的好处，那它同样也可以解释一个基因在基因组内不同位点上有冗余拷贝的好处。通常，如果在基因组内的另一部位见到一个基因的冗余版本，理论学家们就简单地假定这是远古基因复制的产物。他们通常进一步假定复制以后发生了突变和分化。但这都是基于理论上的推导，而非观察的结果。如果一个基因在基因组中重复，在逻辑上同样可以认为这种现象有一个被设计的功能，如后备基因或复杂的基因调节。

如果说只要复制一个基因就是有益的，这种简单的理解是生物学上的天真。确实，基因复制可能增加一个基因的表达程度。事实上，这也是时常所观察到的。但是仅仅增进一个基因的表达通常是有害的（基因表达必须受精密设计的分子系统的准确调节）。再者，复制对于增进基因的表达是远远谈不上有效的一种方式。进化怎么总是这样低效？

最后，实际的基因复制不仅扰乱了被复制基因的表达，还通常扰乱了其他基因的表达。在我的职业生涯中，我曾花许多年的时间致力于基因工程植物的培植。产业界和学术界在这种研究上花费了超过10亿美元。我们很快就发现，多重基因插入总是比单个基因插入的表达水平低。不仅如此，多重插入基因的表达也总是更不稳定（你开始明白基因调节的复杂性了吧）。另外，由于额外DNA的随机插入破坏了基因组内的重要位点，转基因植物中有很一部分表现出其他的基因缺陷。既然基因组有一个功能性和高度特异的组织架构，任何复制或插入从逻辑上讲都会破坏这个架构。这正是植物遗传学家所看到的。

把基因复制看作一种“进化新信息”的方式，这在进化学界已经根深蒂固。我相信一部分原因就是大家都在说：“这必须是正确的！不然还怎么能发生进化？”我也相信，当一句经文被人念诵得多了，听起来就成了不容置疑的真理。但仔细地分析信息究竟是什么，它是如何产生的，再加上一定的常识，就可以让我们知道随机复制总是有害的。我个人认为“通过随机复制而进化”充其量只是个许多人相信的哲学假设，而非站得住脚的科学观察。我相信，虽然这个假设听起来很精深奥妙而且冠冕堂皇，但它经不起诚实的和批判性的推敲。

作者注：基因组最关键的方面是它携带了大量的功能性信息。遗憾的是功能性基因信息常与“沙农信息”被混为一谈。澄清这个疑惑最简单的方式，是解释沙农信息所关注的只是潜在信息，而基因组信息是真正的功能性信息。如果你购买一个造词游戏，其实就是一套字母，这些字母代表了一定量的潜在信息。如果你用这些字母给朋友发一个短讯，那就代表了功能性信息。如果你再买一个造词游戏，你的潜在信息就增加了一倍。但你还没有另外创造出任何功能性信息（这需要对新字母进行智能排列）。需要指出语言所沟通的就是功能性信息。沙农信息只适用于潜在信息（我有什么样的一套字母？），而且总是假定一套线性的、一维的信息码。沙农发明了他的电子通讯系统统计方法（一条电线可以输送多少电子字节？），但他也明确地指

出他的这些概念不适用于功能性生物信息体系。

atcg

附录 5

可能存在的四条异议

在本附录中，我想回答对于本书主题可能存在的四条异议。这些问题并未在本书的主体中论及，因为它们的对象是更高级的读者，如果加在章节之内会影响本书的连贯性和可读性。

异议1 现存的多种生命形式否定了基因熵

有人称，既然我们还活在这里，这就证明基因熵不可能是真实的。鼠类应该比人类退化更快，可是鼠类仍然繁荣昌盛。细菌和病毒早就该灭绝了，但还是无处不在。

此议无效：

这种异议是由于反对者对我所讲的基因熵缺乏清晰的理解。我的主旨并非说万物都在走向灭绝（虽然我也赞同这种观点），我的首要目的是要指出，多重证据表明新达尔文主义是错误的。突变/选择过程不仅不能创造基因组，它甚至不能阻止基因组的持续退化。如果只有进化加选择的新达尔文主义理论，万物必将灭绝。我意识到，除了自然选择之外，可能还有一种对抗基因熵的力量。这种反作用可能是上帝，或外星人，或某种未知的自然力。但是如果只有严格的新达尔文主义理论，那么一切生命形式确实将归于灭绝。

我们为何还活在这里？鉴于基因熵的真实性，如果除了自然选择

之外再也没有对抗基因熵的力量，那么我们之所以还生存，只能提示基因熵还没有足够的时间导致我们灭绝。这符合圣经中对人类历史的记述，但在进化论的圈子里这是大不敬。对那些无法赞同圣经中的时间架构的人，我建议你可以考虑神导进化或其他外部作用，作为潜在的对抗基因熵的力量。

鼠类为何还活在这里？假定鼠类基因退化比人类快是有道理的。鼠类的繁殖周期确实要短得多，所以在一个世纪中会经历更多的世代。但是鼠类每代的突变率低，尽管我不清楚鼠类每年每个个体的突变率是否比人类高。不仅鼠类每一代的低突变率中和了快速的世代交替，鼠类还有两个关键的优势：高繁殖率和频繁的选择。它们每个世纪会发生多轮选择。在人类，每隔20-30年才出现一次选择的机会，每个世纪只有四次选择。在鼠类，每个世纪可能会有几百轮选择。这就极大地提高了选择淘汰有害突变的机率，从而有效地延长了物种走向灭绝的时间，所以人类应该在鼠类之前灭绝。

细菌病毒为何还存在？微生物的问题和鼠类类似。微生物在每个世纪中的时代数目要多得多，但每代的突变率也低得多。人类每代每人有约100个突变，但在细菌中，每1000代（即1000次细胞分裂）只发生1个突变。在这样的微生物中，选择应该跟得上有害突变的出现率。重要的是，在微生物系统中，每一次细胞分裂之后都有选择作用，所以每一个突变都要经历1000轮选择。人类的情况则相反，每100个突变只经受一轮选择。这些优势使得大多数微生物特别有能力抵制基因熵。然而微生物仍然受到基因熵的影响，因为大多数有害突变是近中性的，是选择所看不见的。

RNA病毒，比如流感病毒和埃博拉病毒，情况特殊——它们繁殖期短，而且突变率也高（因为不存在RNA修复酶）。因此，人们普遍认为RNA病毒容易出现误差灾难和自发性突变熔断，对此已有大量文献记载。在RNA病毒中，某些毒株因基因熵而灭绝可以在一个世纪内

见到（如流感），有时候只需要几个月（如埃博拉）。所以在这些特殊情况下我们必须问：RNA病毒为何还存在？答案在微生物学界是众所周知——病毒和细菌能够以休眠状态长期存活，然后从“天然库存”中重新浮现出来。微生物学家常规地将这种知识应用于实验室研究，他们知道持续生长中的毒株在基因上不稳定，所以不断地把研究中的毒株冻存在冰箱里，以备将来之用。如果他们去度假或把一个项目中止10年，他们可以将生长中的毒株销毁，将来从冰箱里（他们的“天然库存”）拿出原始的、突变较少的毒株，复苏以后重新开始。微生物的天然库存方式包括冷冻、干燥、半休眠、或芽孢。如果监测数千年，任何具体的菌株或毒株处在休眠或半休眠状态的时间都可能长于快速生长的时间。一个毒株可能生长一年，然后休眠或半休眠一个世纪。所以说，微生物不是活着的生物钟，它们能够存活数千年，并不一定表示它们经历了大量的细胞分裂或积累了大量的突变。

异议2 超级有利突变

如果偶然发生一些罕见的极为有利的突变，这样的突变可能抵消所有的正在累积的有害突变所带来的破坏，退化可能因此而停止。比如说，一个核苷酸置换可能增加基因组信息量的1%。这样可以有效地补偿功能性基因组1%的突变（甚或缺失）。在这个假设的情形中，一个点突变可能创造3千万个核苷酸位点所包含的信息。如此一来，只要几个超级有利突变，便可在理论上抵消几百万个有害的点突变。

此议无效：

以上方案不可行，原因有四：

1. 基因组的还原主义模式，将基因组基本上看作一个装着基因和核苷酸的袋子，每个基因或核苷酸的效果基本上是相加性的。在这

个模式中，信息只能一次建立一点，像一粒一粒地堆沙子一样。这一模式即便在大段DNA重复的特例中也是一样。直到有利的点突变一个一个地引入重复区以后，重复区域才会产生新信息。还原主义理论家也许嘴上承认相互作用和协同效应的重要性，但他知道在现实中爬上“无望巅峰”的唯一路径是一长串无穷小的步骤。我们可以理性地扪心自问：“我们能从一架喷气飞机组装说明书的拼写错误中，期望得到何种改进呢？”显然，如果有任何改进的话，永远不会很显著。最多有些微小的调整，缓慢增加的改进是基本公理的核心所在。在基本公理下，我们可以很放心地说，基因库的信息是一次添加一点缓慢地填满的。

2. 在一个有着30亿个信息单位的基因组内，一般的有利突变只能增加的信息的30亿分之一。确实，有些有利突变会比其他突变的益处更大，突变效应形成一个自然分布，但是完全没有理由相信，一个有利的点突变能够增加3千万个功能性核苷酸所携带的信息。

有人可能会这样反驳：“肯定存在一些致死的有害突变。在这些情况下，一个点突变就能有效地抹杀30亿个信息单位。公平起见，反过来，有利突变不也应该一样吗？最强的有利突变不也应该等同于30亿个信息单位吗？”这样的想法把我们又带回到那个天真的、对称的钟型突变分布曲线。然而我们知道这种观点被普遍否定了。为什么？

突变效应的极端不对称性是缘于我们要爬“无望巅峰”这一现实。确实可以想象一个错误可以使你向上摔个跟头，但摔出只有几尺远。你永远也不会摔一个跟斗翻上几千尺。然而，一个错误却可以很轻易地使你向下跌出很远。事实上，在攀登无望巅峰时，你可能很容易就冲下很长的距离，造成即刻死亡。同样地，如果你用纸牌来建房子，失败很容易，而且经常是灾难性的，然而你只能一次一张牌地向上垒。很类似地，突变的效果也是极不对称的。

3. 用几个超级有利突变来弥补数百万个其他核苷酸位点上丢失的

信息是不明智的，这会导致荒谬的结论。照此逻辑，只要有100个超级有利突变，每个突变增加1%的信息，就能取代整个基因组。可以删掉基因组内所有30亿个碱基对，只用一个由100个超级有利核苷酸构成的基因组就可替代。再进一步来说，如果我们能够想出一个使基因组信息倍增的突变（致死性突变的镜像），其余的全部基因组就可以删除了。我们还保持着1.0的适合度，但基因组中只有一个核苷酸！如果存在一种机制，可以用少数几个超级有利突变来代替因突变而不断退化的数百万个其它携带信息的位点，那么这种机制就必导致有效基因组的规模持续而迅速地缩减，这显然是不可能的。这就好比说，如果每增加一个新字母就删减1000个字母，仍然可以改进一本书的内容。

4. 经常被引用的貌似“超级有利突变”的例子有极大的误导性。为说明问题，让我们来考虑细菌对抗生素的耐药性、狗的皮毛厚度和果蝇的同源异型突变。

使细菌产生抗生素耐药性的染色体突变，似乎就是超级的有利突变。当环境中存在抗生素时，其他细菌都会死亡，唯有突变株存活。所以突变体的适合度比其他细菌不是增加一倍，而是增加了无穷倍，从0增到1！

如果你把一只（北极）撒摩耶犬放到炎热的沙漠里，它会死掉。而一个“无毛”突变会产生对极度炎热的适应性，所以突变狗会存活。我们再次看到适合度从0增至1，一个无穷大的增长！

就适应某种特殊环境而言，以上当然都是“超级有利突变”的例子。然而两者皆为失去功能并减少基因组净信息的突变。就信息内容而言，两者都还是有害的突变。几乎所有貌似超级有利突变的例子都涉及对新环境的适应。这只不过是一种微调，而不是建立基因组。这类改变的戏剧性并不是由于生物体真正地“升级”了，而仅仅是因为其他个体都死亡了！突变体所谓的“改进”是相对于其他死了的竞争

对手而言的。这类突变并不增加信息，也不创造更多的有意义的复杂性，更不在任何意义上创造更高形式的生命。

令人遗憾的是，进化论学者们将两个非常不同的现象，对环境的适应和向更高生命形式的进化看作一回事。我们不用成为天才也能判断这是两码事。适应通常可以通过信息丢失甚至发育退化（器官消失）来达到。然而，演变成更高的生命形式（表现更多的有意义的复杂性）总是需要增加大量的信息。

有一种特殊类型的突变可以对机体的发育产生重大影响：发生于“同源异型框”基因内的突变。这些突变可以导致大规模的器官重组。例如，一个突变可以把一个昆虫的触角变成一条腿，或者可以让一只苍蝇长出两对翅膀而不是正常的一对。这些突变可称得上是超级突变。这样通过简单突变而带来的体形上的戏剧性改变使得很多进化学家们兴奋不已。这种突变类型创造了一个全新的猜想领域，美其名曰“EvoDevo”（进化性发育）。普遍认为这种突变类型提供了基本公理所需的宏观有利突变，而且可能带来进化跳跃（大跃进）。

确实可以想象同源异型突变带来的宏观改变有时或许是有利的。甚至可以想象它们可能是非常有利的。但实际上这种情况发生的频率有多少？这样的点突变真的能够抵消全基因组范围的退化吗？

就一架喷气式飞机的说明书而言，一个拼写错误有可能把“重复环路3次”的指令转变成“重复环路33次”。或者，一个拼写错误有可能把“将21号组合接至机体A部”转变成“将21号组合接至机体Z部”。这种文字处理错误可能导致飞机形体上很大的改变，但会不会是有利的呢？如果这些改变是有利的，它们能否有效地抵消数百万个使飞机部件降级的其他拼写错误所带来的信息损失呢？

我们可以在理论上承认同源异型突变可能偶尔在某些方面有用。然而，以上所举的实例其实是非常有害的。触角变成腿的苍蝇不过是一个怪物。易位的器官既没有触角的作用也没有腿的功能。在两对翅

膀的苍蝇，多出来的翅膀也是无用的（它们不与肌肉或神经联结），只会干扰正常翅膀的功能，因而突变的苍蝇几乎不能飞。显然说明书中的某些随机改变会在终产品中造成极大的偏差。但这能在任何意义上支持超级有利突变的说法吗？这能揭示一个宏观突变能够增加总基因组信息达1%（相当于人体细胞中3千万个核苷酸）之多吗？这能说明一个这样的突变可以抵消基因组中数百万个突变的退化效应吗？显然不能！

我们的结论是，不能用超级有利突变之说支持基本公理。

异议3 噪音会均衡

如果一个群体几乎无穷大，而且完全均匀，而且如果“噪音”是持续而均一的，而且如果有无限的时间，所有的噪音效应最终会达到均衡，这样即使近乎中性的突变也可以被选择。在这些条件下，可以想象自然选择可能最终遏制退化。

此议无效：

以上列出的消除噪音的基本要求，在自然界中绝不存在：

1. 群体规模从来不会无穷大，对于人类来说，群体规模在最近几千年中才真正庞大起来。进化论者所假设的有效人类进化群体只有1万人。那个小群体从来没有形成均一的基因库，而只可能由孤立的分支群体组成，比如100个部落，每个部落约100人。自然选择主要限于部落之间的竞争。在这种情况下，不可能有显著的噪音均衡。

2. 噪音从来不是均一的。特别是环境噪音在空间上和时间上都非常不稳定。对某个特定区域的一个部落，非遗传噪音最重要的来源可能是极端的气候变化，但对于另外一个区域的一个部落来说，可能是疾病。也许有很多世代，营养不同是干扰选择的的主要噪音来源，随

之却发生了周期性的疾病或战争。在这些情况下，噪音难以均衡。

3. 随着突变的积累和适合度的下降，基因组内部的环境也将改变。对于一个特定核苷酸的选择来说，不断累积的其它分离性突变造成的噪音会逐渐增多。虽然有些环境噪音会随适合度按比例增减（因而随着适合度的下降噪音也按比例减弱），另外一些环境噪音却不会随适合度而改变，比如自然灾害带来的噪音。这些不随适合度下降而相应减弱的噪音，会随着适合度的下降而对选择造成越来越大的干扰。如果噪音持续增大，就谈不上有效地均衡与中和。

4. 当噪音很大时，选择的作用基本上被淹没，导致迅速的和灾难性的突变积累。鉴于我们的低繁殖力和高突变率，在我们这个种族灭绝之前噪音均衡的时间太少了。就算真有噪音均衡，在有效均衡之前也要先有大量的选择。既然假想的人类进化群体很小，要实现大量选择的唯一途径就是跨越很多世代。然而，早在噪音均衡能够有效地提升选择效率之前，进化中的人类群体就已经灭绝了（甚至可能在1000代之内），在达到突变平衡之前适合度早已降到零值。噪音均衡，即便真有其事，似乎也不足以在误差灾难和种族灭绝之前有效地遏制退化的过程。

异议4 - 基本公理的失败对于进化思想不是一个严峻的挑战

如果基本公理有致命的缺陷而且基本上被否定，那又如何？基本公理只是众多进化机制之一，所以对进化理论无关紧要。进化论研究者只需要更多一些时间和更多一些资金，来理顺他们的各种理论中的几个“小枝节”。

此议无效：

这是试图控制损失的说法，是明显错误的：

1. 只有一种进化机制，这个机制就是突变/选择（即基本公理）。如果信息是自发地产生的，除了突变/选择，别无其他可行的机制。说突变/选择是众多进化机制之一是错误的。突变有几种，选择也有几种，但基本进化机制只有一种（突变/选择）。基本公理一旦灭亡，进化理论就失去了有活力的机制。没有了自然主义的机制，进化论和基于信仰的任何宗教便毫无两样了。

2. 达尔文的唯一真正有创意的概念，即是把自然选择定为自然界中的原始创造力。但他那时还没有遗传学或突变的概念，所以还不知道被“选择”的究竟是什么，也不知道被“创造”的是什麼。因而他对于本书所论及的所有问题一无所知。他的大体观点，即自然选择可以解释生物学的所有方面，只是他极力鼓吹的哲学立场而已。直到很多年后，新达尔文主义者才综合了遗传学、突变和自然选择，创立了群体遗传学这一领域。至此，达尔文主义才以真正科学的面目出现。自从突变/选择机制被建立，直至今日，它一直是贯穿达尔文主义思想各个方面的唯一轴心。

3. 退化恰是进化理论的对立命题。因此基因熵的真实性对于达尔文主义来说必定是致命的。许多人宣称智能设计的概念不能科学地研究，所以只能是个信仰问题。然而在生物学上，智能设计的“无效假设”显然是突变/选择。我们都知道，否定了无效假设就是对备择假设的有力支持。所以，任何说明突变/选择不能创造或维持基因组的科学证据就是支持智能设计的有力的科学证据。

atcg

附录 6

学术界对这些发现有何反应？

《基因熵》一书

九年前这本书首次发表的时候，我自己也步入了一个新的领地（群体遗传学）。在这个相对封闭的圈子里，我是个不受欢迎的新人。我请求该领域的几个人在本书发表前批评指正，但他们根本不理睬，所以我只好在没有得到多少反馈的情况下发表。《基因熵》从初版至今已有九年，结果如何？

作为一部技术专著，本书在保守人士的圈子里读者不少，评价也不低。我相信许多进化生物学家也读过——这毕竟是公开地挑战他们的模式的主要著作之一。有没有人严肃地反驳本书的结论？从该领域内有资格回应的科学家那里，我听到的只是振聋发聩的沉默。虽然我奉上免费的样本和相关科研论文，礼貌地寻求中肯的反馈，结果还是一样。即使从前愿意与我进行有限的对话的相关科学家，也对这种友好的咨询置之不理。这让我难以理解，因为科学只有在开诚布公的讨论中才能进步。我知道有几个达尔文主义的博客曾做出一些反应——主要是靠人身攻击和谩骂。然而就我所知，至今还没有严肃的辩驳。我断然没有听到该领域内有名望的人做出任何科学上的回应。只有一个博客（一个生物学的外行），写了一个长篇攻击，但他的观点浅薄，他的说法已经在博网上得到了适当的反驳。

学术界对这本书和我在过去十年中发表的科学论文不置一词。我的一位好友，也是一位遗传学家和忠实的进化论者，仔细地阅读了此书。然后我问他：“他们为何不理睬我的论证？”他的回答很简单却有很震撼力：“因为他们没有答案。”我相信他是对的。他们没有答案，这其实不足为奇。他们中间许多人已经在自己的论文里悄悄地承认了我所总结出的一切问题。在我之前，大多数群体遗传学界的泰斗们早已意识到了本书中所提出的这些问题的真实性（见附录1）。我确信，今天该领域的领袖们也理解并在私下里承认我正在研究的这些问题。在我看来，他们之所以回避公开的对话，唯一的原因就是，他们想把这些基本的理论难题当作行业秘密，不宜公开对话，不让大众讨论。如果是这样，我希望他们保不住这些“行业机密”。真正的科学只有在开诚布公的、平心静气的对话中才能进步。科学上不能有至圣神主牌——新达尔文主义也不行。我盼望着新达尔文主义理论的深层问题能暴露于光天化日之下，受到坦诚的、公开的审查。

总而言之，九年之后，这本书基本上还没有污点，不需要收回什么，不需要向谁道歉。事实上，这些年的科学发展极大地强化了本书的主题。迄今，我最大的尴尬是一些打印错误和拼写错误。我最大的遗憾还是学术界的沉默、躲闪和屏蔽，这些做法竟然如此有效地压制着对新达尔文主义的深层问题的开诚布公的审查。

《孟德尔的会计师》程序

《孟德尔的会计师》是用全面的、符合生物学现实的数值模拟来研究突变/选择过程的第一个软件，开创了该领域的先河。它是一个得力的工具，既可用于教学，又可用于探索群体遗传学上一些无可企及的研究课题。八年来，该领域拒绝测试和使用这一重要工具，甚至不愿意承认它的存在。然而《孟德尔》是今天解决群体遗传学上的

大问题的最有效的工具。一个领域为何要存心忽略能促进自身进步的一个新工具？原因其实很明显。首先，这一新技术的研发者“有政治问题”——大半是些保守的基督徒。在西方学术界，虔诚的基督徒就好像长了麻风病。对于这“一小撮”被鄙视的异类，西方学术界所注重的那套准则，包括宽容、学术自由、公正审查、甚至起码的尊重和公道原则，都用不着了。如果出版物的作者是某些“有政治问题”的人，出版物就可以被拒绝或忽略，并不需要什么理由。科学本来不应该是这样做的。如果作者是无神论者、穆斯林、印度教徒、信巫术的、信地母盖亚的、信外星人的，都不要紧。甚至自由派的基督徒也常常被西方学术界接纳——只要他们肯向达尔文主义的机制低头。我们所看到的是一种颇有针对性的偏见，这在大多数校园里是非正式批准的，完全是针对保守的基督徒科学家和学生的。想想真有意思，为什么有这种情况？为什么学术界大多数人不起来反对？

至于该领域无视这一有力的新研究工具的第二个原因，我相信是由于最先进的数值模拟结果总是令该领域多数玩家尴尬。既然新证据表明皇帝是赤身裸体，他们便要不惜一切代价忽略这些发现。他们自己早该看到了，可是他们被大能的“群体思维”蒙住了双眼。他们之所以被蒙蔽，还因为他们把某些“伟人”偶像化，但事实上“伟人”也和我们一样，会犯错误。

《孟德尔》的模拟结果科学地、经验性地显示了新达尔文主义理论行不通。看起来，群体遗传学领域的许多人似乎是故意地将这一简单的事实向科学界大部分人隐藏。若是这样，他们就是在故意拖科学的后腿，他们这样做想必是因为许多人为了捍卫一个垂死的、十九世纪的思维模式倾注了自己的事业和人生。我希望该领域有人为了个人道义和职业道义从众人中站出来，在所有这些课题上开启公开的、坦诚的对话。如果我对基本公理的挑战是误导视听，那就请同僚们指明错误所在，我必将改邪归正。如果他们说不出我在这许多议题上错在

哪里，他们就有必要公开地、真诚地承认：现在占统治地位的模式确实有严重的问题。

atcgg

词汇表

适应 (Adaptation): 生物体与外部环境相互作用，而这种相互作用受基因突变的影响。突变能够使生物体适应环境的能力更好或更差。适应性突变是有利突变的最佳例证，而对这些突变的选择是“微观进化”的基础。然而基因组内的大部分信息与环境适应无关，而是控制一个生物系统奇妙复杂的内部活动。

适应性退化 (Adaptive degeneration): 对特殊环境的适应通常不是靠创造新基因或新基因功能而实现的，而是缘于有用功能或有用变异类型的丧失。适应性突变通常带来信息损失，一般是通过基因或基因启动子的破坏或缺损。

染色体 (Chromosome): 大多数基因组是由称为染色体的亚单位组成的。基因组中的染色体类似于一套大百科全书中的每一卷。人类的每一条染色体通常含有数千万个核苷酸和数千个基因。

选择的代价 (Cost of Selection): 选择的代价是群体中由于选择作用而不能繁殖的百分比。为了让选择发挥作用，每一代都必须有剩余群体，这样才能把不适应的个体淘汰掉。只要群体持续增长，这个代价暂时可以得到控制（每个生物体都能繁衍，只是有些比其他繁殖得更多），然而从长远来看，群体的增长必然只是阶段性的。

熵 (Entropy): 该术语，在最通俗的用法中，是指在没有智能干预的情况下一切物体都要衰退的普遍趋势。具体地说，熵就是混乱度，而混乱度总是不断增加的。熵有时也被定义为“事物从有序到无序的自然倾向”。它也被描述为“机器不能以100%的效率运作”。它也被称为“时间之箭”，因为总的来说熵总是在增加的。熵所带来的退化是普遍的，而且我们天天都能体验到这种力量。

上位效应 (Epistasis): 不同的突变常有相互作用, 这种相互作用就是上位效应。一个有害突变的效应可因其它突变的存在而加重或减轻。上位效应因产生不具遗传力的“噪音”而严重干扰选择的作用。遗传学家们承认上位效应的重要性, 但假定正反面的相互作用大体上互相抵消。

误差灾难 (Error catastrophe): 指一种生物学状况。在此状况下, 有害突变积累的速度比选择消除它们的速度快, 造成每一代的适合度持续下降。除非被逆转, 误差灾难将导致一个种群的灭绝。

功能性信息 (Functional information): 当我们在正常意义上使用“信息”这个词时, 指的就是功能性信息。功能性信息表达一个特定的内涵, 并有特定的目的。数据压缩使我们能以较少的数码或字符来表达较多的功能性信息。

功能门槛 (Functional threshold): 一个字符串需要一定的字符数目才能表达一个信息。孤立的二进制码或孤立的核苷酸是没有信息的。一个典型的句子最少需要十二个字符。一个典型的基因需要1000个以上的字符。正在构建中的新基因在获得了起码的功能之后才能受到正向选择的作用, 也就是说, 在足够数量的核苷酸到位之后才能达到功能门槛。

基因 (Gene): 染色体内具有生物功能的单位。基因是由几千到几百万个核苷酸组成的。染色体中的基因就像一卷书中的每一章。一个典型的人类基因是多功能的, 编码许多不同的RNA和蛋白质。基因是由非常复杂的调节机构, 包括远距离DNA序列来调节的。

基因库 (Gene pool): 基因库的概念抽象, 无现实意义。基因(等位基因、核苷酸)从不存在于自由混合的库里, 而只存在于活生生的有机体中, 存在于染色体内特定的连锁块中。当我们需要形象地描述不同等位基因在一个群体中独立地增多或减少的情形时, 基因库的概念有时是有用的。

基因漂变 (Genetic drift): 新一代生物体都是从前一代所产生的配子中取样形成的。随机取样误差将导致所有等位基因的频率缓慢增加或减

少。孟德尔分离律和随机配子取样不可避免地造成基因漂变。漂变在小群体中尤为明显, 能对抗几乎所有的选择。在一个极小群体中, 选择基本上被中和了(对半致死突变的选择除外)。等位基因随机地迅速流失或固定, 与适合度无关。

基因熵 (Genetic entropy): 广义上熵的概念也适用于生物学和遗传学。除非有智能干预, 独立生活的生物体内的功能性基因组信息必将持续减少(或许有些病毒例外)。如世界上的其他层面一样, 在生物范畴内的“自然之矢”是退化, 混乱度持续地与时俱增。

基因连锁 (Genetic linkage): 染色体是一条线性分子链, 如一长串的文字信息。一对同源染色体能互相交换整块的信息, 好像两本书能交换它们最后的篇章一样。所交换的信息只限于大的DNA连锁块。好像篇章可以交换, 而单个的词或字母通常不能交换。在一个连锁块(篇章)内, 所有的字母和词语是连着的, 不会分离。一个典型的人类基因连锁块包含大约3万个核苷酸, 它们不可分割, 因此一个连锁块是一个独立的遗传单位。

基因负荷 (Genetic load): 一个群体中有害突变的总和为其基因负荷。当基因负荷持续增加时, 这一群体处于误差灾难中。

基因组 (Genome): 一个有机体的全部基因含量。包括所有DNA、所有基因和所有染色体。功能性基因组是除去任何无生物意义(没有功能性信息)的部分之后的整个基因组。

基因型 (Genotype): 基因型是个体化的基因组版本。是存在于特定群体中特定个体内的等位基因的特定组合。

基因型值 (Genotype value): 一个生物体基因组的信息总量, 是个体全部生物功能中由基因组决定的那一部分。由于环境因素(表型噪音)也影响个体生物功能的强弱, 所以基因型值与表现型值不同。某一特定突变对基因型值改变的程度反映了它对适合度的影响。因此基因型值也称基因型适合度。

霍尔丹的困境 (Haldane's Dilemma): 著名的遗传学家霍尔丹在分析选

择的代价时发现，即使一个物种不断产生一定数量的可选择有利突变，这些突变经选择而固定的速度也会极其缓慢（每一个突变需要300代才会被固定）。这意味着正向进化即使在地质时间内也是不现实的。也有其他学者验证了这一问题。

遗传力 (Heritability): 衡量选择过程中信号与噪音比值的传统方法。具体来说，是可遗传变异与总变异的比值。总变异包括可遗传变异（信号）和非遗传变异（噪音）。非遗传变异包括所有环境作用、基因和环境共同作用、上位效应、表观遗传效应和显性效应。

信息 (Information): 该词语最有用的定义就是其通常的字面意义。信息就是语言所交流的内容。由于生命需要大量复杂的交流网络，所以生物信息的形式也多种多样。

整合复杂性 (Integrated complexity): 完美的整合复杂性导致合一现象。一个拼图的每个部分与其它部分完全连接，没有缺失的部分，没有多余的部分，构成了完美的整体。“共存”或“预先适应”都不能解释这种完美（这两种机制所产生的是“偷工减料”的装置）。自然选择不具备近乎无限的分辨率来造就这种完美。一堆废品是一个复杂物，但却不是一个互联整体。一个完好的拼图和一架能飞行的飞机却能够显示出整体复杂性。

木村资生的盒子 (Kimura's box): 穆勒意识到对适合度影响不大的突变是不可选择的。木村资生正式地将这一原则量化并给出“实际中性”（即不可选择的）突变的一个界限。木村认为大多数突变都在此范围内，且“实际中性”的原则对于有利突变和有害突变同样适用。在描绘突变对适合度的影响时，可以看到中值两侧的一个区域（或“方格”）。在一个遗传系统中，非遗传噪音越多，这个“非选择”性方格就越大。木村指出群体大小有助于确定这个非选择区域的范围（小群体因为配子取样而存在更多的噪音）。然而，这个非选择区域也为其他各种噪音所影响，包括环境噪音和繁殖噪音。在大基因组中，绝大多数突变应存在于木村的非选择方格之内。

孟德尔的会计师 (Mendel's Accountant): 第一个全面地、符合现实地模拟基因组在突变/选择作用下随时间而变化的计算机数值模拟程序。

穆勒的棘轮 (Muller's Ratchet): 一个基因组中所发生的好突变与坏突变常呈一定的比例，而坏突变总是占多数。所以，要阻止基因退化，就必须把罕见的好突变和大量的坏突变分离开来。在有性生殖过程中，如果在好突变和坏突变之间有基因重组，这种分离就是有可能的。然而，在缺乏两性重组的基因组中，或在没有重组的基因组区域，好突变与坏突变紧密连锁，不可分割。在这样的基因组或基因组区域内，占大多数的坏突变就不能与罕见的好突变分离，造成一个向下沉沦的棘轮机制，其净效果是基因组的不断退化。

突变等位基因 (Mutant allele): 所有的新突变最初在群体中只有一个拷贝。该拷贝的数目和频率可以随时间而增加。一个特定的突变等位基因在群体中的频率可能增加也可能减少。当一个突变等位基因频率达到1时，就达到固定状态（群体中的所有个体都有两个突变拷贝，非突变等位基因已经在群体中消失了）。

突变等位基因频率 (Mutant allele frequency): 某个基因位点上突变型等位基因占群体中等位基因总数的百分比（在二倍体生物中，等位基因总数是群体规模的两倍）。假如在100个个体组成的群体中有两个突变拷贝，突变等位基因的频率就是1%，非突变等位基因频率为99%。

突变位点 (Mutant locus): 突变在染色体内一个连锁块中的位置。

突变 (Mutation): 基因组的改变，好象一个文本在抄写过程中出现的词语错误，包括任何置换、删除、插入、重复、颠倒等。

突变熔消 (Mutational meltdown): 误差灾难的最后阶段。随着一个群体的适合度逐渐减弱，群体的繁殖力势必下降。但要排除不断累积的有害突变，就需要不断产生多余的个体以供选择。当每个母体产生的后代数量越来越少，可供选择淘汰的多余个体数量也减少。结果导致适合度加速减弱，繁殖力也随之快速降低，选择的作用愈来愈小。最终的退化结果是群体的迅速瓦解和突然性的绝种。

突变适合度效应 (Mutation fitness effect): 突变的生物效应。突变适合度效应可以是正面的，也可以是负面的，影响可大也可小。突变效应

表现在个体的总生物功能因突变而带来的相对变化。一个适合度效应为-0.01的有害突变使个体基因型值减少1%。基因组失去1%的信息，总生物功能降低1%。这些是描述突变生物效应的不同说法。突变效应与环境变异（表型噪音）和繁殖随机性（繁殖噪音）无关。突变效应与传统的选择系数相似但有所不同（参看 Sanford et al., 2007a）。

突变/选择链 (Mutation/selection chain): 真实的生物世界中突变事件和选择事件之间的一系列事件。突变影响连锁块，连锁块影响染色体，染色体影响基因型，基因型影响表现型，表现型影响个体的繁殖适合度，进而影响突变向下一代的遗传。在每个链接之间都有生物噪音。

自然选择 (Natural selection): 自然选择是群体中生物功能较差的个体比生物功能较强的个体繁殖率低的自然趋势。当生物功能较强的个体得以繁衍时，常称为“适者生存”，但称之为“适者繁殖”更确切些。在自然界，繁殖还有大量的随机性。这一重要原则可称为“最幸运者生存（繁殖）”。因为两个原则在自然界中同时运作，适合度最强的个体只是繁殖概率稍微高一点而已，甚至可能会因为随机事件而不能繁衍。这种天然的选择形式也称为“概率选择”。它是一种很低效的选择形式（参看斩截性选择）。

核苷酸 (Nucleotides): 核苷酸是4种不同的分子，是DNA梯状结构里的横档，通常以字母A、T、C、G来代表。高级基因组中的很多核苷酸具有多功能性，同时对多个生物过程产生影响。

表型组 (Phenome): 是整个功能性生物体，包括基因型和表现型的所有方面，是一个基因组中的全部信息在特定环境中的外在表达形式。

表现型 (Phenotype): 实际表现出来的具有生物功能的个体。表现型受基因型和周围环境因素的影响。基因型和表现型互相关联，但有所不同。

表型值 (Phenotype value): 某一个体的实际生物功能的强弱，是相对于群体中其他个体而言的，是由基因型效应和环境效应所共同决定的。表型值是“自然母亲”所实际看到的，是选择的直接作用对象。

它与生物学家通常用的“表型适合度”或“生物适合度”同义。然而表型值（表型适合度）这一概念与群体遗传学家正式定义的“适合度”有所区别。我们将用繁殖适合度来指传统群体遗传学家所定义的适合度（繁殖概率），以便与表型适合度区别开来。

多功能DNA (Poly-functional DNA): 毋庸置疑，高级基因组内存在相互重叠的多种功能性信息系统。这就意味着许多核苷酸有不只一种功能（像一个字母在拼字游戏中可以同时是两个单词的组成部分）。多功能DNA由于受到多重限制，很难发生有利突变。

群体 (Population): 相互交配繁殖的个体的集合。群体大小就是个体的数量。

基本公理 (Primary Axiom): 一切达尔文主义思想的基础信条，即突变加自然选择能够解释生命的各个方面。也称为新达尔文主义理论。

繁殖适合度 (Reproductive fitness): 我的定义是表型值（表型适合度）与“繁殖噪音”的乘积。繁殖噪音的存在是因为繁殖再生不仅取决于生物功能的强弱，也受随机因素的影响。所以表型适合度和繁殖适合度虽互相关联，但不是一样的。表型值和繁殖适合度的相关程度取决于选择的方式。人工斩截将产生最强的相关，而天然的概率选择则产生最弱的相关。“繁殖适合度”有时也称“达尔文适合度”或“赖特适合度”。索维尔·赖特(Sewell Wright)是第一位用数学方法描述“达尔文适合度”的科学家。

选择 (Selection): 选择可看作差异性繁殖，在此过程中，有些个体比其他个体繁殖得少。有些个体被排除而不能产生后代。除非有多余的部分群体（过度繁殖），选择必因一些个体不繁殖导致群体数量缩减。

选择干扰 (Selection Interference): 对群体中一个性状的选择会影响到同一群体中另一性状的选择。

选择门槛 (Selection threshold): 如果一个核苷酸或一个核苷酸系列所引起的生物效应不够大，不能超出选择门槛，那么它就免于选择。该

门槛取决于许多因素，这些因素就是干扰选择的“生物学噪音”。当突变的适合度效应低于选择门槛时，自然选择就看不见它。

沙农熵 (Shannon entropy): 一种复杂性（“非简单性”）的统计方法。这种统计帮助确定潜在信息的下限。按序排列的字母（a, b, c, …）的沙农熵很低，因为这个序列是固定的可预测的。固定顺序限制着它所能携带的信息量。所以低沙农熵意味着潜在信息受限。同样一组字母可以通过重排使之“复杂化”。将字母排列成有内涵的词句增加了沙农熵，但仅把字母随机打乱也会增加沙农熵。第一种情况产生的是真实的功能性信息（通过智能）。第二种情况中只有潜在的信息（字母的随机组合）。因此，沙农熵的增加不应该与功能性信息的增加相混淆。真实的信息沙农熵较高，随机排列的字符也是一样。

沙农信息 (Shannon information): 一种统计潜在信息的方法。增加字母总数和增加字母种类都会增加潜在信息。然而，除非字母按特定的有意义的顺序排列，功能性信息总是等于零，并不能表达什么。打乱一个词句的字母，虽然潜在信息量没有改变，但没有了功能性的（实际的）信息。

协同上位效应 (Synergistic epistasis): 协同上位效应是通常用于解释如何阻止基因组不断退化的一种说法。基本概念是有害突变之间的上位效应（相互作用）可能总是负面的。所以，当突变增多时，每个新的突变平均会有更大的破坏效应。这是与标准的相乘性群体遗传学模型正好相反的，后者认为随着突变数量的增加，每个突变的效应减少（两者必有一个是错误的，或者都是错误的）。协同上位效应模型是极端人为的，不符合生物学现实。即便有这个模型，考虑到突变与非突变之间的连锁和相互作用，这种机制也不能阻止退化。

斩截性选择 (Truncation selection): 在动植物育种中，人工选择的典型方法是斩截性选择。最好的个体一定被选择，最差的个体一定被排除。这种选择形式在自然界中通常无法实现。在特殊情况下，除了少数特定基因型（如抗药性细菌）存活，其他个体死亡殆尽，这就是偶尔发生的斩截性选择，但针对的只是一种特别性状。这是一种很强的选择，自然群体中罕见，只有在针对一个特定性状进行极强力的选择

时，才会偶尔发生。



Anzai, T. et al. 2003. Comparative sequencing of human and chimpanzee MHC class I regions unveils insertions/deletions as the major path to genomic divergence. PNAS 100: 7708-7713.

Barrick, J.E. et al. 2009. Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. Nature 461:1243-1247.

Bataillon, T. 2000. Estimation of spontaneous genome-wide mutation rate parameters: whither beneficial mutations? Heredity 84:497-501.

Bataillon, T. and S.F. Bailey. 2014. Effects of new mutations on fitness: insights from models and data. doi:10.1111/nyas.12460 Ann. N.Y. Acad. Sci. (2014) 1-17.

Baumgardner, J. et al. 2008. Mendel's Accountant: A New Population Genetics Simulation Tool for Studying Mutation and Natural Selection. In A. A. Snelling (Ed.) (2008). Proceedings of the Sixth International Conference on Creationism (pp. 87-98). Pittsburgh, PA: Creation Science Fellowship and Dallas, TX: Institute for Creation Research. <https://www.icr.org/i/pdf/technical/Mendels-Accountant.pdf>

Baumgardner J., W. Brewer and J. Sanford. 2013. Can Synergistic Epistasis Halt Mutation Accumulation? Results from Numerical Simulation, In: Marks II R.J. et al., (eds) *Biological Information – New Perspectives* (pp 312-337). http://www.worldscientific.com/doi/pdf/10.1142/9789814508728_0013

Behe, M. 1996. *Darwin's Black Box: Biochemical challenge to Evolution*. The Free Press. NY, NY.

Behe, M.J. 2007. *The Edge of Evolution*. Free Press. NY, NY.

- Bejerano, G., et al. 2004. Ultraconserved elements in the human genome. *Science* 304:1321-1325.
- Bergman, J. 2004. Research on the deterioration of the genome and Darwinism: why mutations result in degeneration of the genome. Intelligent design Conference, Biola University. April 22-23.
- Bernardes, A.T. 1996. Mutation load and the extinction of large populations. *Physica ACTA* 230:156-173.
- Brewer, W., J. Baumgardner and J. Sanford. 2013a. Using Numerical Simulation to Test the "Mutation-Count" Hypothesis. In: Marks II R.J., et al., (eds) *Biological Information – New Perspectives* (pp 298-311). http://www.worldscientific.com/doi/pdf/10.1142/9789814508728_0012
- Brewer, W., F. Smith and J. Sanford. 2013b. Information loss: potential for accelerating natural genetic attenuation of RNA viruses. In: Marks II R.J., et al., (eds) *Biological Information – New Perspectives* (pp 369-384). http://www.worldscientific.com/doi/pdf/10.1142/9789814508728_0015
- Britten, R.J. 2002. Divergence between samples of chimpanzee and human DNA sequences is 5% counting indels. *PNAS* 99:13633-13635.
- Campbell, C.D. and E.E. Eichler. 2013. Properties and rates of germline mutations in humans. *Trends in Genetics* 29:575-584.
- Carlsen, E., et al. 1992. Evidence for decreasing quality of semen during past 50 years. *BMJ* 305:609-613.
- Carter, R.W. 2007. Mitochondrial diversity within modern human populations. *Nucleic Acids Research* 35(9): 3039-3045.
- Carter, R.W. and J.C. Sanford. 2012. A new look at an old virus: patterns of mutation accumulation in the human H1N1 influenza virus since 1918. *Theoretical Biology and Medical Modeling* 9:42doi:10.1186/1742-4682-9-42. <http://www.tbiomed.com/content/pdf/1742-4682-9-42.pdf>
- Charlesworth, B. 2013. Why we are not dead one hundred times over. *Evolution* 67:3354-3361.

- Chen, J., et al. 2004. Over 20% of human transcripts might form sense-antisense pairs. *Nucleic Acid Research* 32:4812-4820.
- Crow, J.F. 1997. The high spontaneous mutation rate: is it a health risk? *PNAS* 94:8380-8386.
- Crow, J.F. and M. Kimura. 1970. *An Introduction to Population Genetics Theory*. Harper and Row. NY, NY p. 249.
- Dawkins, R. 1986. *The Blind Watchmaker*. Norton & Company, New York.
- Demski, W. 1998. *The Design Inference: Eliminating Chance Through Small Probabilities*. Cambridge University Press.
- Dennis, C. 2002. The brave new world of RNA. *Nature* 418:122-124.
- Elena, S.F., et al. 1998. Distribution of fitness effects caused by random insertion mutations in *E. coli*. *Genetica* 102/103:349-358.
- Elena, S. F. and R.E. Lenski. 1997. Test of synergistic interactions among deleterious mutations in bacteria. *Nature* 390:395-398.
- Ellegren, H. 2000. Microsatellite mutations in the germline: implications for evolutionary inference. *TIG* 16:551-558.
- Eyre-Walker, A. and P. Keightley. 1999. High genomic deleterious mutation rates in Hominids. *Nature* 397:344-347.
- Eyre-Walker A, and P.D. Keightley. 2007. The distribution of fitness effects of new mutations. *Nat Rev Genet* 8:610-618.
- Felsenstein, J. 1974. The evolutionary advantage of recombination. *Genetics* 78: 737-756.
- Flam, F. 1994. Hints of a language in junk DNA. *Science* 266: 1320.
- Gabriel, S.B., et al. 2002. The structure of haplotype blocks in the human genome. *Science* 296:2225-2229.

- Gardiner, K. 1995. Human genome organization. *Current Opinion in Genetics and Development* 5:315-322.
- Gerrish, P.J. and R. Lenski. 1998. The fate of competing beneficial mutations in an asexual population. *Genetica* 102/103: 127-144.
- Gibbs, W.W. 2003. The hidden genome. *Scientific American*. Dec.:108-113.
- Gibson, P., J. Baumgardner, W. Brewer, and J. Sanford. 2013. Can Biological Information Be Sustained By Purifying Natural Selection? In: Marks II R.J. et al., (eds) *Biological Information – New Perspectives* (pp 232-263). http://www.worldscientific.com/doi/pdf/10.1142/9789814508728_0010
- Gitt, W. 1997. *In the Beginning was Information*. Literatur-Verbreitung Bielefeld, Germany.
- Gitt, W., R. Crompton, and J. Fernandez. 2013. Biological Information – What is it? In: Marks II R.J. et al., (eds) *Biological Information – New Perspectives* (pp 11-25). http://www.worldscientific.com/doi/pdf/10.1142/9789814508728_0001
- Hakimi, M.A. 2002. A chromatin remodeling complex that loads cohesion onto human chromosomes. *Nature* 418:994-998.
- Haldane, J.B.S. 1957. The cost of natural selection. *J. Genetics* 55:511-524.
- Higgins, K. and M. Lynch. 2001. Metapopulation extinction caused by mutation accumulation. *PNAS* 98: 2928-2933.
- Hirotsune S., et al. 2003. An expressed pseudogene regulates the mRNA stability of its homologous coding gene. *Nature* 423:91-96.
- Hochedlinger, K., et al., 2004. Reprogramming of a melanoma genome by nuclear transplantation. *Genes and Development* 18: 1875-1885.
- Holladay, P.M. and J.M. Watt. 2001. De-generation: an exponential decay curve in old testament genealogies. *Evangelical Theological Society Papers*, 2001. 52nd Natl. Conf., Nashville, TN Nov. 15-17, 2000.

- Howell, et al. 1996. Evolution of human mtDNA. How rapid does the human mitochondrial genome evolve? *A. J. Hum. Genet.* 59: 501-509.
- Hoyle, F. 1999. *Mathematics of Evolution*. Acorn Enterprises, LLC, Memphis, TN.
- Johnson, J.M., et al. 2005. Dark matter in the genome: evidence of widespread transcription detected by microarray tilling experiments. *Trends in Genetics* 21:93-102.
- Kapranov, P., A.T. Willingham, and T.R. Gingeras. 2007. Genome-wide transcription and the implications for genome organization. *Nature Reviews Genetics* 8:413-423.
- Karlin, S. 1998. Global dinucleotide signatures and analysis of genomic heterogeneity. *Current Opinion in Microbiology*. 1:598-610.
- Kimura, M. 1968. Evolutionary rate at the molecular level. *Nature* 217: 624-626.
- Kimura, M. and T. Ohta. 1971. *Theoretical Aspects of Population Genetics*. Princeton University Press, Princeton, NJ, pp 26-31, p 53.
- Kimura, M. 1976. How genes evolve; a population geneticist's view. *Ann. Genet.*, 19, no 3, 153-168.
- Kimura, M. 1979. Model of effective neutral mutations in which selective constraint is incorporated. *PNAS* 76:3440-3444.
- Kimura, M. 1983. *Neutral Theory of Molecular Evolution*. Cambridge Univ. Press, NY, NY. (p.26, pp 30-31).
- Kondrashov, A.S. 1995. Contamination of the genome by very slightly deleterious mutations: why have we not died 100 times over? *J. Theor. Biol.* 175:583-594.
- Kondrashov, A.S. 2002. Direct Estimate of human per nucleotide mutation rates at 20 loci causing Mendelian diseases. *Human Mutation* 21:12-27.

- Koop, B.F. and L. Hood. 1994. Striking sequence similarity over almost 100 kilobases of human and mouse T-cell receptor DNA. *Nature Genetics* 7:48-53.
- Kruuk, L.E.B. et al. 2000. Heritability of fitness in a wild mammal population. *PNAS* 97:698–703.
- Lee, J. 2003. Molecular biology: complicity of gene and pseudogene. *Nature* 423:26-28.
- Lenski, R.E. 2011. Evolution in Action: a 50,000-generation salute to Charles Darwin. *Microbe* 6(1):30-33.
- Lesecque, Y., P.D. Keightley, and A. Eyre-Walker. 2012. A resolution of the mutation load paradox in humans. *Genetics* 191:1321-1330.
- Loewe, L. 2006. Quantifying the genome decay paradox due to Muller's ratchet in human mitochondrial DNA. *Genetic Research* 87:133-159.
- Lynch, M., et al. 1995. Mutational meltdown in sexual populations. *Evolution* 49 (6):1067-1080.
- Lynch, M., J. Conery, and R. Burger. 1995. Mutation accumulation and the extinction of small populations. *Am. Nat.* 146:489-518.
- Lynch, M. 2010. Rate, molecular spectrum, and consequences of human mutation. *PNAS* 107 (3):961-968.
- Manuelidis, L. 1990. View of interphase chromosomes. *Science* 250:1533-1540.
- Marks, II R.J., M.J. Behe, W.A. Dembski, B.L. Gordon, and J.C. Sanford. 2013. *Biological Information – New Perspectives*. World Scientific Publishing Co., Singapore (pp 1-559). <http://www.worldscientific.com/worldscibooks/10.1142/8818>
- Mattick, J.S. 2001. Non-coding RNAs: the architects of eukaryotic complexity. *EMBO reports* 2:986-991.

- Merila J. and B.C. Shelton. 2000. Lifetime reproductive success and heritability in nature. *The American Naturalist* 155:301-310.
- Montañez, G., R. Marks, J. Fernandez, and J. Sanford. 2013. Multiple overlapping genetic codes profoundly reduce the probability of beneficial mutation, In: Marks II R.J. et al., (eds) *Biological Information – New Perspectives* (pp 139-167). http://www.worldscientific.com/doi/pdf/10.1142/9789814508728_0006
- Morrish, T.A., et al. 2002. DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nature Genetics*:31:159-165.
- Morton, N.E., J.F. Crow, and H.J. Muller. 1956. An estimate of the mutational damage in man from data on consanguineous marriages. *PNAS* 42:855-863.
- Muller, H.J. 1950. Our load of mutations. *Amer. J. Human Genetics* 2:111-176.
- Muller, H.J. 1964. The relation of recombination to mutational advance. *Mutation Research* 1:2-9.
- Nachman, M.W. and S.L. Crowell. 2000. Estimate of the mutation rate per nucleotide in humans. *Genetics* 156:297-304.
- Neel, J.V., et al. 1986. The rate with which spontaneous mutation alters the electrophoretic mobility of polypeptides. *PNAS* 83:389-393.
- Nelson, C.W. and J.C. Sanford. 2011. The Effects of Low-Impact Mutations in Digital Organisms. *Theoretical Biology and Medical Modeling*, Vol. 8, (April 2011), p. 9. <http://www.tbiomed.com/content/pdf/1742-4682-8-9.pdf>
- Nelson, C. and J. Sanford. 2013. Computational evolution experiments reveal a net loss of genetic information despite selection, In: Marks II R.J. et al., (eds) *Biological Information – New Perspectives* (pp 338-368). http://www.worldscientific.com/doi/pdf/10.1142/9789814508728_0014
- Ohno, S., and T. Yomo. 1991. The grammatical rule for all DNA: junk and coding sequences. *Electrophoresis* 12:103-108.

- Ohta, T. 1973. Slightly deleterious mutant substitutions in evolution. *Nature* 246:96-98.
- Ohta, T. 1974. Mutational pressure as the main cause of molecular evolution and polymorphism. *Nature* 252:351-354.
- Ohta, T. 1992. The nearly neutral theory of molecular evolution. *Ann Rev Ecol Syst* 23:263-286.
- Ohta, T. 2002. Near-neutrality in evolution of genes and gene regulation. *PNAS* 99:16134-16137.
- Paley, W. 1802. *Natural theology: evidences of the existence and attributes of the Deity, collected from the appearances of nature.*
- Parsons, T.J. et al. 1997. A high observed substitution rate in the human mitochondrial DNA control region. *Nature Genetics* 15:363-368.
- Patterson, C. 1999. *Evolution.* Comstock Publishing Associates, Ithaca, NY.
- Provine, W.B. 1971. *The Origins of Theoretical Population Genetics.* University of Chicago Press, Chicago. pp 174-177.
- Qui, S., A. McSweeney, S. Choulet, A. Saha-Mandal, L. Fedorova, and A. Fedorov. 2014. Genome evolution by matrix algorithms (GEMA): cellular automata approach by population genetics. *Genome Biol. Evol.* 6(4):988-999.
- Rands, C.M, S. Meader, C.P. Pointing, G. Lunter, 2014. 8.2% of the Human Genome is Constrained: Variation in Rates of Turnover across Functional Element Classes in the Human Lineage. *PLOS Genetics*, 24 Jul 2014 DOI: 10.1371/journal.pgen.1004525.
- ReMine, W. 1993. *The Biotic Message: Evolution versus Message Theory.* St. Paul Science, St. Paul, MN. 538 pages.
- ReMine, W. 2005. Cost of Selection Theory. *Technical Journal* 19:113-125.

- Rupe, C.L. and J.C. Sanford. 2013. Using numerical simulation to better understand fixation rates, and establishment of a new principle: Haldane's Ratchet. ICC. http://media.wix.com/ugd/a704d4_47bcf08eda0e4926a44a8ac9cbfa9c20.pdf
- Sandman, K., et al. 2000. Molecular components of the archaeal nucleosome. *Biochimie* 83: 277-281.
- Sanford, J., J. Baumgardner, P. Gibson, W. Brewer, and W. ReMine. 2007a. Mendel's Accountant: a biologically realistic forward-time population genetics program. *Scalable Computing: Practice and Experience* 8(2):147-165. (<http://www.scpe.org>)
- Sanford, J., J. Baumgardner, P. Gibson, W. Brewer, and W. ReMine. 2007b. Using computer simulation to understand mutation accumulation dynamics and genetic load. In: Shi et al. (Eds.), *International Conference on Computational Science 2007, Part II, LNCS 4488* (pp.386-392), Springer-Verlag, Berlin, Heidelberg. <http://bioinformatics.cau.edu.cn/lecture/chinaproof.pdf>
- Sanford, J.C. 2008. *Genetic Entropy – Classroom Edition.* FMS Foundation, Inc. Waterloo, NY. 227 pages.
- Sanford, J.C. et al. 2008. Using Numerical Simulation to Test the Validity of Neo-Darwinian Theory. In A. A. Snelling (Ed.) (2008). *Proceedings of the Sixth International Conference on Creationism* (pp. 165–175). Pittsburgh, PA: Creation Science Fellowship and Dallas, TX: Institute for Creation Research. <http://www.icr.org/i/pdf/technical/Using-Numerical-Simulation-to-Test-the-Validity-of-Neo-Darwinian-Theory.pdf>
- Sanford, J. and C. Nelson. 2012. The Next Step in Understanding Population Dynamics: Comprehensive Numerical Simulation, *Studies in Population Genetics*, in: M. Carmen Fusté (Ed.), ISBN: 978-953-51-0588-6, InTech, Available from: <http://www.intechopen.com/books/studies-in-population-genetics/the-next-step-in-understanding-population-dynamics-comprehensive-numerical-simulation>.

Sanford, J. 2013. Session II Chair - Biological Information and Genetic Theory: Introductory Comments, In: Marks II R.J. et al., (eds) *Biological Information – New Perspectives* (pp 203-209). http://www.worldscientific.com/doi/pdf/10.1142/9789814508728_others02

Sanford, J., J. Baumgardner, and W. Brewer. 2013. Selection Threshold Severely Constrains Capture of Beneficial Mutations, In: Marks II R.J. et al., (eds) *Biological Information – New Perspectives* (pp 264-297). http://www.worldscientific.com/doi/pdf/10.1142/9789814508728_0011

Sanford J.C. 2014. *Biological Information – New Perspectives: A Synopsis and Limited Commentary* (46 pages). http://www.amazon.com/Biological-Information-Perspectives-Synopsis-Commentary-ebook/dp/B00IKTV2C/ref=sr_1_1?s=digital-text&ie=UTF8&qid=1394477907&sr=1-1&keywords=biological+information+new+perspectives

Seaman, J. 2013. DNA.EXE: A Sequence Comparison between Human Genome and Computer Code. In: Marks II R.J. et al., (eds) *Biological Information – New Perspectives* (pp 385-401). http://www.worldscientific.com/doi/pdf/10.1142/9789814508728_0016

Segal E., et al. 2006. A genomic code for nucleosome positioning. *Nature*. 442(7104):772-778.

Shabalina, S.A., et al. 2001. Selective constraint in intergenic regions of human and mouse genomes. *Trends in Genetics* 17:373-376.

Shapiro, J.A., and R.V. Sternberg. 2005. Why repetitive DNA is essential to genome function. *Biol. Rev.* 80:1-24.

Simonsen L., et al. 1998. Pandemic versus epidemic mortality: a pattern of changing age distribution. *Journal of Infectious Diseases* 178:53-60.

Slack, F.J. 2006. Regulatory RNAs and the demise of 'junk' DNA. *Genome Biology* 7:328.

Storz, G. 2002. An expanding universe of non-coding RNAs. *Science* 296:1260-1263.

Sutherland, G.R. and R.I. Richards. 1995. Simple tandem repeats and human disease. *PNAS* 92: 3636-3641.

Tachida, H. 1990. A population genetic model of selection that maintains specific trinucleotides at a specific location. *J. Mol. Evol.* 31:10-17.

Taft, R.J. and J.S. Mattick. 2003. Increasing biological complexity is positively correlated with the relative genome-wide expansion of non-protein-coding DNA sequences. *Genome Biology* 5(1):P1.

The ENCODE Project Consortium. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447: 799-816.

The ENCODE Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489:57-74.

Tishkoff, S.A. and B.C. Verrelli. 2003. Patterns of human genetic diversity: implications for human evolutionary history and disease. *Annual Review of Genomics and Human Genetics* 4:293-340.

Trifonov, E.N. 1989. Multiple codes of nucleotide sequences. *Bull. of Mathematical Biology* 51: 417-432.

Trifonov, E.N. 1997. Genetic sequences as product of compression by inclusive superposition of many codes. *Molecular Biology* 31(4): 647-654.

Vinogradov, A.E. 2003. DNA helix: the importance of being GC-rich. *Nucleic Acid Research* 31:1838-1844.

Weikart, R. 2004. *From Darwin to Hitler: Evolutionary Ethics, Eugenics, and Racism in Germany*. Palgrave MacMillan.

Wells, J. 2013. Not junk after all: non-protein-coding DNA carries extensive biological information, In: Marks II R.J. et al., (eds) *Biological Information – New Perspectives* (pp 210-231). http://www.worldscientific.com/doi/pdf/10.1142/9789814508728_0009.

Xue, Y., et al. 2009. Human Y chromosome base-substitution mutation rate measured by direct sequencing in a deep-rooting pedigree. *Current Biology* 19:1453-1457.

Yelin, R., et al. 2003. Widespread occurrence of antisense transcription in the human genome. *Nature Biotechnology* 21:379-386.