DOI: 10.13376/j.cbls/2021165
文章编号：1004-0374(2021)12-1469-07

戴俊彪博士，中国科学院深圳先进技术研究院研究员。1997 年于南京大学获得学士学位，2000 年于清华大学获得硕士学位，2006 年于美国爱荷华州立大学 (Iowa State University) 获得博士学位。之后在美国约翰霍普金斯大学医学院从事博士后研究，2011 年回国入职清华大学生命科学学院。2017 年加入中国科学院深圳先进技术研究院。戴俊彪实验室主要从事表观遗传学与合成生物学研究，开发基因和基因组的合成、组装及转移技术，通过基因组的设计构建解析基因组功能，并进行合成生物的改造和优化利用等，是国际合成酵母基因组计划的主要成员。已在 *Cell*、*Nature*、*Science* 等刊物上发表学术论文四十余篇。2017 年 3 月与 Sc2.0 合作团队在 *Science* 杂志上以封面和专刊的形式发表了五篇染色体合成相关文章，入选 2017 年中国科学十大进展、中国高等学校十大科技进展、中国科技进展十大新闻。荣获国家杰出青年基金、谈家桢生命科学创新奖、英国皇家学会牛顿高级学者等。

DNA信息存储的机遇与挑战

强　薇[1]，沈　玥[2]，戴俊彪[1]*
(1 中国科学院深圳先进技术研究院，深圳合成生物学创新研究院，广东省合成基因组学重点实验室，深圳市合成基因组学重点实验室，深圳 518055；2 深圳华大生命科学研究院，广东省高通量基因组测序与合成编辑应用实验室，深圳 518083)

摘　要：在高度信息化的现代社会中，现有的硅基存储资源将难以满足爆炸式增长的信息总量。基于 DNA 介质的信息存储融合了多个学科领域，提供了一种新的存储模式，并在世界范围内取得了相当的进展。该文首先对目前 DNA 信息存储所取得的进展进行梳理归纳，继而总结了 DNA 信息领域现阶段所面临的关键问题，最后对 DNA 信息存储领域未来进行展望，以期为下一步研究提供方向。

关键词：DNA 存储；合成生物学；信息科学；生物信息学；高通量 DNA 合成
中图分类号：Q811.4　　**文献标志码：**A

The opportunities and challenges in DNA digital data storage

QIANG Wei[1], SHEN Yue[2], DAI Jun-Biao[1]*
(1 Guangdong Provincial Key Laboratory of Synthetic Genomics and Shenzhen Key Laboratory of Synthetic Genomics, Institute of Synthetic Biology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China; 2 Guangdong Provincial Key Laboratory of Genome Read and Write, BGI-Shenzhen, Shenzhen, 518083, China)

Abstract: In modern society, the amount of information is exploding. However, current silicon-based storage is limited by the availability of the materials and quickly behinds the needs. Storing information in DNA is an interdisciplinary filed, which has the potential to provide a solution for the increased information. The filed has made many achievements in the past decade. Here, we start the review by briefly introducing the progress of the

收稿日期：2021-10-27
基金项目：广东省合成基因组学重点实验室(2019B030301006)
***通信作者：**E-mail: junbiao.dai@siat.ac.cn

field and highlighting a few recent achievements. Next, we focus our attention on bottlenecks which are currently limiting the application of DNA data storage and discuss potential solutions. Finally, future directions are proposed.
Key words: DNA data storage; synthetic biology; information science; bioinformatics; high-throughput DNA synthesis

互联网和人工智能等信息技术的快速发展导致现代社会所产生的信息量呈指数型增长。据 IBM 统计，人类社会每天创造 2.5 EB (1 EB 相当于 10^{18} bytes)，大约相当于 5 亿部高清电影 [1]。互联网数据中心 IDC 预测，到 2025 年全球将会有 163 ZB (1 ZB 相当于 10^{21} bytes) 的数据产生。当人们享受着数据与人工智能带来的便利时，一个基础问题日益凸显，那就是如何实现这些信息的保存。目前信息数据主要存储于以硅为介质的硬盘中，然而全球硅的储备总量有限 , 据推测 2040 年该资源将被完全耗尽 [2]，从而导致以硬盘为主的存储设备将面临严峻挑战。因此，寻找具有低成本、高稳定性、高容量、高密度等潜力，并且能够承受极端环境条件的新型数据存储介质，是高度信息化的社会亟需解决的关键问题之一。

DNA 是生物体内的遗传物质，具有极高的存储密度，是承载各种生命遗传信息的天然载体。DNA 数据存储技术，是通过编码算法将目前计算机中的 0、1 二进制数据，转换为 A、T、C、G 四种碱基组成的 DNA 序列，进而通过合成含有指定碱基序列的 DNA，实现数据信息的存储。据计算，1 克 DNA 可存储 1 000 亿张 DVD 光盘，存储近 10 亿 TB 的数据。在存储密度方面，理论上 DNA 存储可达到 455 EB/ 克 (4.55×10^{11} GB/ 克)，大约 10^{18} bytes 或 107 GB 每 mm^3，比传统存储介质提高了 5~6 个数量级 [3]。通过生物技术与信息基础交叉融合研究，利用 DNA 实现二进制信息的存储已取得了突破性进展。

1 发展历史

相比于光盘、硬盘和磁盘等传统存储介质，DNA 存储密度高、保存时间长，在信息存储尤其是大规模档案存储方面有望替代传统存储介质。早在 1959 年，分子尺度计算机的概念就已出现 [4]。1994 年，DNA 作为“数据”载体存储信息 [5]，打开了生物计算的大门。1995 年，Baum[6] 提出构建 DNA 大容量存储系统。2001 年，Reif 等 [7] 建立了支持随机访问的 DNA 数据库。然而直到 2012 年，哈佛大学的 Church 团队 [8] 将一本 650 KB 图书存储在 DNA 后，该领域才逐渐吸引了众多研究者的注意。2013 年，欧洲生物信息学研究所的 Goldman 及其同事 [9] 在 DNA 中存储了 739 KB 的五种文件 (文本、PDF、照片、MP3 和霍夫曼编码)。这两项研究实现了数据可行、高容量的 DNA 数据存储，将 DNA 数据存储领域向前推进了一大步，使得 DNA 存储研究的热潮持续升温。2017 年，哥伦比亚大学和纽约基因组中心的 Yaniv Erlich 和其同事 [10] 开发了一种高效可靠的 DNA 存储策略——“DNA 喷泉 (DNA fountain)”，大大提升了 DNA 数据存储能力；同年，哈佛大学的 Church 团队 [11] 利用 CRISPR-Cas 系统将一张黑白图像和一部短的视频文件“写入”大肠杆菌的基因组中。2018 年，华盛顿大学联合微软研究院将超过 200 MB 的数据“写入”DNA 中，并支持 DNA 数字存储的随机访问 [12]。2019 年，Erlich 团队将通过喷泉码编码的信息包裹入 3D 打印材料中，打印出了一只存有遗传信息的斯坦福兔子 [13]；同年，微软与华盛顿大学结合第三代测序技术，开发出了自动化 DNA 存储的系统 [14]。2021 年，麻省理工学院的 Mark Bathe 团队开发了应用于 DNA 存储数据随机读取的快速检索系统 [15]。近六年来，DNA 存储容量大大增加，并呈现指数级的增长趋势，与最初相比提高了将近 3 个数量级。

近年来，我国在 DNA 存储研究领域也取得了较快的发展。国内第一个 DNA 存储领域的专利申请于 2015 年，由苏州泓迅生物科技股份有限公司提交 [16]。到了 2016 年，清华大学的戴俊彪课题组建立了生物体存储的一种“数据 -DNA”编码方法 [17]。2018 年，深圳华大生命科学研究院开发了“阴阳”双编码方法，能够控制输出 DNA 的 GC 含量、最长单碱基重复长度以及二级结构自由能，使其都在指定范围内 [18]。2020 年，天津大学齐浩课题组构建了携带不同短链信息片段质粒的大肠杆菌分布式混菌存储系统，将 445 KB 的数字文件储存在 2 304 Kbps 的合成 DNA 中，实现了目前在体内的最大规模信息存储 [19]；同年，该课题组将携带数据信息的 DNA 原始文库固定在磁珠上，通过使用等温的链置换扩增技术，对大型 DNA 文库进行低偏好性、稳定重复的扩增，实现了数据的稳定可重复

性读取[20]。2021年，深圳华大生命科学研究院开发了一个集成了多种编码方法的评估平台Chamaeleo，通过该平台，可使用编码测试文件来评估编解码方法的特性[21]；深圳先进技术研究院戴俊彪/王洋课题组提出了适用于DNA存储的自包含自解释系统，可以在DNA存储数据恢复的过程中，摆脱外部工具的依赖[22]；清华大学朱听课题组利用手性DNA抗降解的特性，开发了一套高稳定性的DNA存储流程[23]；天津大学元英进课题组通过体内组装，对编码的254 886 bp的人工染色体进行一次写入，实现了稳定的复制和多次检索[24]；陈为刚等[25]提出了一种DNA数据混合错误纠正与数据恢复的方法；中国科学院北京基因组研究所(国家生物信息中心)提出了DNA活字存储系统和方法，构建出内容活字实物库以及索引活字实物库，且能够一次合成，多次使用，大大降低成本[26]；清华大学的刘凯等[27]开发了基于CRISPR-Cas12a-λRed体系的随机重写DNA信息的存储方法。DNA信息存储发展历程总结见图1。

2　现有状况和水平

DNA信息存储主要包含以下四大部分：

2.1　DNA编解码算法

计算机系统的信息存储，是以硅基介质为基础，将各类信息编码为0、1二进制数据存入其中。而对于DNA存储而言，需要将信息编码为A、T、C、G组成的碱基序列。目前，DNA编码算法大多采用将待存信息先转换为0、1二进制数据，再进一步编码为碱基序列(图2)。在DNA编码中，最重要的两点是编码后碱基序列的GC含量以及碱基连续重复情况。

Church等[8]按照A或T编码0，C或G编码1的规则，对二进制信息进行编码，得到碱基序列。这种方法具有较高的灵活性，能够很好地避免碱基连续重复以及矫正GC偏好性。但是，这样的编码规则并不能使4碱基体系包含最大信息量，编码效率与编码密度不高。Goldman等[9]利用霍夫曼编码，对二进制数据进行转换，之后再将转换后的序列编码为DNA序列，将编码密度提升到了1.58 bits/nt；然而，霍夫曼编码实际上为压缩编码，压缩效率受到数据的具体组成影响，所以对于某些数据来说，压缩效果并不好，这样就会导致最终的编码密度不太理想。在Erlich和Zielinski[10]的DNA喷泉码算法中，00、01、10、11分别编码A、C、G、T，并结

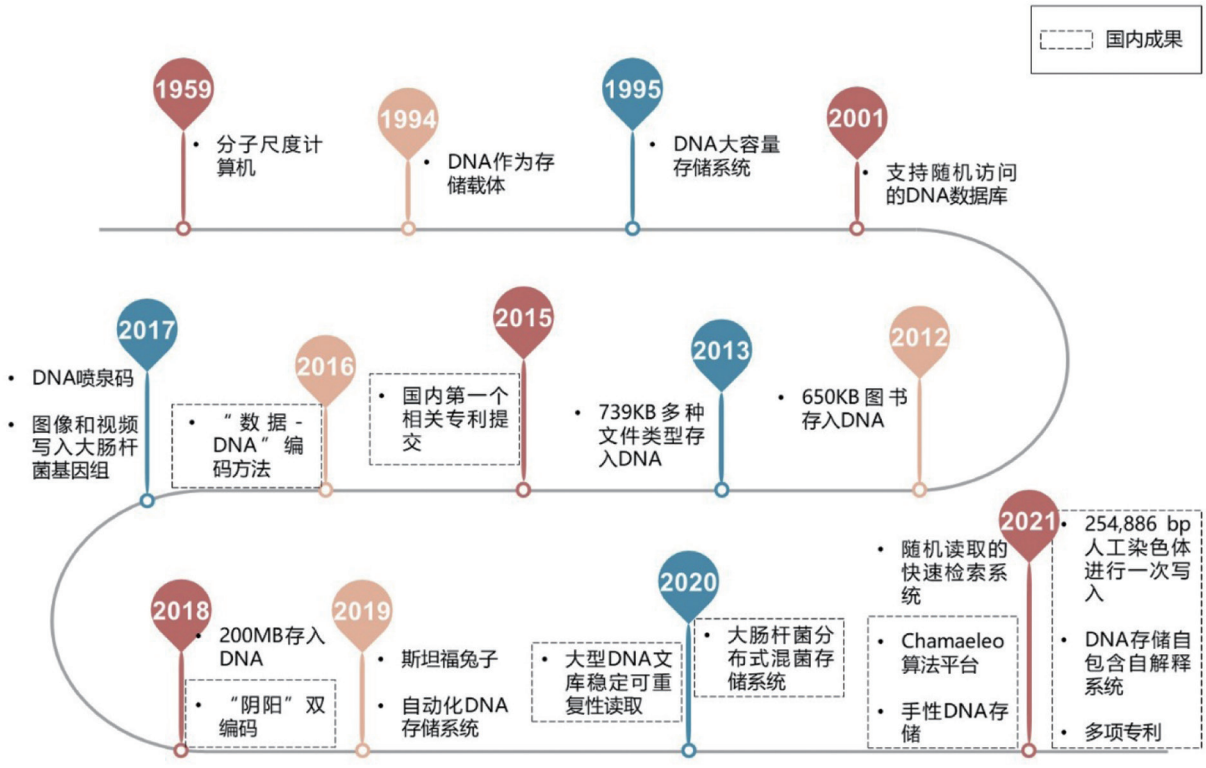

图1　DNA信息存储发展历程

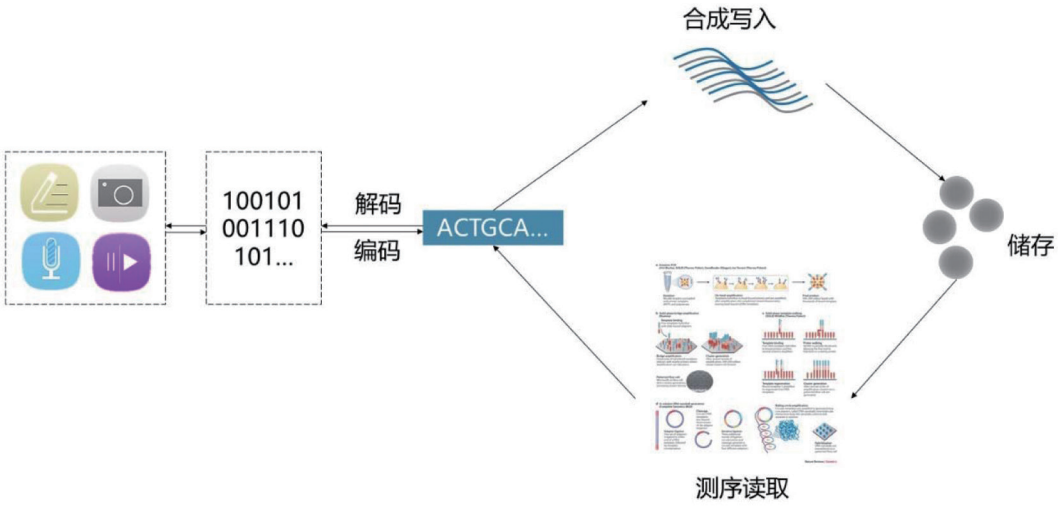

图2 DNA存储流程示意图[28]

合通讯领域的喷泉码[29]，有效地将编码密度提升到了 1.98 bits/nt；但是在实际应用中发现，DNA 喷泉码的数据恢复能力相对较差，当数据有部分缺失时，存在一定概率无法解码。

2.2 DNA高通量合成

目前，已商业化应用的有基于柱式合成和基于微阵列芯片原位合成两种合成方法。基于合成柱的 DNA 合成通量低，DNA 的合成成本在 0.05~0.17 元 / 碱基。DNA 微阵列原位化学合成方法包括原位光刻法、光敏抗蚀层合成法、光致酸法、喷印合成法、软光刻合成法、电致酸法和压印法及基于分选原理合成法等多种方法[30-31]。到 2014 年，基于微阵列芯片原位合成方法的 DNA 合成成本已低于 10^{-4} 元 / 碱基，2020 年基于电致酸法合成的 DNA 成本估计达到 10^{-6} 元 / 碱基[32]。而 2018 年，美国半导体研究会预测，到 2023 年 DNA 合成成本将降低到 10^{-10} 元 / 碱基以下[33]。近些年来，酶促 DNA 合成方法也取得了进一步的进展。2019 年，Lee 等[34] 采用非阻断型的末端脱氧核酸转移酶 (TdT) 开发了一种专用于信息存储的 DNA 酶法从头合成技术。2020 年，Lee 等[35] 进一步利用图案化紫外光快速解离 Co^{2+} 离子激活 TdT，空间选择性合成 DNA，成功将 110 bits 的音乐数据信息编码入 DNA 中，初步验证了在阵列表面实现大规模并行合成的可行性。

2.3 DNA分子保存

DNA 极易受到温度、水分、紫外线、氧气和 pH 的影响，从而被烷基化、水解和氧化，造成 DNA 序列的损坏，导致存储信息的丢失。因此，DNA 的稳定保存是 DNA 信息存储至关重要的一个方面。目前，现有的 DNA 保存方法主要包括液体法、干燥法、加入稳定剂法、封装法和体内存储法。液体法是实验室中最常用的 DNA 保存方法，此方法的优点是 DNA 的取用非常便捷，但是保存在此条件下 DNA 容易被水解，且能耗较大。固体 DNA 相比于液体 DNA 而言更加稳定，而干燥方法主要包括喷雾干燥、空气干燥和冷冻干燥。一般来说，冷冻干燥有效避免了干燥过程出现高温，从而减少了 DNA 的损坏。目前所用到的 DNA 稳定剂有海藻糖、聚乙烯醇 (poly vinyl alcohol, PVA) 和 DNAStable[36-39]。其中，海藻糖是最常用的 DNA 稳定剂，研究表明，可能是海藻糖与磷酸盐结合中和了 DNA 的负电荷，或者海藻糖与 DNA 之间的氢键建立的网络减少了 DNA 结构的波动从而实现了对 DNA 的保护。封装法，即使用二氧化硅或碱金属盐对 DNA 进行包裹，使其不与外界接触，避免环境因素导致 DNA 的损坏。2019 年，Chen 等[40] 以 Fe 为中心，通过 DNA 与聚乙烯亚胺的交替覆盖，将 DNA 载量提升至 7.8 wt%；最外层以二氧化硅保护，使其在室温下可保存 20~90 年。2020 年，Kohll 等[41] 通过磷酸钙、氯化钙和氯化镁等碱金属盐包埋并保护 DNA，其载量可提升至 30 wt%。在体内存储方面，2020 年天津大学齐浩课题组构建了携带不同短链信息片段质粒的大肠杆菌分布式混菌存储系统，将 445 KB 的数字文件储存在 2 304 Kbps 的合成 DNA 中，实现了目前在体内存储的最大规模的信息存储[19]。

2.4 测序信息读取

目前，DNA 存储的主流方式是短片段信息存储 (oligo pool)，读取方式主要是利用高通量 DNA

测序技术 (二代测序技术)。二代测序技术的核心思想是大规模平行测序，一次上样可并行几十万到几百万条 DNA 分子的序列测定，是目前最成熟、应用最广泛的测序技术，然而仪器昂贵，操作复杂。近年来逐渐发展的三代测序技术不依赖 PCR 扩增，并具备读长更长、读取速率更快的显著优势。牛津纳米孔公司开发的 DNA 平均过孔速率为 450 bp/s 的三代测序系列产品，具有袖珍便携的优点。三代测序 MinION 有多达 512 个纳米孔通道能够进行同时测序，而桌面级高通量台式产品 PromethION 48 的数据通量为 7.6 TB (72 h) 量级，数据读取速率相当于 29 MB/s。随着技术更迭和算法升级，三代测序有望用于体内或体外稳定化的长片段 DNA 存储信息的读取，并与当前传统介质的读取速度 (KB~GB/s) 比肩。高通量测序技术的发展使得目前的测序成本相比于最初的 Sanger 测序成本降低了 6 个数量级 [42]。

3　问题与挑战

虽然 DNA 信息存储技术取得了一定的突破，但是其在应用方面仍面临很多挑战。

3.1　“编”

目前，在算法设计上已经能够很好地控制编码后碱基序列的 GC 含量以及碱基连续重复，然而仍然存在较多问题尚未解决。首先，目前所开发的大多数编码算法，需要将待存信息转换为 0、1 二进制数据之后，再进行碱基序列的编码，这样的流程仍然依赖于计算机体系，未能脱离硅基存储；其次，未对串联重复结构进行控制，可能造成在 PCR 扩增的过程中，形成复杂结构，影响序列的扩增效率，从而对解码造成困难。最后，编码过程中需引入索引、纠错码等非信息序列，造成了存储信息密度的下降。

3.2　“写”

化学合成法对于长链 DNA 片段的合成，仍然存在着技术瓶颈，目前只能合成 200 nt 以内的 DNA 序列。“写”DNA 的价格依旧十分昂贵。假设每个碱基存储 1 比特的信息，而使用阵列（高通量）合成 DNA 的成本约为每碱基 0.0001 美元，存储 1TB 的信息至少需要 8 亿美元。相比之下，使用磁带存储同等数据规模的成本仅为 16 美元 [43]。显然，合成 DNA 的高昂成本成为了 DNA 数据存储进入实用阶段的一大短板，削弱了 DNA 分子相比于传统存储介质的优势。以酶合成为基础的第三代 DNA 合成技术目前还处于发展初期，因此目前报道的如基于 TdT 等聚合酶的合成方法还不能进行高通量平行 DNA 合成，其低通量方法也尚未进入实际应用。

3.3　“存”

现有的 DNA 保存技术保护的 DNA 含量少，且能耗大，无法满足存储档案所要求的数据量大、保存时间较长的需求，难以实现 DNA 数据的大规模长期保存。目前的保存方法缺少有效的物理隔离手段，在信息读取过程中，难以准确快速地获得目标信息，数据访问的成本和时间较长，限制了信息的获取速度。据微软报道，目前一个 PCR 反应体系可以操纵 10^6 种不同的 DNA 序列，但是尚未可知一个体系中可以并行操纵的 DNA 序列种数的极限。

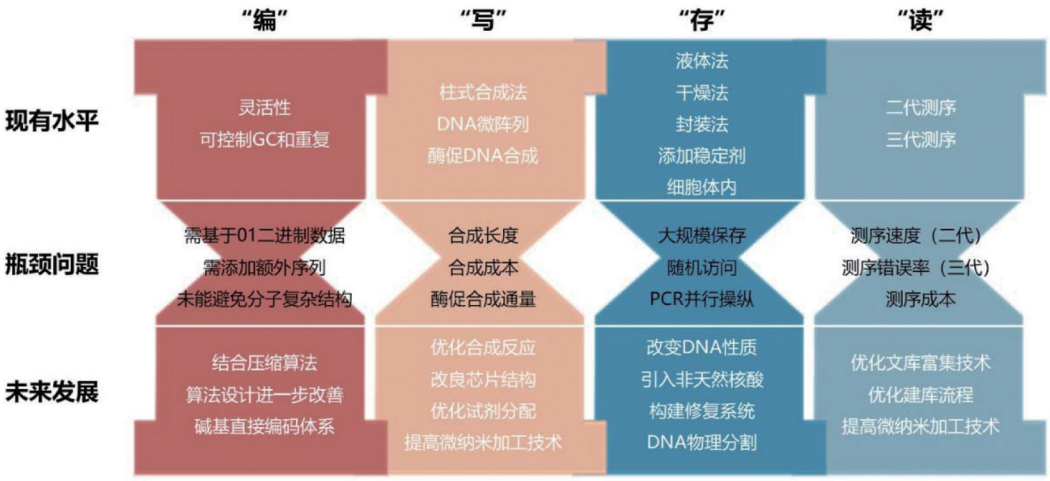

图3　DNA信息存储的现有水平、瓶颈问题及未来发展方向

3.4 “读”

二代测序的核心思想是大规模平行测序，一次上样可并行几十万到几百万条 DNA 分子的序列测定。但二代测序包含建库、读取等流程，一轮需数天时间，无法实现数据的实时读取，限制了全自动化 DNA 存储和读取设备的集成。而三代测序错误率较高，所以需要更多的测序资源以获得正确的数据信息，造成成本的增加。

4 未来与展望

4.1 高效率高质量直接“编”码

受限于目前的合成测序技术以及成本考虑，现有的编码算法均采用将信息分割为 200 nt 以内的序列的方案，以及潜在的碱基错误风险，索引和纠错码序列是必不可少的，由此而导致的存储密度降低，可考虑研发相应的序列压缩算法，从而提升信息存储密度；对于编码后序列的串联重复结构可能带来的影响，可以在今后的算法设计中，将这一因素考虑进去，尽可能地消除由于序列结构本身所带来的问题；同时，也可考虑研发不经过二进制转换的新编码算法，将待存的图片、文本或视频信息直接编码为碱基序列，真正做到脱离硅基存储的依赖。

4.2 低成本高通量信息“写”入

值得注意的是，DNA 信息存储在未来仍有巨大的成本下降潜力。首先，可以从优化合成反应、改良芯片结构、优化试剂分配量、降低单体成本多方面着手，大幅降低合成成本。其次，可通过 DNA 序列人为设计，系统创新，合成步骤容忍更多错误，降低精度与纯度要求，致力于提升总合成量 (通量 × 长度)，大幅降低合成成本。同时，随着微纳加工半导体器件在 DNA 信息存储领域的应用，DNA 合成技术与装备有望快速迭代升级，合成通量快速提升，成本有望实现快速下降。

另外，酶促 DNA 合成虽然还处于发展初期，仍有望大大减少 DNA 合成的时间和成本。2019 年，Lee 等 [34] 使用 TdT 合成 DNA 的时间估计为每周期 40 秒，是化学合成法速度的 6 倍，并且酶促合成法每周期的成本降低为化学合成法使用的亚磷酰胺的每周期的成本的 1/1000。一旦酶反应系统被微型化，预计成本将再下降几个数量级。

4.3 稳定高兼容性分子信息“存”储

DNA 序列的长期稳定保存，对 DNA 信息存储至关重要。首先，可通过改变 DNA 本身的性质，来增强 DNA 存储的物理稳定性。例如，“锁定的”核酸单体在戊糖环的 2-O 和 4-C 之间具有亚甲基桥，可抵抗核酸酶的酶解。环状 -DNA 或甘油 -DNA 可改变糖骨架的化学性质或抵抗核酸酶的酶解，从而提高 DNA 的稳定性。另外，使用非天然核酸、添加 DNA 保护剂以及构建类似于天然生物系统的主动修复系统，也可用于提高存储系统的稳定性。

数据的随机访问也是 DNA 信息存储的一个重要发展方向。未来，可通过将 DNA 文库的物理隔离、数字微流控技术以及索引方法的建立等技术相结合，实现数据的便捷随机读取。

4.4 实时永久性信息稳定“读”取

对于二代测序来讲，因为现有的测序反应要求 DNA 量较多，通常测序使用的模板都是通过对原始样品的核酸序列进行扩增后的产物。而在使用 PCR 技术对大规模的 DNA 文库进行富集时，因 DNA 序列的 GC 含量和二级结构均会影响 PCR 的扩增效率，因此，该过程将不可避免地产生偏好性。所以，优化测序过程中 DNA 文库的富集反应 (优化现有的生化方法或开发新的扩增技术) 是非常必要的。而对于三代测序，进一步开发新的传感设备以提高对信号收集的精准度，提高测序的正确率。未来的测序技术的发展主要依赖于微纳加工技术来实现测序微环境的结构形成，依靠物理学的手段进行识别，从而提高测序的通量。

综上，基于 DNA 介质的新型数据存储作为一种具有划时代意义的存储方式，目前的研究已经取得了一定的成果，但同时也面临着巨大的挑战。随着 DNA 信息存储各个问题的逐步解决，或将打开全球海量数据存储的新纪元，以 BT (生物技术) 协助解决 IT (信息技术) 面临的海量数据存储挑战。

[参 考 文 献]

[1] Rizzatti L, Consultant V. Digital data storage is undergoing mind-boggling growth [EB/OL]. https://www.eetimes.com/digital-data-storage-is-undergoing-mind-boggling-growth/

[2] 戴俊彪. 利用DNA存储还原数据信息的方法[EB/OL]. https://ott.tsinghua.edu.cn/info/1010/1416.htm

[3] Ceze L, Nivala J, Strauss K. Molecular digital data storage using DNA. Nat Rev Genet, 2019, 20: 456-66

[4] Feynman RP. There's plenty of room at the bottom [data storage]. J Microelectromechanical Systems, 1992, 1: 60-6

[5] Adleman LM. Molecular computation of solutions to combinatorial problems. Science, 1994, 266: 1021-4

[6] Baum EB. Building an associative memory vastly larger than the brain. Science, 1995, 268: 583-5

[7] Reif JH, LaBean TH, Pirrung M, et al. Experimental

construction of very large scale DNA databases with associative search capability [M]//Jonoska N, Seeman NC. DNA Computing. Berlin, Heidelberg: Springer, 2002
[8] Church GM, Gao Y, Kosuri S. Next-generation digital information storage in DNA. Science, 2012, 337: 1628
[9] Goldman N, Bertone P, Chen S, et al. Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. Nature, 2013, 494: 77-80
[10] Erlich Y, Zielinski D. DNA fountain enables a robust and efficient storage architecture. Science, 2017, 355: 950-4
[11] Shipman SL, Nivala J, Macklis JD, et al. CRISPR-Cas encoding of a digital movie into the genomes of a population of living bacteria. Nature, 2017, 547: 345-9
[12] Organick L, Ang SD, Chen YJ, et al. Random access in large-scale DNA data storage. Nat Biotechnol, 2018, 36: 242-8
[13] Koch J, Gantenbein S, Masania K, et al. A DNA-of-things storage architecture to create materials with embedded memory. Nat Biotechnol, 2020, 38: 39-43
[14] Takahashi CN, Nguyen BH, Strauss K, et al. Demonstration of end-to-end automation of DNA data storage. Sci Rep, 2019, 9: 4998
[15] Banal JL, Shepherd TR, Berleant J, et al. Random access DNA memory using Boolean search in an archival file storage system. Nat Mater, 2021, 20: 1272-80
[16] 杨平, 孙德斌, 柳伟强, 等. 带有编码信息的人工合成DNA存储介质及信息的存储读取方法和应用[P]. 中国: CN104850760A, 2015
[17] 戴俊彪, 吴庆余, 乃哥麦提·伊加提, 等. 将数据进行生物存储并还原的方法[P]. 中国: CN107798219A, 2018
[18] Ping Z, Chen S, Zhou G, et al. Towards practical and robust DNA-based data archiving by codec system named 'Yin-Yang'. bioRxiv 829721. doi: https://doi.org/10.1101/829721
[19] Hao M, Qiao H, Gao Y, et al. A mixed culture of bacterial cells enables an economic DNA storage on a large scale. Commun Biol, 2020, 3: 416
[20] Gao Y, Chen X, Qiao H, et al. Low-bias manipulation of DNA oligo pool for robust data storage. ACS Synth Biol, 2020, 9: 3344-52
[21] 平质, 张颢龄, 陈世宏, 等. Chamaeleo: DNA 存储碱基编解码算法的可拓展集成与系统评估平台. 合成生物学, 2021, 2: 412
[22] Li M, Wu J, Dai J, et al. A self-contained and self-explanatory DNA storage system. Sci Rep, 2021, 11: 18063
[23] Fan CY, Deng Q, Zhu TF. Bioorthogonal information storage in L-DNA with a high-fidelity mirror-image Pfu DNA polymerase. Nat Biotechnol, 2021, 39: 1548-55
[24] Chen W, Han M, Zhou J, et al. An artificial chromosome for data storage. Natl Sci Rev, 2021, 8: nwab028
[25] 陈为刚, 黄刚, 韩昌彩, 等. 一种DNA数据存储混合错误纠正与数据恢复方法[P]. 中国, CN110442472B. 2021
[26] 陈非, 卜东波, 马灌楠, 等. DNA活字存储系统和方法[P]. 中国: CN111858510B, 2021
[27] 刘凯, 刘杨奕, 张洪杰, 一种可随机重写的DNA信息存储方法[P]. 中国: CN113462710A, 2021
[28] Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet, 2016, 17: 333-51
[29] Byers JW, Luby M, Mitzenmacher M, et al. A digital fountain approach to reliable distribution of bulk data. ACM Sigcomm Computer Communi Rev, 1998, 28: 56-67
[30] 江湘儿, 王勇, 沈玥. DNA 合成技术与仪器研发进展概述. 集成技术, 2021, 10: 80-95
[31] 刘全俊, 陆祖宏, 肖鹏峰, 等. DNA微阵列原位化学合成. 合成生物学, 2021, 2: 354-70
[32] Kosuri S, Church GM. Large-scale *de novo* DNA synthesis: technologies and applications. Nat Methods, 2014, 11: 499-507
[33] Zhirnov VV, Rasic D. 2018 Semiconductor Synthetic Biology Roadmap [EB/OL]. https://www.researchgate.net/publication/328812596_2018_Semiconductor_Synthetic_Biology_Roadmap
[34] Lee HH, Kalhor R, Goela N, et al. Terminator-free template-independent enzymatic DNA synthesis for digital information storage. Nat Commun, 2019, 10: 2383
[35] Lee H, Wiegand DJ, Griswold K, et al. Photon-directed multiplexed enzymatic DNA synthesis for molecular digital data storage. Nat Commun, 2020, 11: 5246
[36] Alkhamis KA. Influence of solid-state acidity on the decomposition of sucrose in amorphous systems. I. Int J Pharm, 2008, 362: 74-80
[37] Zhu B, Furuki T, Okuda T, et al. Natural DNA mixed with trehalose persists in B-form double-stranding even in the dry state. J Phys Chem B, 2007, 111: 5542-4
[38] Simbolo M, Gottardi M, Corbo V, et al. DNA qualification workflow for next generation sequencing of histopathological samples. PLoS One, 2013, 8: e62692
[39] Howlett SE, Castillo HS, Gioeni LJ, et al. Evaluation of DNAstable for DNA storage at ambient temperature. Forensic Sci Int Genet, 2014, 8: 170-8
[40] Chen WD, Kohll AX, Nguyen BH, et al. Combining data longevity with high storage capacity—layer-by-layer DNA encapsulated in magnetic nanoparticles. Adv Funct Materials, 2019, 29: 1901672
[41] Kohll AX, Antkowiak PL, Chen WD, et al. Stabilizing synthetic DNA for long-term data storage with earth alkaline salts. Chem Commun (Camb), 2020, 56: 3613-6
[42] DNA sequencing costs: data from the NHGRI Genome Sequencing Program (GSP) [EB/OL]. https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data
[43] Bioglio V, Grangetto M, Gaeta R, et al. On the fly gaussian elimination for LT codes. IEEE Communi Lett, 2009, 13: 953-5