

特约评述

DOI: 10.12211/2096-8280.2024-066

DNA 存储的关键技术：编码、纠错、随机访问与安全性

徐怀胜¹, 石晓龙², 刘晓光³, 徐苗苗⁴

(¹ 广州商学院现代信息产业学院, 广东 广州 511363; ² 广州大学计算科技研究院, 广东 广州 510006; ³ 广州体育学院运动健康学院, 广东 广州 510620; ⁴ 广州中医药大学体育健康学院, 广东 广州 510006)

摘要：DNA 信息存储是一种利用 DNA 分子作为数据载体的新型存储技术，通过合成特定序列的 DNA 来编码信息，并通过测序技术实现数据的读出。相比于传统的磁性、光学和电子存储介质，DNA 存储在存储密度、数据保存时间和能源效率等方面具有显著优势，且不易受电磁干扰的影响。随着全球数据总量的猛增，DNA 存储以其高效的存储能力、潜在的低维护成本和易于合成的化学特性，逐渐成为研究热点。本文首先介绍了 DNA 存储的基本流程，然后综述了 DNA 信息存储涉及的关键技术，尤其是编码策略、纠错技术、随机访问及 DNA 信息加密的研究进展。探讨了当前 DNA 存储技术的发展现状和主要挑战，如高成本、写入和读取速度慢等问题，并提出了可能的技术改进方向。并展望了 DNA 存储未来的发展前景，强调其在大数据时代的潜在应用和革命性影响，指出了实现商业化应用所需解决的关键技术瓶颈。

关键词：DNA 信息存储；DNA 合成；信息编码；DNA 纳米技术；合成生物学

中图分类号：Q819 文献标志码：A

Key technologies for DNA storage: encoding, error correction, random access, and security

XU Huaisheng¹, SHI Xiaolong², LIU Xiaoguang³, XU Miaomiao⁴

(¹ School of Modern Information Industry, Guangzhou College of Commerce, Guangzhou 511363, Guangdong, China; ² Institute of Computing Science and Technology, Guangzhou University, Guangzhou 510006, Guangdong, China; ³ College of Sports and Health, Guangzhou Sport University, Guangzhou 510620, Guangdong, China; ⁴ School of Physical Education and Health, Guangzhou University of Chinese Medicine, Guangzhou 510006, Guangdong, China)

Abstract: DNA information storage is a new technology that uses DNA molecules as data carriers. It encodes information for synthesizing DNA with a specific sequence and reads out data through sequencing technology. Compared with traditional magnetic, optical, and electronic storage media, DNA storage has significant advantages in data density, retention duration, energy efficiency, and security, since it is not easily affected by electromagnetic interference. With the rapid increase in the total amount of global data, DNA storage has gradually become a research

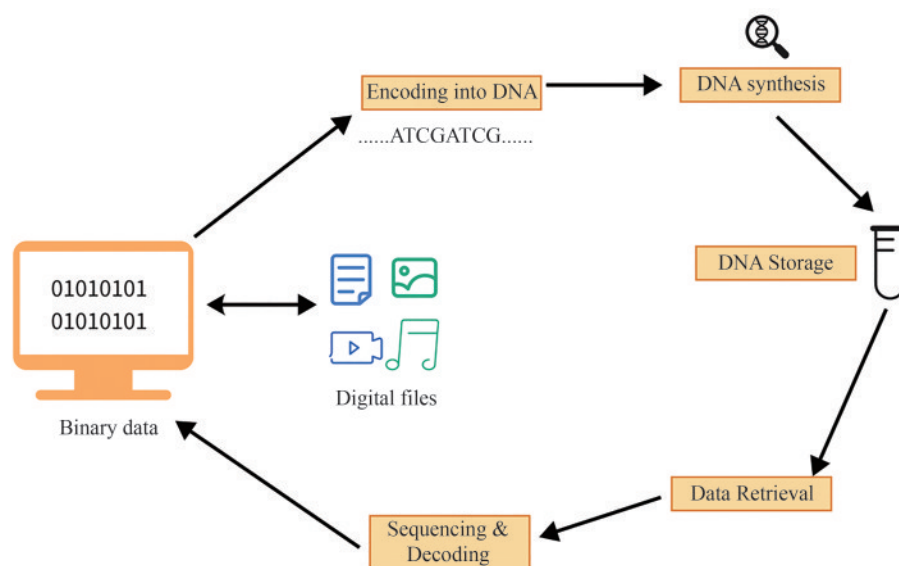
收稿日期：2024-08-26 修回日期：2024-10-15

基金项目：国家自然科学基金 (32300964)

引用本文：徐怀胜, 石晓龙, 刘晓光, 徐苗苗. DNA 存储的关键技术: 编码、纠错、随机访问与安全性[J]. 合成生物学, 2025, 6(1): 157-176

Citation: XU Huaisheng, SHI Xiaolong, LIU Xiaoguang, XU Miaomiao. Key technologies for DNA storage: encoding, error correction, random access, and security[J]. Synthetic Biology Journal, 2025, 6(1): 157-176

hotspot with its efficient storage capacity, low maintenance cost, and unique chemical property for synthesizing easily. However, DNA storage technology is still in its early stages of development and there are still many technical bottlenecks to be addressed. For example, an important advantage of DNA storage is its ultra-high storage density and long-term stability. However, achieving these goals require overcoming many technical challenges, such as reducing the error rate for synthesis and improving the encoding efficiency. Understanding existing key technologies, such as DNA encoding, error correction, random access, and DNA information encryption, can help identify and address those shortcomings, thereby promoting further technological innovation and development in DNA storage. Encoding strategy is one of the core aspects of DNA storage technology, directly determining data storage efficiency, reading accuracy, and error correction capability. To achieve efficient and stable DNA information storage, it is essential to develop more advanced encoding algorithms to enhance storage density, reduce synthesis and sequencing error rates, and ensure data accuracy and integrity. Moreover, the information security of DNA storage is becoming increasingly important, particularly in terms of data and privacy protection. As a potential data carrier, DNA storage needs to address challenges related to data encryption, information security, and tamper-proof to ensure data confidentiality and integrity. Therefore, integrating modern cryptographic techniques with DNA storage to establish a secure and reliable information storage system has become a key research focus in this field. This article first introduces the basic process of DNA storage, and then reviews the key technologies involved in DNA information storage, especially the research progress of encoding strategies, error correction technology, random access and DNA information encryption. In addition, the current development status and main challenges of DNA storage technology are also discussed. For example, the scale of DNA data storage in the laboratory is small, and the operation time for synthesis is long. Moreover, most DNA storage steps rely on experimenters, making it difficult to automate the information storage and reading process. With the advancement of synthetic biology and encoding and decoding methods, we believe that these bottlenecks will be solved in the near future, and promote the transformation of technology from laboratory research to practical applications.



Keywords: DNA information storage; DNA synthesis; information encoding; DNA nanotechnology; synthetic biology

信息技术的飞速发展和社交网络的广泛应用导致数据生成量呈指数级增长,超出了现有存储介质(如磁盘和磁带)的容量。此外,信息存储技术的快速发展意味着,由于硬件和软件的过时以及物理退化,当前存储在磁性或光学介质上的数据可能在一百年甚至更短时间内变得无法恢复^[1],如果要将有价值的信息保存给子孙后代,就需要定期将信息数据转移到新的存储介质中。因此,需要探索新的存储方法,以确保数据在更长时间内仍然可以被检索^[2]。生物存储是一个有前途的新兴技术,DNA作为新型信息存储介质,与传统的存储载体相比有很多优势:第一,保存时间长,在适当的存储条件下,DNA分子可以稳定保存数千年^[3];第二,存储密度极高,理论上,它的存储密度可以达到455 EB/g^[4];第三,随着核酸合成和测序技术的快速发展,DNA的写入和读取成本在不断降低,预示着DNA信息存储会更具有商业竞争力^[5-7];第四,在核酸中编码数据不仅能降低操作所消耗的能量,还能使其效率比闪存设备高出几个数量级^[8]。DNA的这些特性使其成为信息存储的绝佳选择。

近年来,DNA作为信息存储介质逐渐成为研究热点,受到广大科研人员的关注。大量综述对DNA存储技术的各个方面进行了深入讨论,主要涵盖了以下几个方面:

① 编码和解码技术:许多综述介绍了DNA存储的编码与解码技术,包括如何将二进制数据转换为核苷酸序列,以及如何设计高效的纠错码来应对DNA合成、测序过程中出现的错误。例如,Heinis等^[9]对不同的编码策略、纠错机制和信息密度进行了详细的对比,强调了在DNA数据存储中的挑战与解决方案。

② 合成技术:DNA的合成最新进展在存储效率和成本方面起着决定性作用,如Yu等^[10]详细阐述了DNA合成方面的最新发展,并指出了未来DNA存储在降低成本和提高效率方面的潜力。

③ 存储介质的稳定性与可持续性:DNA作为存储介质的耐久性和长期稳定性也是研究的重点。Tan等^[11]的综述探讨了DNA的稳定性,涵盖了DNA在各种环境条件下的存储性能,并提出了化学保存方法的改进建议。

尽管现有综述对DNA存储的各个方面进行了全面讨论,但随着研究的不断深入,编码技术、纠错方法、随机访问以及DNA数据的加密等关键技术领域取得了新的进展。因此,本综述将重点关注这四个方面,为该领域的进一步研究和应用提供参考。DNA信息存储的流程主要包括六个步骤:编码、写入、保存、检索、测序和解码。在过去几十年来,生物技术的重大进步促进了DNA信息存储技术的发展,比如DNA信息存储过程中用到的化学和酶促DNA合成^[12]、用于DNA扩增的聚合酶链反应(polymerase chain reaction, PCR)^[13]和DNA测序技术^[14],这些重大发现已经使得写入、随机访问、读取和编辑DNA序列中的数据程序成为可能^[15]。本文首先介绍了DNA存储的基本流程,接着介绍了DNA存储过程中涉及的关键技术,并对这些关键技术进行归纳总结,最后讨论了DNA信息存储面临的主要挑战和发展趋势。

1 DNA信息存储的基本流程

DNA信息存储技术是将图片、文本、音视频等数据文件编码为特定的DNA序列,以人工合成的DNA为存储介质进行存储,通过生物测序等技术读取数据信息的一种存储技术^[16]。DNA信息存储的基本流程如图1所示,主要包括编码(将信息编码为特定的DNA序列)、写入(合成DNA)、保存(体内或体外保存)、检索(随机访问)、测序(读取)和解码。

1.1 编码

DNA由四种不同的碱基组成,即腺嘌呤(A)、胸腺嘧啶(T)、胞嘧啶(C)和鸟嘌呤(G),它们通过特定的顺序排列。DNA信息存储的首要任务是编码,把数字信息转换为特定的DNA序列^[17],将计算机中存储的以0、1为基元的二进制数通过特定的编码规则映射为碱基序列,同时遵循生物化学约束限制。高效准确的编码算法是DNA信息存储中必不可少的,它会影响DNA存储的稳健性和读写的一致性,在重建DNA数据系统中也起着非常重要的作用^[18]。

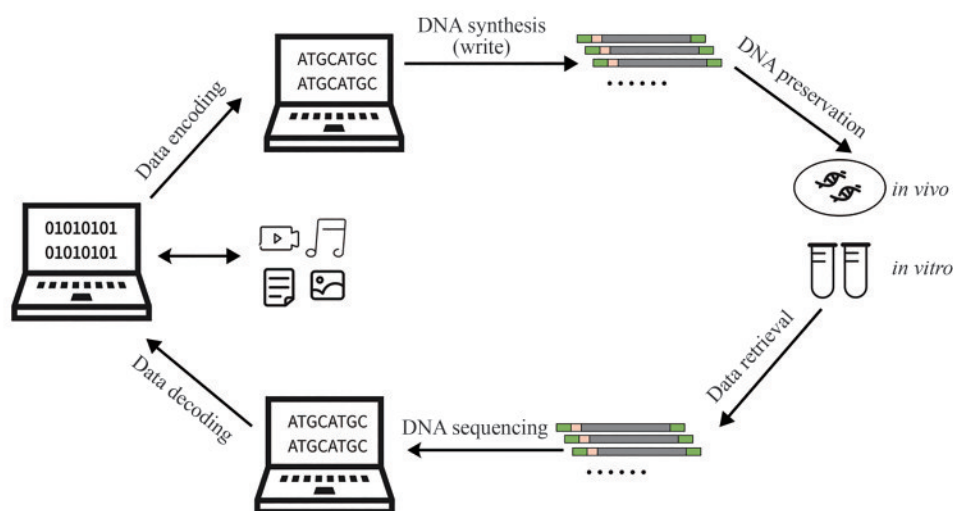


图1 DNA信息存储流程

Fig. 1 Process for DNA information storage

1.2 写入

经过编码后, 需要将生成的DNA序列写入DNA分子中。写入是将数字数据真正转化为分子结构的关键步骤, 通过化学和生物反应将碱基单体偶联, 然后逐个组装成片段的过程。目前, 主流的DNA分子合成方法包括PCR法^[13]、酶促合成法^[12]、芯片法^[19]。DNA合成技术的水平决定了数据存储的整体质量。

1.3 保存

在DNA信息存储过程中, 为了保证DNA序列的高准确性、可重复性和可靠性, 需要延长DNA的寿命。长期保存DNA数据的方法包括体外保存和体内保存, 迄今为止, 大多数DNA信息存储都是在体外进行的。光、水和氧气影响DNA的保存, 会引起化学反应, 导致合成的DNA断裂和突变, 使得编码后的寡核苷酸信息内容难以破译。体外保存通常涉及将合成的单链DNA组装成化学结构更稳定的双链DNA, 随后将其纯化并在低温下进行体外保存^[20], 通常以冷冻粉末、溶液或封装在纳米颗粒中进行保存^[21]。体内保存一般是保存在细菌的质粒或人工染色体中^[22-23], 或者将其整合到活生物体的基因组中进行长期保存^[24]。与体外DNA信息存储相比, 体内存储利用了细胞内高效的DNA复制、校对和长链DNA维护机制, 从而

提供了无组装即可随机访问数据的机会^[25]。

1.4 检索

检索是指从大量DNA库中随机访问特定的DNA数据集, 由于样本池中缺乏物理组织, 因此在分子存储中很难找到带有特定数据的目标DNA片段。随机访问最广泛的方法是向DNA序列中添加经过独特设计的引物, 通过PCR扩增的方式获取所需的DNA序列^[26]。目前已经成功实现从超过1300万个DNA寡核苷酸池中准确无误地检索所需要的数据^[20]。其他的随机访问方法还有基于微阵列的随机存储方法^[27]、DNA分子的固定化^[28]、数字微流控液滴^[29]以及DNA条形码二氧化硅珠^[30]。

1.5 测序

DNA测序技术是指通过生化方法提取存储的DNA分子, 识别目标碱基序列, 获得写入的编码数据。它用于读取和拼接寡核苷酸池中的包含碱基序列的DNA片段。在早期的研究中, 读取DNA中所包含的数据需要对整个分子进行测序, 后来出现的基于阵列的DNA数据存储技术, 通过将合成的DNA固定在阵列的指定位置, 进而可以直接通过芯片进行寻址。例如DNA微盘^[28]、滑动芯片^[31]、边合成边测序等方法可以实现快速访问和较高的可操作性^[32]。将DNA结合在芯片表面, 在

复制过程中不会受损或丢失，易于获取和处理，而且减少了对PCR引物的选择和大规模PCR扩增的需求。

1.6 解码

与编码相对应，解码是DNA信息存储的最后一步，是指使用与编码功能相反的计算机算法将DNA序列转换为原始的二进制数据。

2 DNA信息存储流程中的关键技术

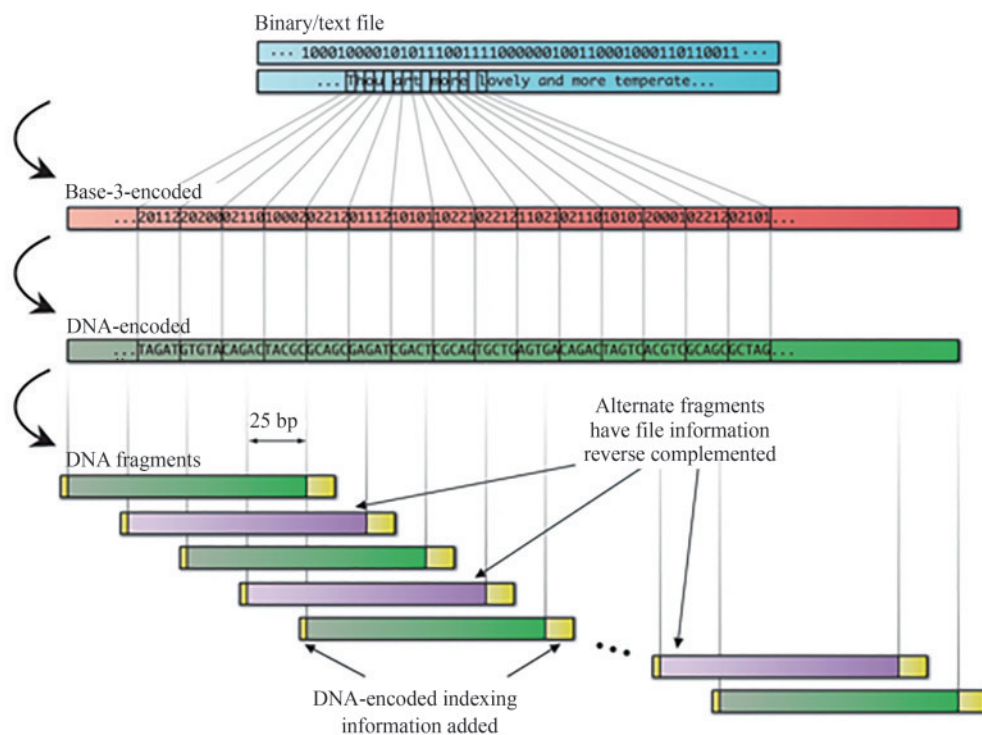
2.1 编码技术

尽管DNA存储有众多的优点，但也有几个缺点需要克服。由于DNA存储的合成成本相对较高，因此需要一种高效的编码算法，能以较少的核苷酸数量存储大量的数据。另外，DNA存储的错误率受DNA生化结构的影响很大。由于这些生物化学约束限制会导致DNA在合成和测序过程中产生高错误率，因此应满足生物化学约束限制，如GC平衡率、均聚物的运行长度。GC平衡率是由G和C核苷酸的数量与整个序列的比例决定的，这个比率需要接近0.5，因为高或低的GC平衡率会导致聚合酶链反应(PCR)错误^[33-34]，平衡这一比率会使合成和测序过程中的错误率降低。均聚物的运行长度是指DNA链中连续相同的核苷酸的最大数量。众所周知，如果均聚物的运行长度超过6个，替换和删除的错误率会增加^[33-34]，这也会造成DNA存储过程中的高错误率。

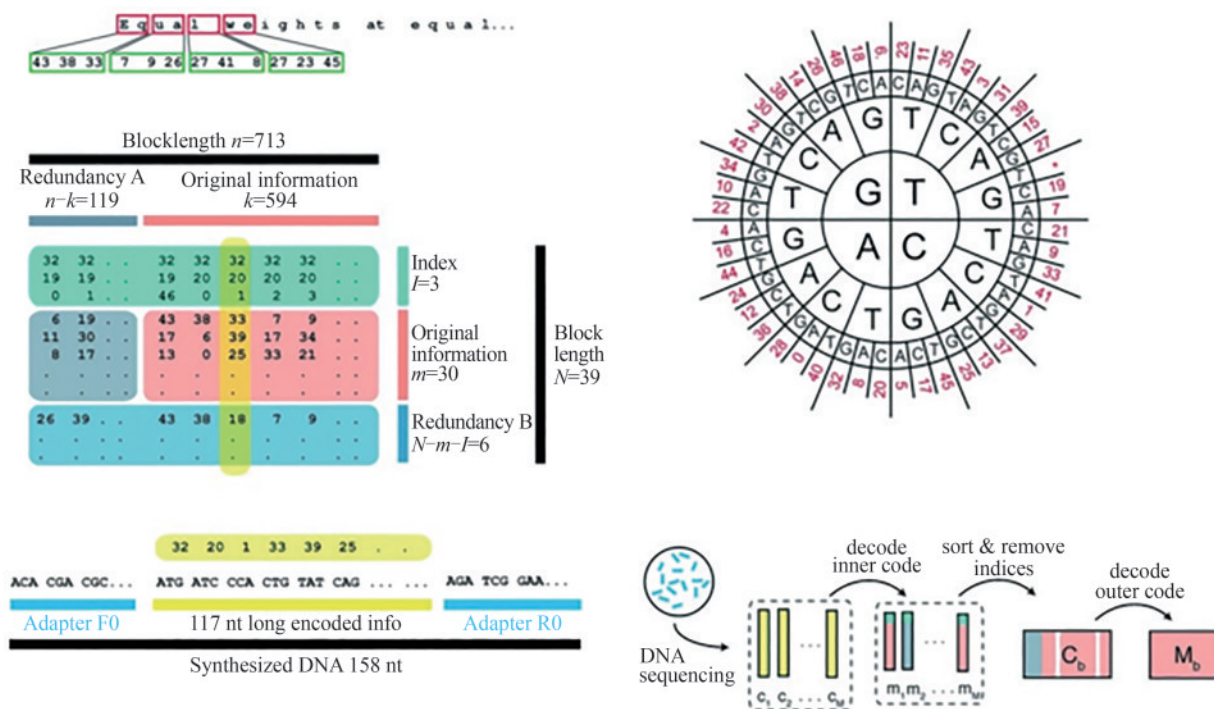
1948年，香农发表了论文《通信的数学理论》^[35]，创建了信息论这一新兴学科理论，为数学和通信工程中信息的传输、编码和处理提供了理论基础和实用方法^[36]。在DNA存储领域，香农信息论同样涉及信息的转换，而编码方法的发展也在这一理论框架下得到了迅速推进。2012年，Church等^[2]提出了一种二进制转换的方法，通过将每个碱基编码为1位数据，其中A或C代表0，G或T代表1，实验中，将5.27 MB的数据编码到DNA中。该编码方案为DNA信息存储技术奠定了坚实的基础，并推动了该领域的研究和应用。2013年，

Goldman等^[4]提出了基于霍夫曼编码的三元信息转换方案，编码时，DNA序列的碱基由当前数据信息和前一位已经被选择的碱基共同决定，该方案同时利用分段保存将每条信息复制到四个不同的分子中，大大提高了稳定性，但是在一定程度上增加了成本[图2(a)]。为了避免在编解码过程中出错的问题，2015年，Grass等^[3]首次引入纠错码，利用里德-所罗门码(Reed-Solomon codes, RS)进行内外部纠错，解决了在编解码过程中碱基出错或丢失的问题，首先将文本文件中每两个字母通过碱基转换映射到47阶伽罗瓦域的三个元素中，在外部编码的步骤中，采用RS编码向各个块中添加冗余A，并使用内部RS编码生成冗余B，将G(47)的每个元素映射为三个核苷酸，两端添加两个接头序列，DNA序列长度达到158个核苷酸。在读取时，DNA序列被转换为G(47)，然后通过内部和外部解码恢复数据信息[图2(b)]。Erlich等^[37]提出了基于Luby变换的喷泉码编码算法，将二进制信息预处理成一系列固定长度、不重叠的片段，随机选择可变数量的序列片段进行异或运算，再添加固定长度的种子形成液滴。然后筛选出满足约束限制的序列，并迭代重复这两个步骤，直到生成足够有效的寡核苷酸。随后，Anavy等^[38]提出了一种使用复合DNA字母的编码方案，大大增加了DNA信息存储的逻辑密度，传统的DNA编码使用四种核苷酸(A、C、G、T)进行编码，而该研究提出了一种将核苷酸以特定比例混合成“复合DNA字母”的方法，从而扩展了编码字母表。这一方法有效地在每个合成周期中存储更多信息，减少了合成周期的数量。在实际应用中，研究人员使用这种编码技术成功地将6.4 MB的数据编码到复合DNA中，合成周期减少了20%。他们进一步通过模拟实验，表明更大的复合字母表可以将合成周期减少多达75%[图2(c)]。

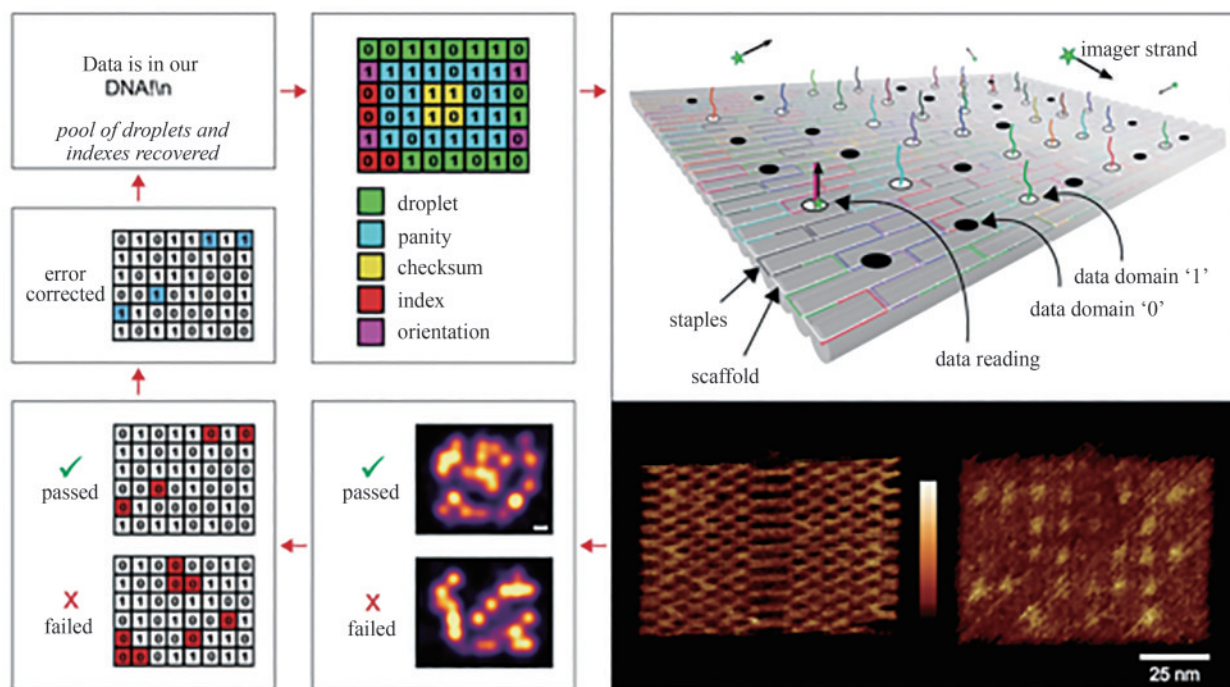
2021年，Dickinson等^[40]通过采用多层纠错方法的系统对数字信息进行编码，数据被分成大小相等的子串，然后使用异或(exclusive OR, XOR)策略组合形成液滴数据块。每个液滴被编码成预定大小的矩阵，然后分别添加索引、方向信息、校验和奇偶校验位，该研究展示了DNA折纸技术



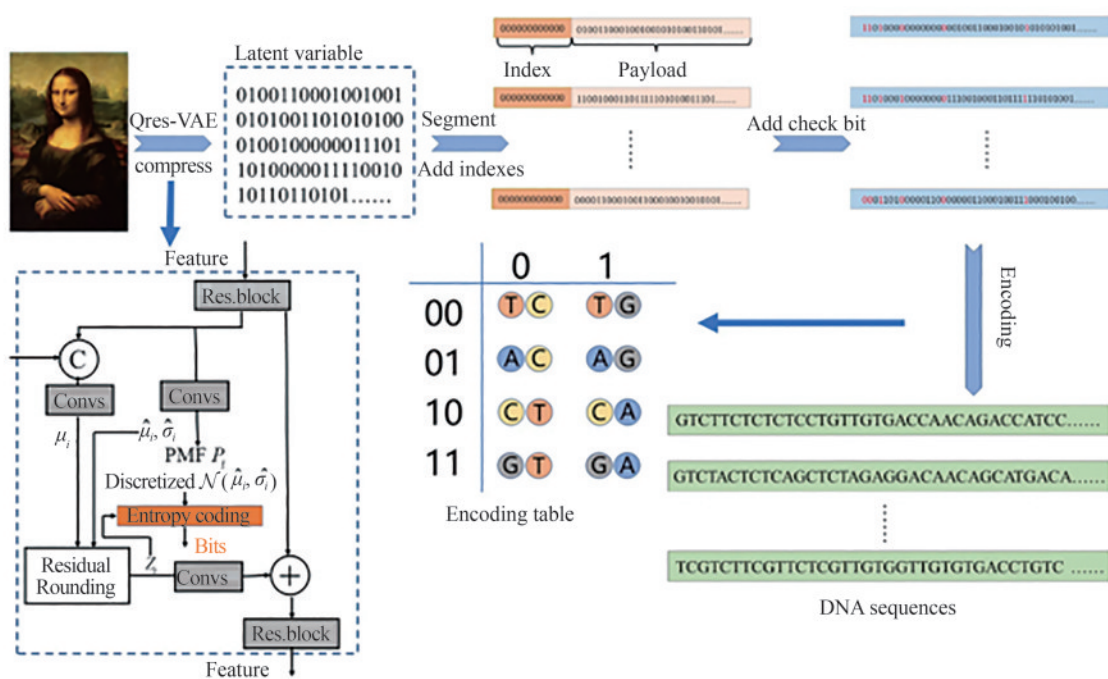
(a) DNA中的数字信息编码
(a) Digital information encoding in DNA



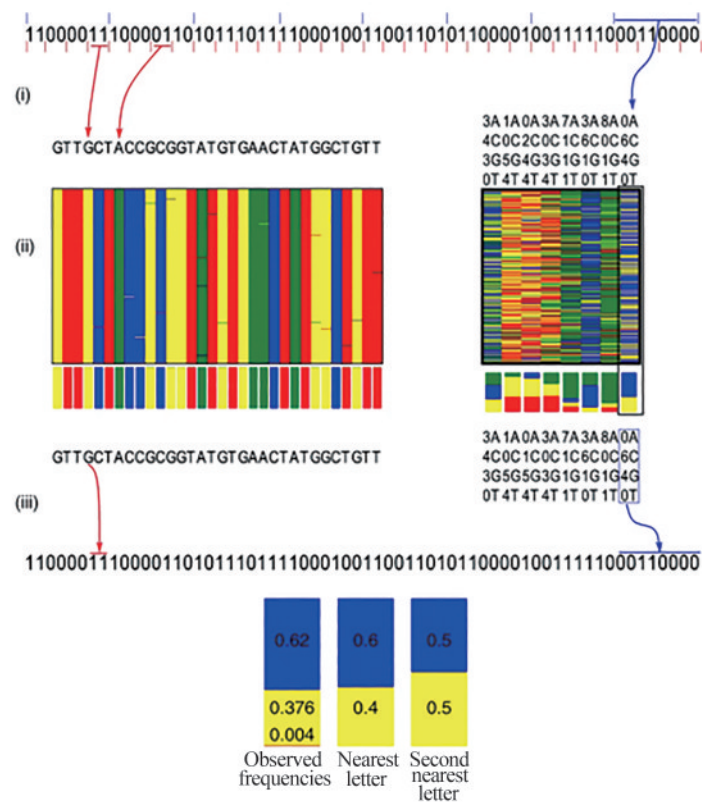
(b) 通过里德-所罗门编码将文本编码到DNA中
(b) Encoding text to DNA by Reed-Solomon coding



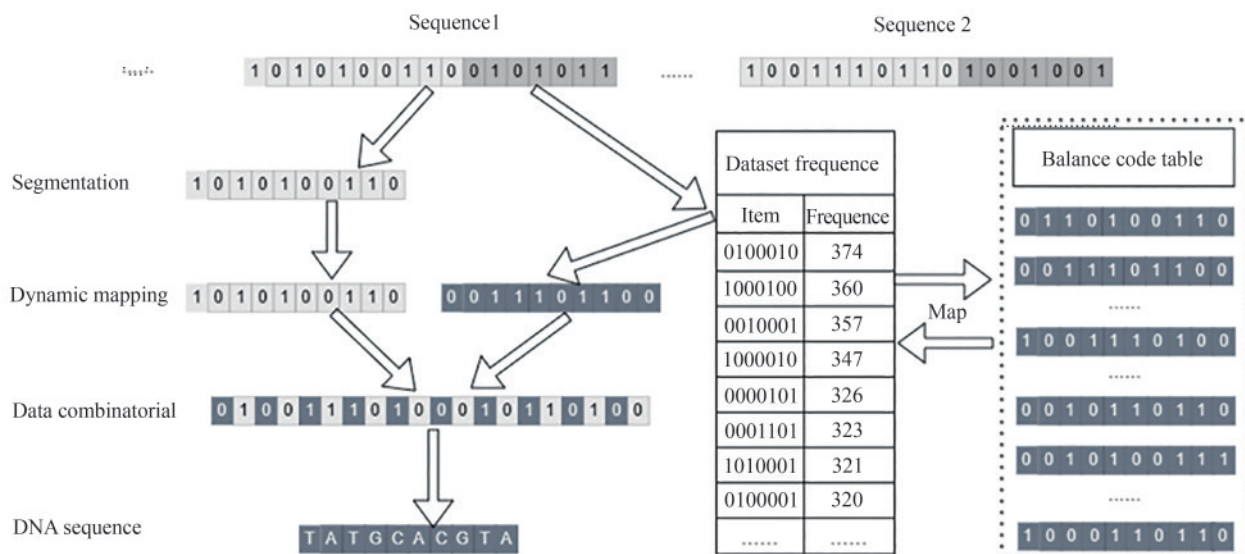
(c) 二进制dNAM概述
(c) Binary dNAM overview



(d) DNA-QLC编码方案流程图
(d) Flowchart of the DNA-QLC encoding scheme



(e) 使用标准和复合DNA对二进制消息进行编码
(e) Encoding a binary message using standard and composite DNA



(f) DNA-base128编码过程流程管道
(f) Pipeline of the DNA-base128 encoding process

图2 DNA 编码方案比较 [3-4, 38-41]

Fig. 2 Comparison of DNA coding schemes [3-4, 38-41]

在数字存储中的潜力,提供了一种无需测序的新方法,将编码与读取的准确性和效率显著提升,为DNA存储编码技术提供了新思路和研究方向[图2(d)]。为了解决DNA存储过程中净信息密度低和错误率高的问题,Zheng等^[42]提出了改进的藤壶交配优化器和有效载荷编码(modified barnacles mating optimizer and payload encoding, MOPE)编码算法。该算法具有较高的网络信息密度并满足相应的生物约束条件,可以有效降低DNA合成与测序的成本。

DNA序列的二级结构是由序列向后折叠而形成的,这有时会导致DNA分子的双螺旋结构变得更加脆弱,影响DNA的复制和解旋过程。Chu等^[43]介绍了一种显式构造避免二级结构代码的方法,并使用递归和贪婪算法探索了能够生成具有良好速率的m-SSA代码的良好生成集。2022年,Park等^[44]提出了一种新型的DNA存储迭代编码算法,通过使用贪婪算法来满足GC平衡和运行长度的约束,与随机生成的映射方法相比,该方案通过提出的一种新的映射方法使平均比特误差减少了20.5%。同年,Ping等^[45]提出了一种名为阴阳编解码器的转码算法,通过使用两种不同的规则将两个二进制位编码为一个核苷酸,用N1、N2、N3和N4分别表示四种核酸A、T、C和G,生成满足约束条件的DNA序列,阴阳编解码器编码算法的优势在于,通过阴阳规则最终可以产生1536种适用于各种数据类型的编码方案,并且可以有效消除长均聚物的生成且GC含量也满足约束条件。2023年,Lu和Kim^[46]提出了一种结合弱互相关代码(weakly mutually uncorrelated codes, WMU)与均聚物的运行长度约束的编码技术,来提高DNA存储中引物设计的效率和准确性,该研究将WMU代码与均聚物的运行长度约束相结合,来增强编码的鲁棒性。与先前的研究相比,他们所提出的编码方案始终能够保证均聚物的运行长度约束,提升了引物设计的精确性与效率。2024年,Wang等^[39]提出了通过base128编码方式将图像存储在DNA中,在数据写入阶段,进行数据切分和概率统计,并将数据块频率与约束编码集关联来实现编码。当需要恢复图像时,DNA-base128通过阈值设置和漂移比较完成内部纠错[图2(e)]。与

前人的工作相比,DNA-base128编码结果显示,不良模体减少了71.2%~90.7%,局部鸟嘌呤-胞嘧啶含量方差降低了3倍,表明了DNA-base128能够更稳定地存储图像信息。为了实现更高效的DNA图像存储,同年,Zheng等^[41]提出了一种名为DNA-QLC(QRes-VAE和Levenshtein Code)的编码方案,旨在解决DNA存储中编码密度低的问题,该方案采用了QRes-VAE(量化ResNet VAE)模型对图像进行压缩,并结合Levenshtein代码(LC)进行错误校正,从而实现了高效且可靠的DNA存储[图2(f)]。实验结果表明,DNA-QLC编码方案在保持信息密度的同时,确保了DNA序列满足生物化学约束条件,避免了出现不期望的DNA序列。

综上所述,基于简单映射关系的编码算法在固定的映射规则下,确定数据比特与碱基之间的映射规则。解码速度快^[47],但过于模式化,合成时容易引入模式错误,同时非有效载荷和校验冗余的存在也会在一定程度上降低编码的信息密度。基于约束的编码方法虽然相对复杂,但可以一次性满足DNA存储的均聚物、GC含量、避免二级结构等多种生物化学约束,大大保证了合成和测序的准确性。该类算法的编码性能比较容易受二进制输入文件和参数选择的影响,难以设计出各类型约束的组合。从整体编码角度看,目前编码算法的研究思路主要集中在提高编码速率、信息密度、生化约束兼容性^[48]、数据的随机读取以及无误差的数据重建。随着计算机和信息技术的发展,有望出现更加鲁棒的DNA编码方案,以进一步提高编码质量。

2.2 纠错技术

在DNA信息存储的合成和扩增步骤中会出现错误,因此,想要将数据准确无误地解码出来面临着巨大的挑战。虽然扩增主要引入的是替换错误(例如,将G读成C),但DNA的合成主要会导致缺失(例如,缺失一个碱基),每碱基发生率约为0.2%~1%^[49-50]。插入(添加额外的碱基)错误并不常见,通常每碱基发生率不到0.1%^[49]。表1为DNA存储纠错技术研究对比。

表1 DNA存储纠错技术研究对比

Table 1 Comparison of error correction technologies with DNA storage

作者	纠错码类型	码率	纠错类型	解码算法	参考文献
Bornholt等	XOR+重叠编码	约0.67	插入、删除和替换错误	聚类、对齐、多数投票	[1]
Sun等	LDPC	0.5, 0.9	替换错误	非对称错误感知BP(ABP)解码算法	[51]
Antkowiak等	RS纠错码	0.4	插入、缺失和替换错误	聚类、对齐、多数投票	[52]
Ding等	软判决译码方法+RS纠错码	0.83, 0.92, 0.945	插入、缺失和替换错误	软判决解码策略	[53]
Lu等	LDPC+LLR	—	插入、删除和替换错误	LDPC码的和积算法	[54]
Fei等	LDPC+类Turbo	0.5	插入、删除和替换错误	基于LDPC的解码算法	[55]

其中对于替换错误，Sun等^[51]提出了一种改进的BP算法，通过使用两个二进制低密度奇偶检查码（low density parity check code, LDPC code）来编解码一个四进制DNA序列，此外，外部信息被用作附加信息，在解码过程中校准LLR值，从而提高解码性能。Sun等^[56]探讨了一种新方法来解决PCR扩增错误导致测序数据不准确的问题，作者发现，PCR错误会影响大规模和单细胞测序数据。为了解决这个问题，他们提出使用三核苷酸块来合成UMI，并通过多数投票法进行纠错。这一方法比传统的单体UMI方法更有效地纠正了错误，提升了测序的准确性。除了扩增产生的偏差之外，存储过程中的主要问题是序列拷贝数分布会发生变化，随着时间的推移，可能会导致某些序列不可恢复地丢失，例如，94%的DNA衰变后（即四个半衰期）会丢失8%^[57]，这表示序列将永远不能完全恢复。因为解码过程主要依赖于序列信息，因此单个序列的丢失和这些序列中引入错误都会影响解码的正确性。虽然可以通过增加簇序列信息的物理冗余来缓解这个问题，但是它会大大降低信息密度，所以这样做是不可取的，而且效率低下^[50]。因此，针对DNA信息存储过程中的错误问题，可以通过一定的编解码算法解决，通过现代纠错码组合引入的逻辑冗余可以大大降低错误率。但是，针对以删除错误为主的错误率，一般使用物理冗余来恢复原始信息^[52]。良好的纠错码可确保以最小的冗余度来恢复数据。纠错码可以实现的最大信息速率在理论上有界限的^[35]，这个界限称为信道容量。Goldman等^[4]使用了一个简单的纠错码，将每个部分的信息写在四个后续DNA序列上，当丢失的后续序列少于四个，则可以纠正序列，但是，在输入的字符串中，每

一个块重复四次会产生大量冗余。2016年，Bornholt等^[1]在Goldman纠错方案的基础上提出了XOR运算，该方法通过将两条DNA链中的有效载荷A和B进行异或，并记录异或后的结果C，这样可以通过A、B、C三条链中任意两条来恢复第三条链，而且每个核苷酸仅平均重复1.5次，存储密度远高于Goldman提出的方案。虽然这种方法可以提高数据传输和存储的可靠性，但整体的纠错能力相对比较有限。对于更复杂的纠错要求，可以考虑RS纠错码，它是一种利用多项式运算和有限域（又称伽罗瓦域）的概念进行编码和解码的级联编码。在Meiser等^[58]的编码方案中，内码和外码均采用RS纠错码，用来纠正DNA合成和测序过程中可能出现的错误。Antkowiak等^[52]提出了一种基于RS码的DNA序列校正方法，将原始数据流划分为10977个长度为14×6位的序列，在内外编码器均采用RS纠错码的条件下，提出了一种基于光定向合成的低成本DNA存储系统，该系统即使在高错误率的情况下也能实现高可靠的信息存储。DNA存储系统中ECC的选择需要在纠错能力和冗余度之间取得平衡。为了解决这个问题，2024年，Ding等^[53]提出了一种软判决译码方法（derrick），主要利用误差预测模型来减少未知变量的数量，使RS码的纠错性能提高一倍。这种软判决译码方法在体外和计算机实验中均显示出较高的纠错能力。

目前像经典的纠错码如喷泉码^[37]、Bose-Chaudhuri-Hocquenghem码（BCH）^[59]、Reed-Solomon码都被广泛应用于DNA信息存储中^[20]。除此之外，Lu等^[54]提出了一种基于观测统计数据的对数似然比（LLR）处理方案，可以减少DNA存储通道中LDPC解码的读取开销。Fei等^[55]针对

DNA 存储通道中的二进制 LDPC 码提出了一种类 turbo 解码算法, 称为外部信息辅助信念传播 (EBP) 解码算法, 该解码器可以纠正纳米孔测序引起的替换错误。然而, 解码过程中通道信息并未得到充分利用。

在 DNA 存储中, 合成和扩增步骤引入的错误对解码的准确性形成了巨大挑战。合成过程主要产生碱基缺失, 扩增过程则常引入替换错误。为解决这些问题, 研究人员开发了多种纠错技术。例如, 通过 LDPC 码改善解码性能, 并通过多数投票法提高了对于 PCR 扩增错误的纠错效果。尽管物理冗余能够减少序列丢失问题, 但效率低下。因此, 现代纠错码如 Reed-Solomon (RS) 码和喷泉码在存储系统中得到广泛应用, 其中软判决译码和 LDPC 改进算法展现了卓越的纠错能力。

综合来看, DNA 存储的纠错技术正逐步趋向提高纠错能力与存储密度之间的平衡, 以确保数据的高效和准确恢复。在未来, DNA 存储的纠错技术可能会朝着综合应用多种方法的方向发展, 以便同时兼顾存储密度、纠错能力和计算效率。结合不同纠错编码的优势, 优化 DNA 数据存储系统的整体性能。

2.3 随机访问

虽然 DNA 信息存储的大量工作集中在改进编码方案上, 但信息存储系统的另外一个重要方向是有效检索特定文件或文件任意子集的能力。随机访问是指访问数据存储中的任意位置时所花费的时间是相等的, 与顺序访问相对, 在顺序访问中, 访问特定数据所需的时间取决于数据在存储介质中的位置。在 DNA 存储中, 随机访问允许直

接访问存储在 DNA 中的特定数据段^[60], 不需要依次逐个读取整个序列, 从而减少了数据重建延迟^[61]。分子随机存取在很大程度上依赖于传统的聚合酶链反应 (PCR)^[62], 为了频繁访问存储信息可以将 DNA 存储在水溶液中^[63]。目前主流的随机访问方法有 PCR 引物寻址和 PCR-free 两种。PCR 引物识别目标序列, 可以实现对特定数据的访问^[30]。PCR-free 随机接入采用直接序列标记和识别策略来提取靶标信息, 避免了引物设计、消耗和序列特异性低的限制^[21]。这两类方法各有优势, 在不同的应用场景中发挥着重要作用, 共同推动了 DNA 存储随机存取策略的进展。

基于 PCR 的 DNA 存储随机访问主要依靠预先设计的特定引物来识别目标序列。使用 PCR 或变异 PCR 进行 DNA 存储随机访问的研究对比如表 2 所示, 虽然可以通过增加引物的长度来扩大容量, 但这样做 DNA 信息存储密度会受到影响。

为了在 DNA 存储中实现随机访问, 2015 年, Hossein Tabatabaei Yazdi 等^[66]提出了一种既能随机访问数据块又能重写存储块内任意数据信息的存储架构, 该架构克服了只能读的缺点。将存储信息的编码按照地址序列的前缀进行拼接, 从而避免了在实际编码序列中出现地址, 这样仅通过地址和 PCR 扩增技术就可以准确定位和检索目标块。Bornholt 等^[64]对四个图像文件进行了编码, 大小从 5 KB 到 84 KB 不等, 来证明在 DNA 存储中使用随机访问功能的可行性。2018 年, Organick 等^[20]为了实现大规模 DNA 存储中的随机访问, 设计并验证了一个大型引物库, 成功地将存储在 1300 万个 DNA 寡核苷酸中的 200 多兆字节的数据成功编码并实现了随机访问, 而且可以单独无错误地恢复每个文件。近年来, 越来越多的研究人员将目光转向 DNA 纳米结构、微阵列技术等手段, 希望

表2 使用PCR或变异PCR进行DNA存储随机访问的研究对比

Table 2 Comparison of studies using PCR or variant PCR for the random access of stored DNA

作者	数据大小	测序技术	覆盖率	随机访问	参考文献
Organick 等	200.2MB	Illumina/Nanopore	5X/36X	ePCR	[20]
Bögels 等	25MB	Illumina	30X	Thermoconfined PCR	[21]
Hossein Tabatabaei Yazdi 等	3KB	Nanopore	200X	PCR	[26]
Erllich 等	2.15MB	Illumina	250X	PCR	[37]
Bornholt 等	150KB	Illumina	40X	PCR	[64]
Lau 等	—	Nanopore	30X	PCR	[65]

通过这些技术手段能够进一步拓展地址空间。El-Shaikh等^[27]通过喷泉码对数据进行编码,将设计的大量差异化的DNA探针作为索引。借助微阵列技术,可以在单个DNA池中索引数万到数百万个数据对象。此外,还使用局部敏感的哈希算法高效计算DNA序列之间的相似性,并通过同时合成多个探针实现对多个对象的并行访问。Lau等^[65]提出了基于磁性DNA的随机存取存储器存储策略,该策略采用聚合物树形条形码方案。与传统的为每个文件合成单独寻址引物的方法不同,该聚合物条形码寻址策略大大减少了所需的引物数量,并使大规模随机存取成为可能。通过将合成DNA与磁性琼脂糖珠结合,能够实现重复读取数据,同时保留原始DNA样本,并确保数据读取的质量。

PCR引物寻址主要通过引物扩增目标DNA信息片段,然后将其从储存中提取出来^[42]。对于实现随机存储来说,设计高度特异性的引物至关重要,随着存储数据量的增加,对引物的需求也随之增加。在设计引物库时,重要的是同时考虑质量和数量。Limbachiya等^[67]采用利他算法构建满足GC含量和均聚物约束的引物库,该算法可枚举满足约束的DNA码字,并且在数量上表现出显著的提升。后来,一些研究发现了一种基于智能算法驱动的引物设计方法^[68-70],利用元启发式算法和组合约束来构建高质量的引物库并增加引物数量,从而提高随机访问能力。例如,Yin等^[71]采用非线性控制参数和随机对立学习策略,在满足约束的前提下可以有效避免陷入局部最优解,从而提高引物库的下界。同时,将DMVO算法与汉明距离、GC含量、均聚物约束相结合,构建引物库^[68],保证引物数量的同时,避免了PCR过程中引物的非特异性杂交。Cao等^[72]提出了一种热力学最小自由能约束,并将该方法应用于DNA存储中,其中布朗多元宇宙优化器(BMVO)算法在多元宇宙优化器(MVO)算法的基础上,融入了布朗运动的思想 and Nelder-Mead方法,用于设计更优的DNA存储编码集。与以前的工作相比,该编码集在规模上增加了4%~50%,并具有更好的热力学性质。随着DNA编码集质量的提高,读写的准确性和DNA存储系统的鲁棒性也增强。2021年

Rasool等^[73]提出了MFOL算法,通过基于对立学习策略来增强三维搜索空间的最优解,从而提高引物库的下限。随后,该团队又创建了基于神经网络的计算模型^[74],并结合生物化学约束设计大型引物库,从而能够用更短的引物检索出更多的存储信息。在此基础上,Cao等^[75]通过将图神经网络与约束编码深度结合,提出了一种基于GNN的DNA存储随机接入引物设计算法,该方法大大减少了引物构建时间,提高了引物的丰富度。随后,Wu等^[76]提出了通过末端约束来提高序列的稳定性,并引入随机开关和双重权重后代策略,将引物库的下界提高了9%~37%。这些研究方法旨在实现高质量的引物库设计,为随机访问的实现提供了更多的可能。

近年来,研究人员探索了多种非PCR策略,来克服引物数量限制而导致的可扩展性问题。Shipman等^[77]成功利用CRISPR-Cas系统将信息存储在DNA中,并将像素值编码到活细菌群体的基因组中。CRISPR不仅具有数据编码和重写的潜力,而且在DNA信息存储系统中还被用于数据的高级搜索功能^[78]。此外,研究人员还通过利用天然DNA聚合物实现了随机访问,并利用内置并行性来加速信息的快速处理^[79],通过将DNA与其他非生物材料结合来实现新的数据访问模式。2021年,Fan等^[80]提出了镜像DNA信息存储,利用合成的镜像Pfu DNA聚合酶,研究团队在I-DNA中存储了路易·巴斯德的一个段落信息,并成功地从存储库中提取和解码。实现随机访问除了通过酶促技术外,近年来,Newman等^[29]提出了一种使用数字微流控技术检索DNA数据的方法。该方法首先将脱水的DNA存储在连接到数字微流控芯片的玻璃板上的一个子库中,当请求一个文件时,通过芯片上嵌入的电极阵列实现的电润湿技术将一滴水溶液移动到目标位置,然后DNA在水滴中溶解并被运输,方便进行后续的数据读取。该系统利用了高通量数字微流控技术,有望构建自动化的大规模DNA数据库。多子池的物理分布解决了在高密度单一DNA数据库中分子拥挤的困境,然而,物理分离会降低存储密度,因为独立分区占用的空间比混合大型数据库所需的更大。此外,物理分区方法通常具有较小的吞吐量,因

此不适合大规模存储。

在众多的随机访问技术中, 基于PCR的随机访问技术仍然是高通量并行随机访问最有前景的技术。通过使用引物设计的基于PCR的随机访问方法更适合在DNA中存储大量文件或大型单个文件。DNA数据存储在实际应用中仍然面临一些挑战, 包括天然扩增与特定数据检索之间的差距, 以及扩增技术在错误率和保真度之间的平衡问题^[81]。此外, DNA扩增技术需要具备更大的灵活性和多样性, 以满足不同应用场景的需求, 例如从个人数据存档到大规模数据中心的操作。随着技术的高速发展, 基于随机访问能力的DNA信息存储仍在迅速发展, 未来的研究有望通过改进分子设计、引入更高效的生物酶系统或开发新的化学标记方法来提升随机访问的速度和准确性。

3 DNA信息存储的安全性

如今, 数据安全问题已成为社会面临的关键问题, 许多存储的信息具有时效性和机密性, 在未经授权的情况下应禁止访问。密码学通过数学和逻辑设计对信息进行加密, 从而确保了存储信息和通信的安全性^[82]。近年来, 随着计算能力的快速提升, 密码学方法很可能在暴力破解的攻击下被攻破, DNA分子展现出许多独特的化学和物理特性, 例如可编程性和可修改性。如果将这些特性与密码学结合起来, DNA信息存储的安全性将会大大提高^[83-84]。DNA信息加密主要可以分为基于序列的加密和基于结构的加密两类。

在基于序列的加密中, 信息以DNA序列的形式存储在DNA分子中, 信息加密是通过改变或隐藏真实序列信息来实现的。利用DNA的核苷酸序列作为信息编码的载体, 将明文和伪随机序列编码成DNA序列, 再用碱基计算规则将两段DNA序列进行运算形成密文。目前, 大多数DNA隐写术方法都使用引物序列作为密钥, 在发送者和接收者之间秘密共享, 来维持私人信息的连接^[85]。1999年, Clelland等^[86]将加密后的信息与人类的DNA相混合, 并将其加载到滤纸上形成微点, 从而开发出了一种双重隐写技术, 可以用来发送加密后的信息。这项工作的结果表明, DNA可以用

来隐藏和传递信息。然而, 存在一个安全问题, 当引物序列信息被破译后, 编码后的信息可以被任何人读取。后续研究人员提出了各种各样的改进措施, 来增强DNA隐写技术的安全性和隐藏效果。2020年, Cui等^[85]设计了一种更为先进的DNA隐写术方案, 将携带信息的DNA与含有限制性酶位点的随机DNA库混合, 从而增加了一层额外的保护。此外, 他们利用PCR扩增引入的随机性实现了自毁功能和监听检测机制, 类似于量子密钥分发。即使引物被破解, 这种方法也使得信息很难被非法读取。后来, Gehani等^[87]以DNA为信息载体, 通过生化技术实现了对DNA分子的一次一密加密。Lu等^[88-89]提出了基于DNA的对称/非对称机密技术, 可以抵御超级计算机的攻击。随后, 越来越多的研究人员开发了各种各样的加密技术和隐写术。2021年, Fan^[80]通过在DNA信息存储系统中引入镜像核苷酸, 开发了手性隐写术。在正交信息库中, 利用天然聚合酶读出错误信息, 通过镜像PCR过程解码目标信息。Grass等^[90]还提出了一种新的信息隐写术方案, 利用遗传短串联重复序列作为密钥, 采用AES-256加密方案保护DNA中存储的数字信息。该方法将加密信息嵌入合成DNA中, 然后使用NGS技术读取个体DNA和合成DNA以检索密钥和编码信息。由于STR信息的高熵特性, 这种生物密钥对暴力解密具有很强的抵抗力。2023年, Yao等^[91]通过借鉴分子生物学中的信息处理机制, 提出了一种面向DNA存储的图像加密算法, 基本思想是通过基因杂交进行像素替换, 通过像素扩散和基因突变实现双重扩散。加密后合成密文DNA图像并存储在DNA存储系统中。该方法能抵御常见的攻击, 并对DNA存储通道中的序列丢失和碱基替换错误表现出很强的鲁棒性。

在基于结构的加密方面, 通过折叠DNA链形成复杂的三维纳米结构, 将信息嵌入到这些结构中。加密复杂度更高, 抗破解能力更强, 且信息容量更大。与序列信息加密不同的是, 结构中存储的信息可以被原子力显微镜 (atomic force microscope, AFM) 或修饰的纳米孔直接读取, 通过引入钥匙链片段, 使DNA结构发生改变, 从而读出结果, 隐藏在结构中的信息可以解密出来^[92]。

DNA折纸是一种生物分子自组装技术^[93-96],通过数百条短“钉书针”链的帮助下折叠一条长“支架”链来生成基于DNA的纳米结构。其纳米级可寻址性允许将分子和纳米物体精确组织成复杂的图案^[97-101]。近年来,DNA折纸技术已经被用来增强DNA信息的安全性。2014年,Yang等^[102]提出了一种利用DNA自组装结构的DNA信息加密方法,通过利用DNA自组装链识别和链置换进行逐位异或运算,以一次性密码本的形式实现信息的加密和解密。DNA折纸加密通信方法利用DNA自组装技术设计手印式DNA纳米图案。将信息编码到适应的DNA支架中,然后通过自组装形成手印图案以加密信息。该方法创建了一个超过700位的巨大密钥空间,并通过在不同的DNA折纸之间建立连接,该方法可以确保接收信息的完整性并实现差异化访问控制^[103],此外,DNA折纸加密技术被证明是信息加密传输的通用方法,可以实现音符和像素图像的传输。后来,Fan等^[104]展示了一种可重构的DNA折纸加密技术,信息首先存储在“之前”的构象中,在加入解密密钥链后,阵列构象发生转变,解密信息在“后”构象中可视化。随后,为了增加系统的复杂性和安全性,他们发展了基于立足点介导的链置换反应的两步解码策略。在解码过程中,依次添加两个解密密钥来恢复加密信息。随后展示了一种基于DNA条形码纳米结构的信息加密方法^[105],通过添加加密密钥段,可以在主载体上构建四元信息发夹结构,并精确地添加和移除。纳米孔读出的信号随着发夹结构从加密形式到解密形式的转变而变化。同样,DNA异构算法也是以基因作为密钥,对数据进行加密,并利用DNA进行存储,这种DNA加密方法计算速度更快,安全性更高^[106]。2022年,Zhu等^[107]提出了一种可操作的DNA链置换加密方法。该研究方案在每次加密中调整生化反应的参数,并根据DNA链的浓度变化来获取密钥。同年,Teng等^[108]提出了一种带深度学习的加密算法,利用深度学习技术生成DNA加密的密钥并生成密文,生成的密钥的权重矩阵存储在DNA中,增加了系统的安全性,减少了密钥中的数据量。此方法在编码阶段减少了原始文件中的重复序列,均平衡了GC含量,增加了系统的容错能力。

综上所述,基于序列的加密方法能够实现较高的存储密度,这种方法依赖于现有的DNA合成和测序技术,由于目前这些技术比较成熟,被广泛使用,而且易于实施和扩展,但是如果没有任何额外的加密措施,只要获取到DNA样本并进行测序,就容易被破解,从而获得存储在序列中的信息。基于结构的加密方法是通过DNA的三维结构(如DNA折纸、纳米结构等)的变化来存储和隐藏信息。这种方法不依赖于DNA序列本身,而是利用DNA分子的折叠、组装等物理结构特性来加密数据。由于信息是通过DNA的三维结构隐藏的,破解这些信息需要了解具体的结构设计和折叠路径,即使获取到DNA样本也不容易解码。此外,与基于序列的加密不同,基于结构的加密难以通过常规的DNA测序技术直接读取,需要专门的设备进行读取,增加了数据的安全性。而且基于结构的加密方法提供了新的信息隐藏维度,尤其适用于需要更高隐蔽性和安全性的应用场景。但因其需要精确控制DNA的折叠和组装过程,技术实施难度大,并且目前的技术仍然处于研究阶段,尚未完全成熟。此外,由于信息是通过空间结构编码的,无法达到与基于序列的加密同样的存储密度,这使得它更适合用于安全性高于数据密度要求的场景。DNA折纸技术因为其灵活的可编程性和能够通过设计精确控制结构的特性,被认为是未来信息安全和大数据存储的重要方向。当前研究集中于如何提高读写效率和准确性,以及探索更为实用的信息恢复方法。此外,随着合成生物学的迅速发展,预计将来会开发出更多创新的DNA加密策略。

4 结论与展望

未来随着硬件技术、编码理论和存储应用的不断拓展,DNA存储有望成为未来最有前景的信息存储技术之一。本文首先介绍了DNA信息存储的基本流程,重点综述了DNA信息存储中涉及到的关键技术,尤其是编码技术、纠错技术、随机访问和DNA信息加密技术的研究进展。在编码技术方面,不仅总结了传统的DNA编码方法,还重点讨论了近期出现的新型编码策略,这些新兴技

术在提高信息密度和读取效率方面展现出巨大的潜力,为DNA存储的实际应用提供了新的思路。在纠错部分,深入探讨了当前最先进的DNA纠错技术,对这些技术进行了系统的比较与分析,指出了它们在实际应用中的优势和局限性,并讨论了如何将不同的纠错策略与编码方法相结合,以实现更高的稳定性和准确性。本文关注了最新的随机访问方法,如利用CRISPR系统进行精确定位和提取,以及DNA纳米结构辅助的存取技术。这些方法有望实现DNA存储的快速、准确随机访问,克服了传统串行读取的限制。在信息安全日益重要的背景下,本文深入分析了DNA存储中信息加密的最新技术,并关注了目前对抗量子计算攻击的新型加密算法、隐写术和基于DNA序列复杂性的动态加密方案。本文分析了这些加密技术在数据传输和存储安全性方面的潜力,并提出了如何将生物学特性与传统加密技术相结合的创新思路,这为提升DNA存储的安全性提供了新的策略。由于DNA合成成本昂贵,因此编码方案的选择对DNA信息存储系统的设计尤为重要,在设计编码方案时,应该考虑的重要方向之一是信息密度,它是衡量每个核苷酸可以存储的比特信息数,有许多因素都会影响信息存储密度,如向包含原始数据的寡核苷酸中添加额外信息会导致信息密度总体下降等,编码方案的设计对信息密度的大小起着决定性的作用。纠错是数据编码和解码阶段不可或缺的功能之一,因为DNA的合成和测序都容易出错,这些错误是由于链断裂或丢失以及链内发生插入、缺失和替换而产生的。除此以外,均聚物和高GC含量会导致测序困难和合成产量低。在设计DNA信息存储系统时,应该满足生物化学约束限制。此外,随机访问能力也是开发存储介质的一个重要考虑因素,在海量存储数据面前,随机访问能够快速定位并读取目标数据信息,大大提高了读取效率。在长期存储数据的过程中,由于数据可能被复制或者在传输过程中被拦截,篡改数据的风险大大增加,DNA信息加密技术就显得尤为重要,加密技术不仅可以保护数据的机密性,还可以通过加密算法确保数据的完整性,使得任何篡改都能被检测到。

DNA信息存储因其极高的存储密度促使大量

的科研人员和投资机构寻求先进的方法和实验。尽管目前DNA信息存储取得了巨大的进步,但是仍然无法与传统的存储技术竞争,需要克服重大挑战,特别是在写入速度、随机访问和成本方面。首先,目前DNA信息存储工作流程中的一个瓶颈是合成成本高,合成时间长^[25]。DNA合成的成本目前约为每个碱基0.10~0.30美元。这意味着以1比特/碱基的密度存储1GB的DNA数据将花费8亿~20亿美元,这一成本要高很多^[109-110]。根据Goldman等和其他人的研究^[4],DNA合成的成本在未来几年应该会下降几倍,但目前尚未看到重大进展。虽然存储的数据的大小已显著增加,但目前的DNA信息存储的记录仍在200MB左右,单次合成运行持续约24h^[1]。为了实现TB级数据的存储,人们正在大力开发编码方案、读写过程和存储程序^[111]。DNA信息存储有较高的存储密度,但是在实际存储保存过程中,还要考虑温度和湿度等环境因素,来确保DNA存储的稳定性和完整性。其次,DNA存储技术不允许在编码后进行更改或重写^[112],实现高效、快速的随机存取将是DNA存储走向实际应用必须克服的挑战之一。目前的随机存取时间与其他存储技术相比仍然没有优势,随机存取速度很慢,仅进行一次PCR就需要超过1h。数据随机存取速度慢是DNA数据存储系统的主要瓶颈之一,目前还无法将随机存取时间从几小时缩短到几分钟甚至几秒。在生物研究中,微流控技术被广泛认为是具有多种应用的强大平台,有望解决这些问题^[113]。此外,在PCR过程中,由于扩增引物的序列偏好性,DNA序列中可能会出现扩增偏差^[114-115]。如果在PCR的早期阶段发生碱基错误(替换、插入和缺失),则它会呈指数级扩增直至反应结束,因此,在DNA存储过程中还需要精确的随机访问技术。对于数据读取,DNA测序正在迅速发展,但目前的DNA测序技术都需要聚合酶等分子机器,这为每种酶的吞吐量设定了基本限制,意味着即使进行大规模并行化,测序速率也存在上限。

目前,大多数的DNA存储步骤都依赖于实验人员的参与,展望未来,自动化数字信息的写入和读取过程会大大提高DNA存储的可移植性和可扩展性,这有助于推动DNA存储商业化发展。在

随机访问方面,微流控芯片的相关开发有望实现DNA存储信息的快速、准确读取。此外,DNA存储的另外一个优势是并行性,可以并行使用数百万个细胞和数十亿个分子^[116]。DNA存储的应用在DNA计算的兴起中也可以看到,DNA计算通过利用DNA杂交过程来实现并行计算^[117],例如逻辑门^[118]和神经网络^[119]。研究人员已经证明,杂交反应可以形成DNA链置换的复杂级联反应^[120],从而改变DNA拓扑结构,通过DNA拓扑结构的修改形式编码信息,打开了新的计算可能性。用于信息存储的DNA纳米结构在密码学、隐写术和其他领域也会有更多机会^[121-122]。DNA信息存储与神经网络等数据分析技术的结合也将为越来越多的领域带来更多机遇^[123-124]。此外,提高可扩展性和可移植性是实现DNA数据存储在日常生活中广泛应用的重要方面^[125]。总之,DNA信息存储是一个非常具有前途的新研究领域,在未来,随着高通量DNA合成和测序技术的发展,DNA信息存储系统的存储密度和读取速度将进一步得到提升,数据写入和数据读取时间将大大缩短。然而,要保证DNA存储数据的快速、准确,还需要在编码算法等多方面取得突破。

参 考 文 献

- [1] BORNHOLT J, LOPEZ R, CARMEAN D M, et al. A DNA-based archival storage system[C/OL]//Proceedings of the Twenty-First International Conference on Architectural Support for Programming Languages and Operating Systems. New York, NY, USA: Association for Computing Machinery, 2016: 637-649. (2016-03-25)[2024-06-25]. <https://dl.acm.org/doi/10.1145/2872362.2872397>.
- [2] CHURCH G M, GAO Y, KOSURI S. Next-generation digital information storage in DNA[J]. *Science*, 2012, 337(6102): 1628.
- [3] GRASS R N, HECKEL R, PUDDU M, et al. Robust chemical preservation of digital information on DNA *in silica* with error-correcting codes[J]. *Angewandte Chemie International Edition*, 2015, 54(8): 2552-2555.
- [4] GOLDMAN N, BERTONE P, CHEN S Y, et al. Towards practical, high-capacity, low-maintenance information storage in synthesized DNA[J]. *Nature*, 2013, 494(7435): 77-80.
- [5] KOSURI S, CHURCH G M. Large-scale *de novo* DNA synthesis: technologies and applications[J]. *Nature Methods*, 2014, 11(5): 499-507.
- [6] LEE H H, KALHOR R, GOELA N, et al. Terminator-free template-independent enzymatic DNA synthesis for digital information storage[J]. *Nature Communications*, 2019, 10(1): 2383.
- [7] PALLUK S, ARLOW D H, DE ROND T, et al. *De novo* DNA synthesis using polymerase-nucleotide conjugates[J]. *Nature Biotechnology*, 2018, 36(7): 645-650.
- [8] ZHIRNOV V, ZADEGAN R M, SANDHU G S, et al. Nucleic acid memory[J]. *Nature Materials*, 2016, 15: 366-370.
- [9] HEINIS T, SOKOLOVSKII R, ALNASIR J J. Survey of information encoding techniques for DNA[J]. *ACM Computing Surveys*, 2024, 56(4): 1-30.
- [10] YU M, TANG X H, LI Z H, et al. High-throughput DNA synthesis for data storage[J]. *Chemical Society Reviews*, 2024, 53(9): 4463-4489.
- [11] TAN X, GE L Q, ZHANG T Z, et al. Preservation of DNA for data storage[J]. *Russian Chemical Reviews*, 2021, 90(2): 280-291.
- [12] LEE H, WIEGAND D J, GRISWOLD K, et al. Photon-directed multiplexed enzymatic DNA synthesis for molecular digital data storage[J]. *Nature Communications*, 2020, 11(1): 5246.
- [13] KUBISTA M, ANDRADE J M, BENGTSSON M, et al. The real-time polymerase chain reaction[J]. *Molecular Aspects of Medicine*, 2006, 27(2-3): 95-125.
- [14] SHENDURE J, BALASUBRAMANIAN S, CHURCH G M, et al. DNA sequencing at 40: past, present and future[J]. *Nature*, 2017, 550(7676): 345-353.
- [15] MEISER L C, NGUYEN B H, CHEN Y J, et al. Synthetic DNA applications in information technology[J]. *Nature Communications*, 2022, 13(1): 352.
- [16] CEZE L, NIVALA J, STRAUSS K. Molecular digital data storage using DNA[J]. *Nature Reviews Genetics*, 2019, 20(8): 456-466.
- [17] WELZEL M, SCHWARZ P M, LÖCHEL H F, et al. DNA-Aeon provides flexible arithmetic coding for constraint adherence and error correction in DNA storage[J]. *Nature Communications*, 2023, 14(1): 628.
- [18] LÖCHEL H F, WELZEL M, HATTAB G, et al. Fractal construction of constrained code words for DNA storage systems[J]. *Nucleic Acids Research*, 2022, 50(5): e30.
- [19] GAO X, LEPROUST E, ZHANG H, et al. A flexible light-directed DNA chip synthesis gated by deprotection using solution photogenerated acids[J]. *Nucleic Acids Research*, 2001, 29(22): 4744-4750.
- [20] ORGANICK L, ANG S D, CHEN Y J, et al. Random access in large-scale DNA data storage[J]. *Nature Biotechnology*, 2018, 36: 242-248.

- [21] BÖGELS B W A, NGUYEN B H, WARD D, et al. DNA storage in thermoresponsive microcapsules for repeated random multiplexed data access[J]. *Nature Nanotechnology*, 2023, 18(8): 912-921.
- [22] HAO M, QIAO H Y, GAO Y M, et al. A mixed culture of bacterial cells enables an economic DNA storage on a large scale[J]. *Communications Biology*, 2020, 3(1): 416.
- [23] CHEN W G, HAN M Z, ZHOU J T, et al. An artificial chromosome for data storage[J]. *National Science Review*, 2021, 8(5): nwab028.
- [24] SUN F J, DONG Y M, NI M, et al. Mobile and self-sustained data storage in an extremophile genomic DNA[J]. *Advanced Science*, 2023, 10(10): e2206201.
- [25] DONG Y M, SUN F J, PING Z, et al. DNA storage: research landscape and future prospects[J]. *National Science Review*, 2020, 7(6): 1092-1107.
- [26] HOSSEIN TABATABAEI YAZDI S M, GABRYS R, MILENKOVIC O. Portable and error-free DNA-based data storage[J]. *Scientific Reports*, 2017, 7(1): 5011.
- [27] EL-SHAIKH A, WELZEL M, HEIDER D, et al. High-scale random access on DNA storage systems[J]. *NAR Genomics and Bioinformatics*, 2022, 4(1): lqab126.
- [28] CHOI Y, BAE H J, LEE A C, et al. DNA micro-disks for the management of DNA-based data storage with index and write-once-read-many (WORM) memory features[J]. *Advanced Materials*, 2020, 32(37): e2001249.
- [29] NEWMAN S, STEPHENSON A P, WILLSEY M, et al. High density DNA data storage library *via* dehydration with digital microfluidic retrieval[J]. *Nature Communications*, 2019, 10(1): 1706.
- [30] BANAL J L, SHEPHERD T R, BERLEANT J, et al. Random access DNA memory using Boolean search in an archival file storage system[J]. *Nature Materials*, 2021, 20(9): 1272-1280.
- [31] XU C T, MA B, GAO Z L, et al. Electrochemical DNA synthesis and sequencing on a single electrode with scalability for integrated data storage[J]. *Science Advances*, 2021, 7(46): eabk0100.
- [32] BENTLEY D R, BALASUBRAMANIAN S, SWERDLOW H P, et al. Accurate whole human genome sequencing using reversible terminator chemistry[J]. *Nature*, 2008, 456(7218): 53-59.
- [33] ROSS M G, RUSS C, COSTELLO M, et al. Characterizing and measuring bias in sequence data[J]. *Genome Biology*, 2013, 14(5): R51.
- [34] SCHWARTZ J J, LEE C, SHENDURE J. Accurate gene synthesis with tag-directed retrieval of sequence-verified DNA molecules[J]. *Nature Methods*, 2012, 9(9): 913-915.
- [35] SHANNON C E. A mathematical theory of communication[J]. *The Bell System Technical Journal*, 1948, 27(3): 379-423.
- [36] REN Y B, ZHANG Y, LIU Y W, et al. DNA-based concatenated encoding system for high-reliability and high-density data storage[J]. *Small Methods*, 2022, 6(4): e2101335.
- [37] ERLICH Y, ZIELINSKI D. DNA fountain enables a robust and efficient storage architecture[J]. *Science*, 2017, 355(6328): 950-954.
- [38] ANAVY L, VAKNIN I, ATAR O, et al. Data storage in DNA with fewer synthesis cycles using composite DNA letters[J]. *Nature Biotechnology*, 2019, 37(10): 1229-1236.
- [39] WANG K, CAO B, MA T, et al. Storing images in DNA *via* base128 encoding[J]. *Journal of Chemical Information and Modeling*, 2024, 64(5): 1719-1729.
- [40] DICKINSON G D, MORTUZA G M, CLAY W, et al. An alternative approach to nucleic acid memory[J]. *Nature Communications*, 2021, 12(1): 2371.
- [41] ZHENG Y F, CAO B, ZHANG X K, et al. DNA-QLC: an efficient and reliable image encoding scheme for DNA storage [J]. *BMC Genomics*, 2024, 25(1): 266.
- [42] ZHENG Y F, CAO B, WU J Q, et al. High net information density DNA data storage by the MOPE encoding algorithm[J]. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2023, 20(5): 2992-3000.
- [43] CHU H, WANG C, ZHANG Y. Improved constructions of secondary structure avoidance codes for DNA sequences[EB/OL]. *arXiv*, 2023: 2304.11403v1. (2023-04-22) [2024-06-25]. <https://doi.org/10.48550/arXiv.2304.11403>.
- [44] PARK S J, LEE Y W, NO J S. Iterative coding scheme satisfying GC balance and run-length constraints for DNA storage with robustness to error propagation[J]. *Journal of Communications and Networks*, 2022, 24(3): 283-291.
- [45] PING Z, CHEN S H, ZHOU G Y, et al. Towards practical and robust DNA-based data archiving using the Yin-Yang codec system[J]. *Nature Computational Science*, 2022, 2(4): 234-242.
- [46] LU X Z, KIM S H. Weakly mutually uncorrelated codes with maximum run length constraint for DNA storage[J]. *Computers in Biology and Medicine*, 2023, 165: 107439.
- [47] WANG Y X, NOOR-A-RAHIM M, ZHANG J Y, et al. High capacity DNA data storage with variable-length oligonucleotides using repeat accumulate code and hybrid mapping[J]. *Journal of Biological Engineering*, 2019, 13: 89.
- [48] WANG Y X, NOOR-A-RAHIM M D, GUNAWAN E, et al. Construction of bio-constrained code for DNA data storage[J]. *IEEE Communications Letters*, 2019, 23(6): 963-966.
- [49] XAVIER KOHLL A, ANTKOWIAK P L, CHEN W D, et al. Stabilizing synthetic DNA for long-term data storage with earth alkaline salts[J]. *Chemical Communications*, 2020, 56(25): 3613-3616.
- [50] HECKEL R, MIKUTIS G, GRASS R N. A characterization of the DNA data storage channel[J]. *Scientific Reports*, 2019, 9

- (1): 9663.
- [51] SUN Y, HAN G J, LIU C, et al. An asymmetric-error-aware LDPC decoding algorithm for DNA storage[J]. IEEE Communications Letters, 2023, 27(1): 32-36.
 - [52] ANTKOWIAK P L, LIETARD J, DARESTANI M Z, et al. Low cost DNA data storage using photolithographic synthesis and advanced information reconstruction and error correction [J]. Nature Communications, 2020, 11(1): 5345.
 - [53] DING L L, WU S G, HOU Z H, et al. Improving error-correcting capability in DNA digital storage *via* soft-decision decoding[J]. National Science Review, 2024, 11(2): nwad229.
 - [54] LU X Z, JEONG J, KIM J W, et al. Error rate-based log-likelihood ratio processing for low-density parity-check codes in DNA storage[J]. IEEE Access, 2020, 8: 162892-162902.
 - [55] FEI P, WANG Z Y. LDPC codes for portable DNA storage[C/OL]//2019 IEEE International Symposium on Information Theory (ISIT). July 7-12, 2019, Paris, France. IEEE, 2019: 76-80. (2019-09-26) [2024-06-25]. <https://ieeexplore.ieee.org/document/8849814>.
 - [56] SUN J F, PHILPOTT M, LOI D, et al. Correcting PCR amplification errors in unique molecular identifiers to generate accurate numbers of sequencing molecules[J]. Nature Methods, 2024, 21(3): 401-405.
 - [57] CHEN K K, ZHU J B, BOŠKOVIĆ F, et al. Nanopore-based DNA hard drives for rewritable and secure data storage[J]. Nano Letters, 2020, 20(5): 3754-3760.
 - [58] MEISER L C, ANTKOWIAK P L, KOCH J, et al. Reading and writing digital data in DNA[J]. Nature Protocols, 2020, 15(1): 86-101.
 - [59] CHANDAK S, TATWAWADI K, LAU B, et al. Improved read/write cost tradeoff in DNA-based data storage using LDPC codes[C/OL]//2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton). September 24-27, 2019, Monticello, IL, USA. IEEE, 2019: 147-156. (2019-12-05) [2024-06-25]. <https://ieeexplore.ieee.org/document/8919890>.
 - [60] CAO B, ZHANG X K, CUI S, et al. Adaptive coding for DNA storage with high storage density and low coverage[J]. NPJ Systems Biology and Applications, 2022, 8(1): 23.
 - [61] ZHANG S F, WU J J, HUANG B B, et al. High-density information storage and random access scheme using synthetic DNA[J]. 3 Biotech, 2021, 11(7): 328.
 - [62] RANU N, VILLANI A C, HACHOEN N, et al. Targeting individual cells by barcode in pooled sequence libraries[J]. Nucleic Acids Research, 2019, 47(1): e4.
 - [63] MATANGE K, TUCK J M, KEUNG A J. DNA stability: a central design consideration for DNA data storage systems[J]. Nature Communications, 2021, 12(1): 1358.
 - [64] BORNHOLT J, LOPEZ R, CARMEAN D M, et al. Toward a DNA-based archival storage system[J]. IEEE Micro, 2017, 37(3): 98-104.
 - [65] LAU B, CHANDAK S, ROY S, et al. Magnetic DNA random access memory with nanopore readouts and exponentially-scaled combinatorial addressing[J]. Scientific Reports, 2023, 13(1): 8514.
 - [66] HOSSEIN TABATABAEI YAZDI S M, YUAN Y B, MA J, et al. A rewritable, random-access DNA-based storage system[J]. Scientific Reports, 2015, 5: 14138.
 - [67] LIMBACHIYA D, GUPTA M K, AGGARWAL V. Family of constrained codes for archival DNA data storage[J]. IEEE Communications Letters, 2018, 22(10): 1972-1975.
 - [68] CAO B, LI X, ZHANG X K, et al. Designing uncorrelated address constrain for DNA storage by DMVO algorithm[J]. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2022, 19(2): 866-877.
 - [69] WENGER A M, PELUSO P, ROWELL W J, et al. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome[J]. Nature Biotechnology, 2019, 37(10): 1155-1162.
 - [70] YIN Q, ZHENG Y F, WANG B, et al. Design of constraint coding sets for archive DNA storage[J]. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2022, 19(6): 3384-3394.
 - [71] YIN Q, CAO B, LI X, et al. An intelligent optimization algorithm for constructing a DNA storage code: NOL-HHO[J]. International Journal of Molecular Sciences, 2020, 21(6): 2191.
 - [72] CAO B, ZHANG X K, WU J Q, et al. Minimum free energy coding for DNA storage[J]. IEEE Transactions on Nanobioscience, 2021, 20(2): 212-222.
 - [73] RASOOL A, QU Q, JIANG Q S, et al. A strategy-based optimization algorithm to design codes for DNA data storage system[M/OL]// LAI Y X, WANG T, JIANG M, et al. Algorithms and architectures for parallel processing: lecture notes in computer science. Cham: Springer International Publishing, 2022: 284-299. (2022-02-23) [2024-06-28]. https://doi.org/10.1007/978-3-030-95388-1_19.
 - [74] RASOOL A, QU Q, WANG Y, et al. Bio-constrained codes with neural network for density-based DNA data storage[J]. Mathematics, 2022, 10(5): 845.
 - [75] CAO B, WANG B, ZHANG Q. GCNSA: DNA storage encoding with a graph convolutional network and self-attention [J]. iScience, 2023, 26(3): 106231.
 - [76] WU J Q, ZHENG Y F, WANG B, et al. Enhancing physical and thermodynamic properties of DNA storage sets with end-constraint[J]. IEEE Transactions on Nanobioscience, 2022, 21(2): 184-193.
 - [77] SHIPMAN S L, NIVALA J, MACKLIS J D, et al. CRISPR-Cas encoding of a digital movie into the genomes of a

- population of living bacteria[J]. *Nature*, 2017, 547(7663): 345-349.
- [78] ZHANG J Y, HOU C Y, LIU C C. CRISPR-powered quantitative keyword search engine in DNA data storage[J]. *Nature Communications*, 2024, 15(1): 2376.
- [79] BENCUROVA E, AKASH A, DOBSON R C J, et al. DNA storage-from natural biology to synthetic biology[J]. *Computational and Structural Biotechnology Journal*, 2023, 21: 1227-1235.
- [80] FAN C Y, DENG Q, ZHU T F. Bioorthogonal information storage in L-DNA with a high-fidelity mirror-image *Pfu* DNA polymerase[J]. *Nature Biotechnology*, 2021, 39(12): 1548-1555.
- [81] ZHANG J L, CHEN M Y, LIN G H, et al. Advanced DNA amplification for efficient data storage[J]. *ACS Applied Materials & Interfaces*, 2024, 16(37): 48870-48879.
- [82] ANDERSON R, MOORE T. The economics of information security[J]. *Science*, 2006, 314(5799): 610-613.
- [83] BRUNET T D P. Aims and methods of biosteganography[J]. *Journal of Biotechnology*, 2016, 226: 56-64.
- [84] DE SILVA P Y, GANEGODA G U. New trends of digital data storage in DNA[J]. *BioMed Research International*, 2016, 2016: 8072463.
- [85] CUI M Y, ZHANG Y X. Advancing DNA steganography with incorporation of randomness[J]. *Chembiochem*, 2020, 21(17): 2503-2511.
- [86] CLELLAND C T, RISCA V, BANCROFT C. Hiding messages in DNA microdots[J]. *Nature*, 1999, 399(6736): 533-534.
- [87] GEHANI A, LABEAN T, REIF J. DNA-based cryptography [M/OL]//JONOSKA N, PÄUN G, ROZENBERG G. Aspects of molecular computing: essays dedicated to Tom Head, on the Occasion of His 70th Birthday. Berlin, Heidelberg: Springer, 2003, 2950: 167-188[2024-06-25]. https://doi.org/10.1007/978-3-540-24635-0_12.
- [88] LAI X J, LU M X, QIN L, et al. Asymmetric encryption and signature method with DNA technology[J]. *Science China Information Sciences*, 2010, 53(3): 506-514.
- [89] LU M X, LAI X J, XIAO G Z, et al. Symmetric-key cryptosystem with DNA technology[J]. *Science in China Series F: Information Sciences*, 2007, 50(3): 324-333.
- [90] GRASS R N, HECKEL R, DESSIMOZ C, et al. Genomic encryption of digital data stored in synthetic DNA[J]. *Angewandte Chemie International Edition*, 2020, 59(22): 8476-8480.
- [91] YAO X Y, XIE R Z, ZAN X Z, et al. A novel image encryption scheme for DNA storage systems based on DNA hybridization and gene mutation[J]. *Interdisciplinary Sciences, Computational Life Sciences*, 2023, 15(3): 419-432.
- [92] MARRAS A E, ZHOU L F, SU H J, et al. Programmable motion of DNA origami mechanisms[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2015, 112(3): 713-718.
- [93] ROTHEMUND P W K. Folding DNA to create nanoscale shapes and patterns[J]. *Nature*, 2006, 440(7082): 297-302.
- [94] VENEZIANO R, RATANALERT S, ZHANG K M, et al. Designer nanoscale DNA assemblies programmed from the top down[J]. *Science*, 2016, 352(6293): 1534.
- [95] BENSON E, MOHAMMED A, GARDELL J, et al. DNA rendering of polyhedral meshes at the nanoscale[J]. *Nature*, 2015, 523(7561): 441-444.
- [96] DOUGLAS S M, DIETZ H, LIEDL T, et al. Self-assembly of DNA into nanoscale three-dimensional shapes[J]. *Nature*, 2009, 459(7245): 414-418.
- [97] ZHANG Y N, CHAO J, LIU H J, et al. Transfer of two-dimensional oligonucleotide patterns onto stereocontrolled plasmonic nanostructures through DNA-origami-based nanoimprinting lithography[J]. *Angewandte Chemie International Edition*, 2016, 55(28): 8036-8040.
- [98] WANG D F, VIETZ C, SCHRÖDER T, et al. A DNA walker as a fluorescence signal amplifier[J]. *Nano Letters*, 2017, 17(9): 5368-5374.
- [99] KNUDSEN J B, LIU L, BANK KODAL A L, et al. Routing of individual polymers in designed patterns[J]. *Nature Nanotechnology*, 2015, 10(10): 892-898.
- [100] LANGECKER M, ARNAUT V, MARTIN T G, et al. Synthetic lipid membrane channels formed by designed DNA nanostructures[J]. *Science*, 2012, 338(6109): 932-936.
- [101] VOIGT N V, TØRRING T, ROTARU A, et al. Single-molecule chemical reactions on DNA origami[J]. *Nature Nanotechnology*, 2010, 5(3): 200-203.
- [102] YANG J, MA J J, LIU S, et al. A molecular cryptography model based on structures of DNA self-assembly[J]. *Chinese Science Bulletin*, 2014, 59(11): 1192-1198.
- [103] ZHANG Y N, WANG F, CHAO J, et al. DNA origami cryptography for secure communication[J]. *Nature Communications*, 2019, 10(1): 5469.
- [104] FAN S S, WANG D F, CHENG J, et al. Information coding in a reconfigurable DNA origami domino array[J]. *Angewandte Chemie*, 2020, 132(31): 13091-13097.
- [105] ZHU J B, ERMANN N, CHEN K K, et al. Image encoding using multi-level DNA barcodes with nanopore readout[J]. *Small*, 2021, 17(28): e2100711.
- [106] SIDDARAMAPPA V, RAMESH K B. DNA-based XOR operation (DNAX) for data security using DNA as a storage medium[M/OL]//KRISHNAA, SRIKANTIAH K, NAVEENA C. Integrated intelligent computing, communication and security: studies in computational intelligence. Singapore: Springer Singapore, 2019, 771: 343-351. (2028-09-15) [2024-03-01]. https://doi.org/10.1007/978-981-10-8797-4_36.

- [107] ZHU E Q, LUO X H, LIU C J, et al. An operational DNA strand displacement encryption approach[J]. *Nanomaterials*, 2022, 12(5): 877.
- [108] TENG Y, YANG S, LIU L Y, et al. Nanoscale storage encryption: data storage in synthetic DNA using a cryptosystem with a neural network[J]. *Science China Life Sciences*, 2022, 65(8): 1673-1676.
- [109] FONTANA R E JR, DECAD G M. Moore's law realities for recording systems and memory storage components: HDD, tape, NAND, and optical[J]. *AIP Advances*, 2017, 8(5): 056506.
- [110] HUGHES R A, ELLINGTON A D. Synthetic DNA synthesis and assembly: putting the synthetic in synthetic biology[J]. *Cold Spring Harbor Perspectives in Biology*, 2017, 9(1): a023812.
- [111] BLAWAT M, GAEDKE K, HÜTTER I, et al. Forward error correction for DNA data storage[J]. *Procedia Computer Science*, 2016, 80: 1011-1022.
- [112] ZHOU Y, BI K, GE Q Y, et al. Advances and challenges in random access techniques for *in vitro* DNA data storage[J]. *ACS Applied Materials & Interfaces*, 2024, 16(33): 43102-43113.
- [113] LUO Y, CAO Z, LIU Y F, et al. The emerging landscape of microfluidic applications in DNA data storage[J]. *Lab on a Chip*, 2023, 23(8): 1981-2004.
- [114] LI X Y, CHEN M X, WU H M. Multiple errors correction for position-limited DNA sequences with GC balance and no homopolymer for DNA-based data storage[J]. *Briefings in Bioinformatics*, 2023, 24(1): bbac484.
- [115] PAN W J, BYRNE-STEELE M, WANG C L, et al. DNA polymerase preference determines PCR priming efficiency[J]. *BMC Biotechnology*, 2014, 14: 10.
- [116] AKASH A, BENCUROVA E, DANDEKAR T. How to make DNA data storage more applicable[J]. *Trends in Biotechnology*, 2024, 42(1): 17-30.
- [117] YANG S, BÖGELS B W A, WANG F, et al. DNA as a universal chemical substrate for computing and data storage[J]. *Nature Reviews Chemistry*, 2024, 8(3): 179-194.
- [118] ZHANG D Y, SEELIG G. Dynamic DNA nanotechnology using strand-displacement reactions[J]. *Nature Chemistry*, 2011, 3(2): 103-113.
- [119] CHERRY K M, QIAN L L. Scaling up molecular pattern recognition with DNA-based winner-take-all neural networks[J]. *Nature*, 2018, 559(7714): 370-376.
- [120] CHEN C Z, WEN J D, WEN Z B, et al. DNA strand displacement based computational systems and their applications[J]. *Frontiers in Genetics*, 2023, 14: 1120791.
- [121] ZHANG Y Y, LI F, LI M, et al. Encoding carbon nanotubes with tubular nucleic acids for information storage[J]. *Journal of the American Chemical Society*, 2019, 141(44): 17861-17866.
- [122] COUDY D, COLOTTE M, LUIS A, et al. Long term conservation of DNA at ambient temperature. Implications for DNA data storage[J]. *PLoS One*, 2021, 16(11): e0259868.
- [123] BEE C, CHEN Y J, QUEEN M, et al. Molecular-level similarity search brings computing to DNA data storage[J]. *Nature Communications*, 2021, 12(1): 4764.
- [124] QIN Y, ZHU F, XI B, et al. Robust multi-read reconstruction from noisy clusters using deep neural network for DNA storage[J]. *Computational and Structural Biotechnology Journal*, 2024, 23: 1076-1087.
- [125] WANG S, MAO X, WANG F, et al. Data storage using DNA[J]. *Advanced Materials*, 2024, 36(6): e2307499.



通讯作者：徐苗苗(1992—)，女，师资博士后。研究方向为合成生物学及DNA存储。

E-mail: miaomiaoxu@gzucm.edu.cn



第一作者：徐怀胜(1996—)，男，硕士，助教。研究方向为DNA信息存储。

E-mail: hsxu@gcc.edu.cn