

DNA 存储技术的中美对比研究

袁 芳, 郑彦宁, 郑 佳, 李 稔, 傅俊英

(中国科学技术信息研究所, 北京 100038)

摘 要: DNA 存储作为新兴技术, 与传统的数据存储方法相比, 具有超强的存储容纳性、稳定性和安全性, 在海量数据存储、机密数据存储与传递中具有巨大应用前景。本文从论文发表情况、专利申请情况、相关政策以及重点企业 4 个方面, 对比研究了中国与美国在 DNA 存储技术领域上的差异。最后, 结合中国的实际情况, 为中国 DNA 存储技术的发展提出了建议。

关键词: DNA 存储; 技术分析; 对比研究

中图分类号: G306 **文献标识码:** A **DOI:** 10.3772/j.issn.1009-8623.2019.04.011

提起信息数据存储的媒介, 人们一般想到硬盘、光盘、U 盘, 或者早先的磁带、录像带等等。如今, 人类产生越来越多的信息, 数据存储也变得越来越困难。2018 年, 全球的数据量为 3.52×10^{22} 位^[1], 到 2040 年全世界的数据总量将达到 3×10^{24} 位^[1, 2]。随着大数据、互联网、人工智能的发展, 这个数字将会继续增长并无逆转迹象。传统数据存储方法存在诸多问题: 首先, 存储海量的数据资源, 需要占用巨大空间资源; 其次, 海量数据存储中心的运行, 需要耗费巨大电力资源来保持其恒温恒湿^[3]; 最后, 硅元素在自然界存在有限, 会有消耗殆尽的时候^[2]。近年来, 大自然最古老的、迄今容量最大的存储媒介——DNA, 被广泛关注并成为一种具有吸引力的潜在数据存储介质^[4]。

1 DNA 存储技术的概念

DNA 存储技术就是以 DNA 分子为存储介质, 将信息存储于 DNA 分子上, 从而模拟存储器的数据读取和写入操作^[5]。DNA 作为一种大自然的

遗传信息存储媒介, 已沿用了 35 亿年^[6]。DNA 是由腺嘌呤(A)、胸腺嘧啶(T)、胞嘧啶(C)和鸟嘌呤(G) 4 个碱基构成双螺旋结构, 来保存生物体的遗传信息^[7]。因此, 将数据文件编码到碱基中, 使用各种算法, 可以形成非常密集又节省空间的 DNA 数据存储器。与传统的数据存储方法相比, DNA 存储具有超强的存储容纳性、稳定性和安全性^[8, 9], 有望解决大数据时代传统数据存储难以满足的需求, 成为海量数据存储的颠覆性技术。

计算机系统数据是采用二进制来表示的, 即每位只有 0 和 1 两种形式。DNA 存储使用的新型算法, 不再是传统的计算机二进制计算模式。要将计算机中的信息存入 DNA 中, 实际上就是把计算机中的二进制数据流, 转换为 DNA 中的碱基序列编码的数据存储计算模式。例如, 可以使用二进制, 用 0 代表碱基 A 或 C, 1 代表碱基 T 或 G^[10]; 也可以使用三进制, 将数据编码存储来降低出错率^[11]; 或者使用其他将数字编码转化为碱基化学编码的数据存储计算模式^[12]。

第一作者简介: 袁芳 (1989—), 女, 博士, 主要研究方向为产业竞争情报研究。

项目来源: 青海省“互联网+”绿色产业发展 (2018-ZJ-614)、博士后基金第 61 批之“颠覆性技术的识别方法研究” (2017M610970)。

收稿日期: 2019-03-21

DNA 存储的数据写入采用 DNA 合成技术, 合成特定碱基序列的 DNA 分子。数据读取采用 DNA 测序技术, 解读碱基序列的编码信息, DNA 存储技术的全过程具体如图 1 所示^[3]。存储数据的 DNA 分子可以导入细菌等生物体内, 也可以存放于干净的储存容器中进行保存。

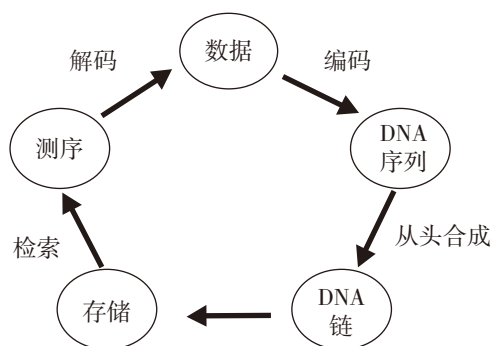


图 1 DNA 存储涉及的步骤

2 DNA存储技术的中美对比情况

2.1 论文发表情况对比

从 Web of science 数据库中的“Science Citation Index Expanded (SCI-EXPANDED)”检索 1900 年至 2018 年有关 DNA 存储技术的 ARTICLE 文献类型, 共有 225 篇论文, 检索日期为 2019 年 1 月 3 日。其中, 来自美国的 DNA 存储技术相关论文有 92 篇,

占全球总量的 40.89%。来自中国的 DNA 存储技术论文有 25 篇, 发表数量位居全球第二。

如图 2 所示, 美国 DNA 存储技术相关论文发表数量的变化趋势与全球的变化趋势高度一致。美国在 2004 年发表 8 篇与 DNA 存储技术相关的论文, 导致全球 DNA 存储技术相关论文的发表数量在 2004 年达到一个小的峰值 12 篇。这表明, 美国在 2004 年前后十分重视 DNA 存储技术的发展。但从 2005 年开始, 美国 DNA 存储技术相关论文发表数量开始逐渐下降, 表明 DNA 存储技术在 2005 年后遇到了瓶颈。直到近两年, 美国 DNA 存储技术相关论文的发表数量又开始迅速增长, 表明美国在 DNA 存储技术领域有了新的突破, DNA 存储技术又开始了新的迅猛发展阶段。

由图可以看出, 在 DNA 存储技术领域, 中国相比美国, 起步较晚。2004 年, 中国才在 DNA 存储技术领域发表了第一篇相关论文。2017 年, 中国 DNA 存储技术相关论文的发表数量达到峰值 7 篇。2017 年以外的其他年份, 中国 DNA 存储技术相关论文的发表数量均不足 5 篇。这表明, 中国在 DNA 存储技术领域仍处在萌芽阶段, 与美国的 DNA 存储技术有一定差距。

2.2 专利申请情况对比

在 Innography 数据库共检索到 DNA 存储技术

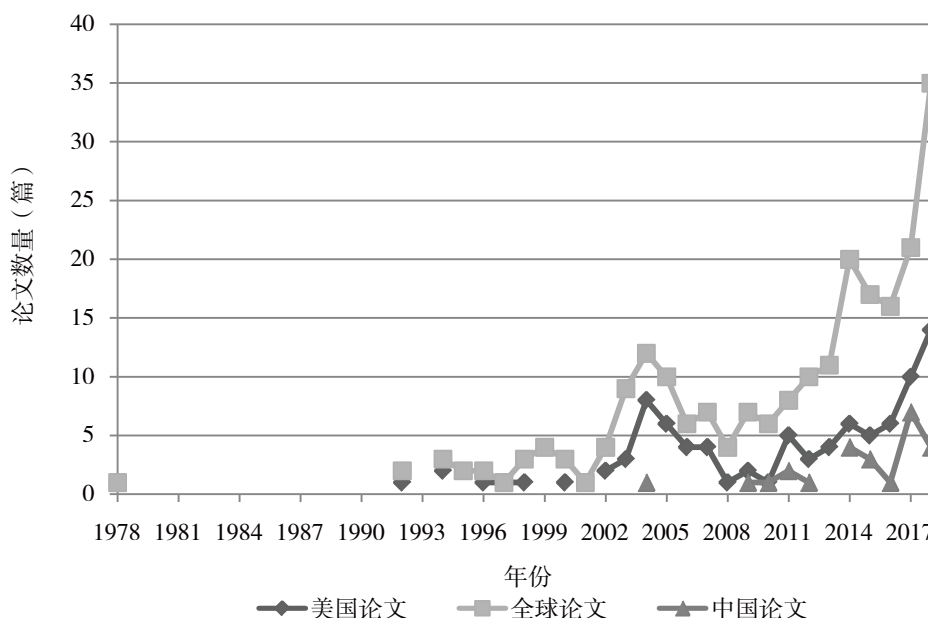


图 2 DNA 存储相关论文的逐年发表情况

相关专利 171 件, 检索日期为 2019 年 1 月 3 日。其中, 来自美国的 DNA 存储技术相关专利有 51 件, 全球排名第一; 来自中国的 DNA 存储技术相关专利有 34 件, 全球排名第二。如图 3 所示, 全球 DNA 存储技术相关专利申请数量在 2006 与 2007 年达到一个最高峰 19 篇。2013 年至 2017 年, 全球 DNA 存储技术相关专利申请数量又出现逐年上升趋势。受专利审查制度的影响, 专利从申请到公开一般会有 2~3 年的延迟^[13], 这与全球 DNA 存储技术相关论文的变化趋势大致相符。

由图 3 可以看出, 来自美国的第一件 DNA 存储技术相关专利在 2003 年申请, 在之后的 5 年内, 美国 DNA 存储技术的相关专利每年申请量都在 5 件左右。从 2008 年开始, 美国 DNA 存储技术的相关

专利每年申请量下降到 3 件以内。但在 2018 年, 美国 DNA 存储技术的相关专利数量达到峰值 10 件。这表明, 2003 年至 2007 年, 美国 DNA 存储技术在应用方面得到了短暂发展, 之后进入发展停滞期。在 2009 年及 2017 年, 美国并没有 DNA 存储技术相关专利的申请, 但在 2018 年美国又加大对 DNA 存储技术的应用研究, DNA 存储技术相关专利的申请数量突破 10 件。。

虽然中国的第一件 DNA 存储技术相关专利在 2002 年申请, 比美国的第一件相关专利早了一年, 但在 2003 年、2004 年、2010 年及 2012 年, 中国并没有 DNA 存储技术相关专利的申请。从 2015 年开始, 中国才加大了对 DNA 存储技术的应用研究。2015 年至 2018 年, 中国 DNA 存储技术相关专利每年申请

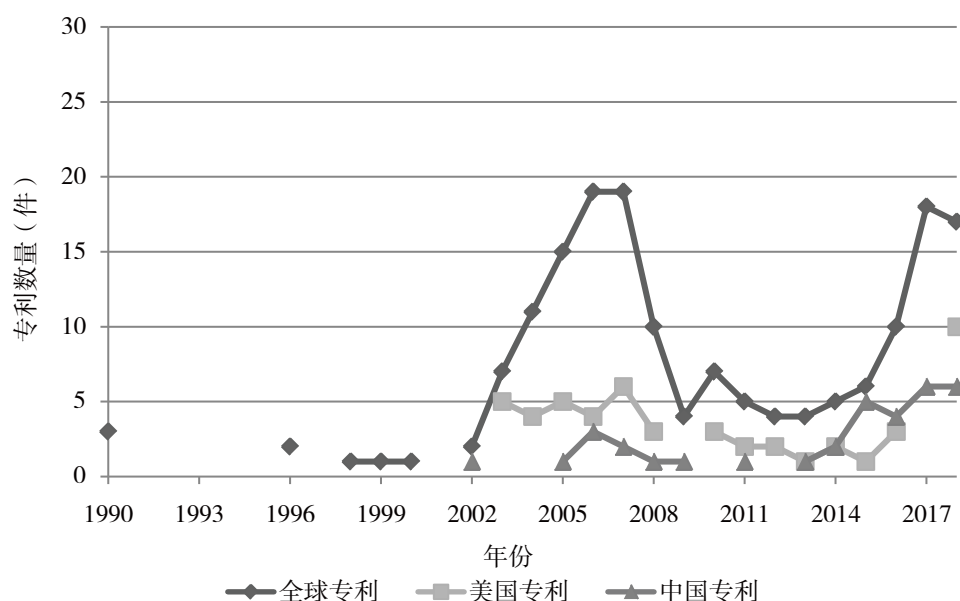


图 3 DNA 存储相关专利的逐年申请情况

量达到 5 件左右。

2.3 相关政策对比

DNA 存储的数据写入采用 DNA 合成技术, 需要人工合成编码信息的 DNA 双链。存储数据的 DNA 除可以放置于干净的特定容器中外, 还可以被导入细菌等生物体内进行存放。因此, DNA 存储的相关政策会涉及人工合成的 DNA 以及重组 DNA 的微生物等方面的相关政策。

美国在 DNA 存储技术领域有较为完善的法律指南监管以及政策资金支持。首先, 美国国立卫生

研究院 (NIH) 有研究者熟知的涉及重组 DNA 分子研究的指南。其次, 美国环境保护署、农业部、健康和人类事务部、商务部等重要部门都对合成的 DNA 以及重组 DNA 微生物的使用、销售、运输等方面做了明确的法律约束及监管^[14-16]。最后, 近两年美国政府发布了多个与 DNA 存储技术相关的项目公告, 用以解决海量数据存储问题。2017 年, 美国国防部高级研究计划局 (DARPA) 发布了“分子信息学”项目公告, 寻求开发一种全新的数据存储技术, 能够在分子和化学层面处理来自侦察、电

子战、信号情报、持续监视等数据密集型军事应用领域的大量信息流。2018年5月,美国情报高级研究计划局(IARPA)发布了“分子信息存储”项目公告,旨在研究一种系统,用于在顺序控制的聚合物(比如人类DNA)上存储艾字节(Exabyte, $1\text{EB}=10^{18}\text{B}$)量级的数据。2018年7月,美国国家科学基金会(NSF)计划投资1200万美元,并联合半导体研究公司(SRC)开展“信息处理和存储技术的半导体合成生物学”项目,创建将合成生物学与半导体技术相结合的存储系统。

目前,中国在人工合成的DNA以及重组DNA的微生物等方面的相关政策主要包括科技部出台的关于重组DNA的生物安全条例、农业部出台的遗传改造生物的安全管理条例及贸易和标识细节、环保部出台的病原微生物实验室生物安全环境管理办法以及各部门出台相关的菌种保藏管理办法^[17,18]。DNA存储技术属于新兴技术,并且还未有大规模的应用。因此,中国与美国相似,没有专门针对DNA存储技术的法律法规,只是在其涉及的相关领域(如重组DNA)遵循现有的相关法律法规。美国近两年开始大力支持与发展DNA存储技术,而中国目前还没有明确的政策与规划用以大力发展DNA存储技术。

2.4 重点企业对比

目前,DNA存储的存储成本、数据读取速度和随机存取仍在限制DNA存储的产业发展。因此,DNA存储的早期使用将仅限于特殊情况,对于需要数据紧急处理及实时处理的情况并不适用。鉴于DNA存储技术超强的存储容纳性、稳定性和安全性等优点,DNA存储技术有以下两方面不可忽视的应用前景:一是建立针对不经常访问数据的DNA存储中心,可有效节约传统数据中心的成本与资源;二是将DNA存储的加密系统与密码学知识结合来对机密信息进行存档,DNA存储的信息会更难以被破译。随着DNA数据存储成本的下降和速度的提高,以及用户能够很容易地将文件、图像甚至神经活动保存到DNA中,新的商业机会将会出现。

美国除了哈佛大学^[10,19]等高校在DNA存储技术领域做出了突破性研究外,国际巨头微软公司及众多初创企业也开始着手推进DNA存储技术产业化进程。2016年,微软利用DNA存储技术完成了约200MB($1\text{MB}=10^6\text{B}$)数据的保存,其中包括百

部经典文学作品^[20]。2018年,微软又解决了DNA存储的随机存取数据技术,标记每一个文件在DNA序列上的地址,就如同硬盘的存储路径一样^[21]。微软公司在DNA存储技术的发展方面做出了诸多国际领先成果,并计划2020年在数据中心内建立一个以DNA为基础的操作存储系统,以满足日益增长的数据存储需求。目前,微软明确了基于DNA的存储设备的发展计划,并打算未来几年实现商业化。此外,一系列初创公司也纷纷获得融资用于DNA存储技术及其相关技术的开发与应用,例如美国的Catalog公司、Molecular Assemblies公司、Twist Bioscience公司、Iridia公司及Synthomics公司等。

中国在DNA存储技术领域刚刚起步,虽然企业或高校的研究水平目前都不突出,但也有某些企业开始将目光转向DNA存储技术。苏州泓迅生物科技有限公司是一家专注于合成生物学的DNA技术公司,该公司在DNA存储技术领域申请了2件相关专利,有自主知识产权的编码转制软件,精确、快速的大规模和高通量的合成DNA序列,应用于下一代的信息存储。并且,泓迅生物拥有第三代合成生物学技术平台,能够大幅度降低DNA合成的生产成本。此外,上海吐露港生物科技有限公司与上海生命科学院于2018年在《ACS合成生物学》上共同发表了一篇关于辅助的DNA隐写术对DNA编码信息的储存和转移的安全处理的文章^[22]。该研究结果可以巧妙地利用Cas12a的trans切割特性,对DNA数据进行加密。

3 小结与建议

从论文发表数量与专利申请数量上看,美国的DNA存储技术处于全球领先水平。这是因为美国政府大力支持DNA存储技术,美国科研机构也比全球其他国家更关注DNA存储技术,该领域的多个突破性成果产出于美国。而中国目前还没有明确的政策与规划用以大力发展DNA存储技术,并且中国科研机构在该领域的关注度与研究水平仍有待提高。DNA存储作为新兴技术,在海量数据存储、机密数据存储与传递中具有巨大应用前景。提前做好对DNA存储技术的战略布局,有利于在新兴产业中提前立足、在国防安全中提前防御。为促进中国DNA存储技术的科学发展,提出以下建议。

(1) 制定 DNA 存储的相关扶持政策与发展规划, 加大政府的支持与资助。中国在 DNA 存储领域实力仍旧薄弱, DNA 存储技术作为新兴生物电子先进技术应被国家大力支持发展。中国政府应加大对 DNA 存储技术的支持与资助, 尤其是要重视 DNA 存储在海量数据存储产业化中所面临的关键技术问题, 例如 DNA 合成技术、DNA 存储快速读取技术或随机读取技术。另外, 也应增强相关领域的军民融合项目研究, 支持 DNA 存储技术在国防安全中的潜在应用项目及其相关的安全预警防御项目。

(2) 制定 DNA 存储的安全预警及安全防御战略。新兴生物技术为更好更快的治疗方案、经济发展、更清洁的环境和更高的生活质量带来了希望, 但同时也带来了新的安全风险。世界军事强国纷纷把生物安全作为国家安全战略的重要组成部分, 防范新兴生物技术对未来国家安全的潜在影响。DNA 存储技术在情报、黑客等方面可能给未来国家安全带来隐患。因此, DNA 存储作为新兴的生物技术, 应该提早做出国家安全预警布局, 减少并及时应对可能出现的安全危害。

(3) 促进多学科研究人员之间的交流, 推进多学科的相互融合。未来超低能量计算系统可能建立在化学、生物学和工程学交叉点的有机系统的原理之上。针对新兴的 DNA 存储技术, 应鼓励大学等研究机构建设生物与计算机信息科学的交叉学科研究团队。相关政府部门应做好牵头组织作用, 促进多学科研究人员的交流, 增强交叉学科基础研究, 推进多学科间相互融合。■

参考文献:

- [1] Panda D, Molla K A, Baig M J, et al. DNA as a digital information storage device: hope or hype? [J] Biotech. 2018, 8(5): 239.
- [2] Zhirnov V, Zadegan R M, Sandhu G S, et al. Nucleic acid memory[J]. Nat. Mater, 2016, 15(4): 366-370.
- [3] Williams E D, Ayres R U, Heller M. The 1.7 kilogram microchip: energy and material use in the production of semiconductor devices[J]. Environ. Sci. Technol., 2002, 36(24): 5 504-5 510.
- [4] Castillo M. From hard drives to flash drives to DNA drives [J]. AJNR. Am. J. Neuroradiol, 2014, 35(1): 1-2.
- [5] 崔光照, 刘玉琳, 张勋才. 数据存储新方向: DNA 分子存储技术 [J]. 计算机工程与应用, 2006, 42 (26): 29-32.
- [6] 吕之品. DNA: 精妙的信息存储体 [J]. 大科学·科学之谜, 2013 (8): 36-38.
- [7] 周谷成, 范艳艳, 肖义军. DNA 存储技术的研究概述 [J]. 生物学通报, 2018, 53 (8): 10-12.
- [8] Dietrich A, Been W. Memory and DNA[J]. J. Theor. Biol., 2001, 208(2): 145-149.
- [9] Garzon M H, Neel A, Chen H. Efficiency and reliability of DNA-based memories[C]. Genetic and Evolutionary Computation Conference. Springer, Berlin, Heidelberg, 2003: 379-389.
- [10] Church G M, Gao Y, Kosuri S. Next-generation digital information storage in DNA[J]. Science, 2012, 337(6 102): 1 628.
- [11] Bogard C M, Rouchka E C. DNA media storage[J]. Prog. Nat. Sci., 2008, 18(5): 603-609.
- [12] Erlich Y, Zielinski D. DNA Fountain enables a robust and efficient storage architecture[J]. Science, 2017, 355(6 328): 950-954.
- [13] 雷鸣, 郑彦宁, 段黎萍. 中国化学药品制剂制造企业海外专利布局研究 [J]. 全球科技经济瞭望, 2018, 33 (5): 59-68.
- [14] Bedau M. Policy-making and systemic complexity[J]. Hastings Cent. Rep., 2014, 44(S5): S29-S30.
- [15] Hagen K. Science Policy and concomitant research in synthetic biology—some critical thoughts[J]. Nanoethics, 2016, 10(2): 1-13.
- [16] Kelle A. Synthetic biology and biosecurity[J]. Embo Reports, 2009, 10(S1): 23-27.
- [17] 中科院武汉文献情报中心, 生物安全战略情报研究中心. 生物安全发展报告——科技保障安全 [M]. 北京: 科学出版社, 2014: 107-123.
- [18] 田德桥, 陆兵. 中国生物安全相关法律法规标准选编 [M]. 北京: 法律出版社, 2017: 1-153.
- [19] 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(7 663): 345-349.
- [20] Bornholt J, Lopez R, Carmean D M, et al. A DNA-based

- archival storage system[J]. ACM SIGPLAN Notice, 2016, 51(4): 637-649.
- [21] Organick L, Ang S D, Chen Y J, et al. Random access in large-scale DNA data storage[J]. Nat. Biotechnol., 2018, 36(3): 242-248.
- [22] Li S, Liu J K, Zhao G, et al. Cads: CRISPR/Cas12a-assisted DNA steganography for securing the storage and transfer of DNA-encoded information[J]. ACS Synth. Biol., 2018, 7(4): 1 174-1 178.

A Comparative Study between China and the United States in the Field of DNA Storage Technology

YUAN Fang, ZHENG Yan-ning, ZHENG Jia, LI Nong, FU Jun-ying
(Institute of Scientific and Technical Information of China, Beijing 100038)

Abstract: Compared with traditional data storage methods, DNA storage technology as an emerging technology has superior storage capacity, stability and security. It has great application prospects in massive data storage and confidential data storage and transmission. This paper compares the differences between China and the United States in the field of DNA storage technology from four aspects: paper publication status, patent application status, relevant policies and key enterprises. Finally, based on the actual situation in China, some suggestions are provided for the development of DNA storage technology in China.

Key words: DNA storage; technology analysis; comparative study

(上接第53页)

Management Mechanism and Enlightenment of International First-class National Laboratories of University: Example of the Scripps Institution of Oceanography

FENG Ze¹, WANG Qiao², CHEN Kai-hua^{1,3}

(1. School of Public Policy and Management, University of Chinese Academy of Sciences, Beijing 100180;

2. School of Economics and Management, Southeast University, Nanjing 211189;

3. Institutes of Science and Development, Chinese Academy of Sciences, Beijing 100190)

Abstract: This paper describes the characteristics of management and operation mechanism of Scripps Institution of Oceanography (SIO), which detailed in five aspects: organizational structure, management mode, equipment and facilities management, funding and personnel management. This study found that more than 90% of SIO's research funding comes from federal institutions' basic research contracts and grants, and the rest comes from state governments, private contracts and research funding within the University of California. SIO is also outstanding in personnel management, resource sharing and organizational structure. Its flexibility and stability, open and shared management and streamlined organizational structure have certain reference significance for China's national laboratory construction.

Key words: national laboratory; Scripps Institution of Oceanography; management and operation mechanism