

# DNA 存储技术国际发展态势分析\*

宋琪\*\* 丁陈君 吴晓燕 陈方  
(中国科学院成都文献情报中心,成都 610041)

**摘要:**随着互联网和人工智能等信息技术和数字技术的快速发展,信息量呈指数级飞快增长,传统的硅基存储设备将面临发展瓶颈,而 DNA 存储具有数据存储密度大、稳定性好、能效低、存储时间长、易备份等优势,可成为数据存储的有效工具。本文对 DNA 存储技术的发展历程及发展现状进行了概述,对在 DNA 存储技术领域方面有政策规划与举措的主要国家和地区进行梳理,同时围绕 DNA 存储技术领域采用文献计量开展国际研发态势研究,结果显示 DNA 存储技术相关论文发表和专利申请持续增加;国家及地区方面,美国、欧洲及我国已经引起重视并进行了相应布局,美国在 DNA 存储技术的论文和专利研发上占绝对优势,我国在该领域活跃程度也相对较强。此外,本文进一步分析了该技术领域的发展现状和存在的问题与挑战,并对我国相关技术发展提出了对策建议。

**关键词:**DNA 存储;战略规划;发展态势

**DOI:**10.16507/j.issn.1006-6055.2020.12.022

## Analysis on the Development Strategies and Trends of DNA Storage Technology\*

SONG Qi\*\* DING Chenjun WU Xiaoyan CHEN Fang

(Chengdu Library and Information Center, Chinese Academy of Sciences, Chengdu 610041, China)

**Abstract:** With the rapid development of information technology and digital technology, such as the Internet and artificial intelligence (AI), the amount of information is growing exponentially, and the development of traditional silicon-based storage devices will face bottlenecks. DNA storage can become an effective tool for data storage since it has the advantages of high data storage density, good stability, low energy efficiency, long storage time and easy backup. In this paper, the development history and current situation of DNA storage technology are summarized, and the major countries and regions that have policy plans and measures in the field of DNA storage technology are sorted out. Meanwhile, the international research and development situation in the field of DNA storage technology is studied by adopting bibliometric methods. The results show that the number of papers published and patent applications for DNA storage technology continues to increase. In terms of countries and regions, the United States, Europe and China have paid attention to it and made corresponding layouts. The United States has an absolute advantage in research papers and patents development on DNA storage technology, while China is relatively active in this field. In addition, the development status, existing problems and challenges in this field are summarized and prospected, and some countermeasures and suggestions are put forward for the development of related technologies in China.

**Keywords:** DNA Storage; Strategy; Development Trend

\* 中国科学院成都文献情报中心 2019 年度中心主任基金项目“生物科技与信息科技间技术融合的演化特征分析”(Y9Z00104)

\*\* Email: songq@clas.ac.cn

二十世纪以来,随着互联网和人工智能等信息技术和数字技术的快速发展,信息量呈指数级飞快增长。据统计,全球数据信息总量将从2018年的30ZB增长到2025年的163ZB<sup>[1]</sup>,传统的硅基存储设备(如硬盘、闪存等)将面临发展瓶颈,而脱氧核糖核酸(Deoxyribonucleic Acid, DNA)中的信息由碱基(A、G、C、T)组成的代码进行存储,可携带大量遗传信息,因此DNA可成为数据存储的有效工具在信息数据存储方面具有极大潜力。DNA存储是将二进制文件通过编码映射成DNA里的碱基序列,按序列顺序通过人工合成技术形成长链DNA来保存数据,DNA存储的写入过程将数字数据映射到DNA核苷酸序列,合成(制造)相应的DNA分子,然后将其存储起来,读取数据涉及对DNA分子进行测序,然后将信息解码回原始数字数据。

在存储信息方面生物存储和传统硅基存储有很多相似之处<sup>[2]</sup>,硅基存储的输入通过输入设备将数据转换成0和1表示,而生物存储的输入是将分子信息通过许多分子传感器转换成内部分子表示,这些分子通过特定调节途径等机制可以很容易地进行重新配置以实现不同的算法,这为DNA作为存储介质奠定了基础。而与传统硅基存储介质相比,DNA存储又具有诸多优势<sup>[3]</sup>:如在数据存储密度方面,硬盘存储每立方厘米约为 $10^{13}$ 位(bits),闪存存储约为 $10^{16}$ 位,而DNA存储可达约 $10^{19}$ 位;在存储时间方面,硬盘、磁带等进行存储最多只能保留约10多年的数据,而DNA则至少可保留上百年的信息,一个显著的例子为2013年*Nature*上发表的文章,宣布从永久冻土捕获的骨头中解码了70万年前马的基因组<sup>[4]</sup>;在能耗方面,1GB数据硬盘存储能耗大约为0.04W,而DNA存储的能耗则可小于 $10^{-10}$ W。2019年,

《科学美国人》杂志将DNA存储技术列为十大新兴技术之一,同时与以金属为材料的微型透镜并称为榜单上最具颠覆性的科技创新技术。在此背景下,本文对DNA存储技术发展态势进行了分析和讨论,并对我国相关技术发展提出了对策建议。

## 1 DNA存储技术发展历程及现状

在1964年的一段采访中,用遗传信息存储数据的概念被首次提出<sup>[5]</sup>,但当时只是预想信息会类似遗传内存(Genetic Memory)一样在计算机中完成。同时期Neiman讨论了开发长链聚合物分子中信息记录和读取的电磁、电子和离子方法的初步研究的可能方向<sup>[6]</sup>,数字信息在生物遗传上进行映射的概念也在Dawkins的*The Blind Watchmaker*书中简要提及<sup>[7]</sup>。这些都只是在理论上说明DNA数据存储的概念,具有里程碑式的相关进展是Joe Davis为“Female Earth”编码了一张古老日耳曼符文35位图像的生物艺术品“Microvenus”<sup>[8]</sup>,首次从实验上证明信息可以在DNA中进行存储。自从在生物艺术领域实践后,DNA存储相关概念再被提及和实践是Clelland等人<sup>[9]</sup>开发的一种基于DNA的双重隐写技术,研究人员创建了一条编码的DNA链并将其隐藏在打印文档中,然后将文件密封并通过美国邮政服务寄送,最终嵌入式的消息(JUNE 6 INVASION; NORMANDY)在实验室环境中成功恢复。自此不断有实验将数字<sup>[10]</sup>、文字<sup>[11-15]</sup>、诗歌<sup>[16-17]</sup>、图片<sup>[14,18]</sup>、歌曲<sup>[14]</sup>等信息写入DNA并进行数据恢复。

2012年,Church等<sup>[19]</sup>开发了一种策略,采用二进制转换将一本书的信息(包含53426个单词、11个JPG图像文件和1个JavaScript程序,共5.27MB)存入DNA中;次年,Goldman等<sup>[20]</sup>利用霍夫曼编码、四倍重叠法、三进制编码等,将五种

格式的文件(154首莎士比亚十四行诗、一张欧洲生物信息研究所的JPG格式图片、一份DNA分子结构的PDF学术论文、26秒MP3格式的马丁路德金演讲片段以及一个霍夫曼编码系统文档)共739kb的硬盘存储空间和523106位Shannon信息编码的计算机文件编码为DNA代码存入DNA中。以Church和Goldman等为代表的研究,进一步明确了DNA作为存储介质的可行性,DNA数字存储逐步进入快速发展期。

此后,各研究团队不断对DNA存储技术领域进行探索。Church研究团队2016年首次证明CRISPR-Cas适应系统可用于在基因组中存储任意序列的DNA寡核苷酸<sup>[21]</sup>;2017年其利用CRISPR系统在活细胞中编码了一个人骑着马狂奔的视频短片,证实了CRISPR系统能够编码诸如数字化视频一样复杂的信息<sup>[22]</sup>;2019年该团队从头开始酶促合成了包含144位的DNA链,并通过流式纳米孔测序演示检索过程,同时设计了一种数字编码器来降低对合成精度和测序覆盖度的要求,实验性地显示了从不完美合成链中获取可靠数据的能力<sup>[23]</sup>;2020年该团队开发了“DNA打孔卡”技术进行多重酶促DNA合成,并将《超级马里奥兄弟》主题音乐Overworld Theme的一部分编码成12个独特的DNA寡核苷酸序列以存储数据<sup>[24]</sup>。麻省理工学院卢冠达(Timothy K Lu)研究团队设计出一种“基因组录音机”,用于记忆活细胞群中的信息,这种基于重组酶的记忆存储系统使研究人员可以将早期信息写入到所有的细菌DNA序列中<sup>[25]</sup>;随后该团队又开发首个可以记录人类细胞中事件持续时间和/或强度的模拟记忆存储系统<sup>[26]</sup>以及一种基于DNA的有序存储器 and 迭代网络运行器(DNA-based Ordered Memory and Iteration Network Operator, DOMINO)的分子记

录仪<sup>[27]</sup>,旨在响应小分子或光照等信息,用DOMINO系统使用CRISPR-Cas9酶的变体产生明确突变。2016年微软与华盛顿大学合作,提出了一种基于DNA的存储系统体系架构(a DNA-based Archival Storage System)——键值存储(Key-value store),同时进一步形成了新的编码方案,该方案能提供可控制的冗余度,在可靠性与密度之间进行权衡<sup>[28]</sup>;两年后两单位又合作创建了一个大型现代数据库,在分子链上存储了破纪录的200MB数据<sup>[29]</sup>,实现了令人惊叹存储容量;微软与华盛顿大学2019年的研究显示能够在DNA中存储1000MB的数据,不仅存储量翻了几倍,同时实现了从存储到提取信息的重大突破<sup>[30]</sup>。除前面提到的几个团队外,瑞士苏黎世联邦理工学院的Robert Grass团队<sup>[31]</sup>引入里德-所罗门(Reed-Solomon, RS)纠错编码技术实现了错误校验,提升了鲁棒性;该团队又与以色列的Erlich团队合作<sup>[32]</sup>设计了一种“DoT”(DNA-of-Things)的存储架构来生产具有不变内存的材料,能够很好地代替传统存储介质的功能,研究人员把二进制的信息转码为由DNA碱基编码的序列信息,将含特定信息的DNA序列包裹在硅纳米颗粒中防止降解,然后将硅纳米颗粒为原材料进行3D打印处理,得到一个兔子形状的玩偶;而此前,Erlich团队也曾在DNA存储技术领域有一定成果,Erlich等<sup>[33]</sup>曾设计了DNA喷泉码(DNA fountain)压缩算法用于DNA存储,研究人员将字符串随机地封装成所谓的“液滴”,并将每个小滴中的1和0映射到DNA中的A、G、C、T中,该算法会自行删除易产生错误的字母组合,同时每个液滴添加了额外的标签,以便于解码时以正确的顺序重组这些文件,编码密度获得极大突破,大大降低了DNA存储的成本。2020年9月25日,在凯茜·弗里曼(Cathy Free-

man)悉尼奥运会女子400米决赛中获得金牌这一历史性比赛的20周年纪念日之际,澳大利亚国家电影和声音档案馆(National Film and Sound Archive of Australia, NFSA)通过保存弗里曼的合成DNA比赛录像来创造历史,这是DNA数据存储技术首次用于长期保存运动图像<sup>[34]</sup>。

## 2 主要国家/地区 DNA 存储技术领域规划与举措

在全球数据信息总量呈指数级增长的背景下,DNA存储技术开始在不同领域探索应用,各国逐渐认识到未来DNA作为存储介质的应用前景以及开发相关新技术的重要性。高通量DNA合成、测序以及编码作为DNA存储技术三个主要的技术领域,成为各国政策规划布局和技术研发的重点。

### 2.1 美国

美国是全球范围内率先对DNA存储技术领域进行研发布局的国家,其多项政策规划均将DNA存储技术领域的相关布局作为一项重要组成部分。2017年3月,美国国防高级研究计划局(The Defense Advanced Research Projects Agency, DARPA)启动分子信息学计划,旨在发现和明确分子在信息存储和处理中可以发挥的功能<sup>[35]</sup>,同时为哈佛大学、布朗大学、伊利诺伊大学和华盛顿大学提供约1500万美元的资助,致力于研究和利用各种分子的结构特征和特性来编码和处理数据。同年5月,美国国家科学基金会(National Science Foundation, NSF)发布“针对信息存储和检索技术的半导体合成生物学(SemiSynBio)”项目指南,拨款400万美元用于探索合成生物学与半导体技术之间的协同作用,促进两大领域的新技术突破,增强信息处理和存储能力。2018年7

月,NSF公布投入1200万美元资助包括基于DNA的可读取电子存储器、使用嵌合DNA的纳米级芯片存储系统、基于纳米孔读取的高度可扩展随机访问DNA数据存储、核酸内存等在内的8个项目进行研究<sup>[36]</sup>。美国情报高级研究计划局(Intelligence Advanced Research Projects Activity, IARPA)2018年7月发布了分子信息存储(Molecular Information Storage, MIST)计划<sup>[37]</sup>,旨在开发可部署的存储技术,减少物理占用空间、功耗和成本。同年10月,在美国国家标准与技术研究院支持下,半导体合成生物学联盟制定第一版《半导体合成生物学路线图2018》<sup>[38]</sup>,该路线图描述了包含基于DNA的大规模信息存储在内的五个技术领域的技术目标。

2019年以来,美国对DNA存储技术领域仍旧加紧布局。在2019年战略框架报告中,DARPA在推动科学技术基础研究战略中明确提出重点关注基于分子信息学的新计算方法,并表示将在更广泛的领域去探索除DNA 4个基本分子以外的更多的数据编码处理新方法<sup>[39]</sup>。2020年2月,NSF发布SemiSynBio-II期的项目招标指南<sup>[40]</sup>,将继续开发与利用结合半导体技术的新兴合成生物学以实现下一代信息存储。DARPA小企业项目办公室(Small Business Programs Office, SBPO)4月发布SBIR/STTR机会(SBIR/STTR Opportunity, SBO)招标合同<sup>[41]</sup>,邀请提交生物医学技术创新研究概念,拟研发快速、灵活地制造用于合成生物学和治疗应用的DNA分子技术,以能够快速有效地合成高精度千碱基对长度的DNA构建体。此外,佐治亚理工学院、麻省理工学院和哈佛大学、洛斯阿拉莫斯国家实验室、桑迪亚国家实验室和美国陆军研究实验室2020年也获IARPA资助以进行包括“写入”、测序读取等

与 DNA 存储相关技术的研发<sup>[42]</sup>。

## 2.2 欧盟

欧盟未明确出台与 DNA 存储相关的政策文件,但对 DNA 存储技术领域的规划大多通过未来和新兴技术(Future and Emerging Technologies, FET)欧盟计划下的 FET Open 进行拨款,资助优瑞卡姆(Eurecom)、法国国家科学研究中心(Centre National de la Recherche Scientifique, CNRS)以及 DNA 合成初创公司海力克斯沃克斯(Helixworks)等开展研究。同时, FET Open 下 OLI-GOARCHINVE 项目为期 3 年(从 2019 年 10 月 1 日到 2022 年 9 月 30 日)聚焦智能 DNA 存储系统的新技术研究,涉及从编码到测序解码的全领域,将为开发构建智能 DNA 存储系统所需的基本技术铺平道路。

## 2.3 国外其他国家

除美国和欧洲外,国外其他国家在 DNA 存储和合成生物学领域也有一定的行动和布局。在合成生物学方面,日本采取了一系列为合成生物学研究人员建立一个共同体的行动,比如 2005 年成立了日本细胞合成研究协会,其中胚胎科学与技术前期研究(Precursory Research for Embryonic Science and Technology, PRESTO)为合成生物学项目提供特殊资金等,日本将合成生物学视为其未来科学政策的重要组成部分,并力争在该领域跻身国际前列<sup>[43]</sup>;2016 年日本丰田汽车公司通过“独特的基因样本调整方法”和“下一代基因测序仪”等的成功研究,开发出了快速、低成本 DNA 解析新技术 GRAS,并且与具有丰富 DNA 解析实绩的日本公益财团上总 DNA 研究所达成协议,准备对该技术开展进一步的验证评价<sup>[44]</sup>;2019 年 5 月由 16 所合成生物设施机构联合发起的国际合成生物设施联盟(Global Biofoundry Alliance, GBA)

在日本神户成立,旨在促进全球合成生物学相关发展等<sup>[45]</sup>。澳大利亚联邦科学与工业研究组织(Commonwealth Scientific and Industrial Research Organisation, CSIRO)表示建立包括合成生物学在内的六个未来科学平台,并为之每年投资超过 5200 万澳元<sup>[46]</sup>,CSIRO 投资创建的合成生物学未来科学平台(SynBio FSP)旨在支持多领域的创新等来提高澳大利亚的竞争力<sup>[47]</sup>。

## 2.4 中国

我国高度重视 DNA 存储技术领域的研发,通过对合成生物学等领域专项进行部署和资助。2018 年国家重点研发计划合成生物学重点专项共有 36 个项目,总经费接近 7.98 亿元。其中专门设置了与 DNA 存储技术相关的项目。“高通量脱氧核糖核酸(DNA)合成创新技术及仪器研发”项目由中国人民解放军军事科学院军事医学研究院牵头,开发化学法 DNA 合成新技术、复杂结构序列的高效合成技术和大片段 DNA 高效组装技术,研制基于高通量芯片的原位组装控制系统及仪器。“使用合成 DNA 进行数据存储的技术研发”项目由南方科技大学牵头,上海交通大学、中国科学院长春应用化学研究所、福州大学、同济大学联合申报。项目拟开发利用合成 DNA 高效快速、高密度数据加密编码转码,随机读取,无损解读新方法;开发多类型数据存储 DNA 介质;通过合成 DNA 开发快速编码,存储及数据读取的集成型软件系统。该项目旨在利用新型存储技术应对大数据的爆炸式增长,解决数据快速增长与数据有效存储和利用之间的矛盾,推动我国在 DNA 数据存储基础研究领域的原始创新和科学突破<sup>[48]</sup>。2020 年,中国科学院深圳先进技术研究院牵头获批 7 个国家科技部重点研发计划项目,获批“合成生物学”等三个重点专项中总经费 8683 万元,在

“合成生物学”重点专项中,深圳先进院获批4个项目,其中“多方协同合成基因信息安全存取方法研究”项目主要针对DNA存储过程中多方协同操作和安全性问题提出混合加密方法和增量编码技术,进一步探究如何保障合成基因信息多方安全协同与提高DNA存储信息高效管理能力,实现合成基因在复杂信息存储需求场景中的存储与可靠读取<sup>[49]</sup>。

### 3 计量分析

#### 3.1 数据来源与方法

通过将Web of Science与Scopus以及ACM全文数据库相比较,为检索出更多与“DNA存储”技术相关的研究,了解研究趋势、国际竞争格局与热点等情况,论文分析选择了Web of Science数据库,利用Web of Science平台的Web of Science™核心合集,在科学引文索引扩展版(Science Citation Index Expanded, SCIE)数据库中进行检索(检索日期为2020年5月19日),基于Web of Science数据库、科睿唯安公司的Derwent Data Analyzer(DDA)数据分析工具以及网络分析软件VOSviewer开展包括发文年度变化趋势、主要国家、主要研究机构、主要研究学科方向等文献计量分析。在专利分析部分,以北京合享智慧科技有限公司的incoPat专利数据库作为检索来源,通过关键词和专利分类号进行组合检索(检索日期为2020年5月15日)。

文献计量检索式:TS=(( "DNA data storage" OR "DNA storage" OR "DNA-based storage" OR "DNA digital storage" OR ("DNA digital data storage") OR "information storage in DNA" OR (data storage in DNA) OR (storing data in DNA) OR (molecular NEAR/5 (long-term data storage)) OR

((long-term storage) AND information AND DNA) OR (((DNA representation) AND ((long term storage) AND DNA)) AND ("digital information" OR "digital data")) OR ((DNA representation) AND digital information) OR ((storage of information) AND DNA AND binary) OR ((organic data memory) AND (DNA approach)) OR (secret signatures inside genomic DNA) OR (poetic gene) OR ("DNA writing" AND "living cell populations") OR ("DNA nanostructure" AND "information storage") OR ("data storage device" NEAR/5 DNA) OR "DNA memory" OR (digit \* AND watermark AND DNA and information) OR ((DNA watermark) AND (digital data)) OR (("secret-message" OR steganography OR encryption OR cryptography) AND DNA) OR ((molecular-level storage) AND (digital information)) OR "DNA microdots" OR "DNA archiving" OR ((storage OR CRISPR) AND ("molecular record \*")) OR "DNA-recording" OR "DNA based memory" OR (DNA AND (storage AND (DOMINO OR recorder))) OR "DNA fountain" OR (molecule AND (genetic information) AND "store the world's data") OR ("error correction" AND "DNA data storage") OR ("genetic recording" AND "CRISPR-Cas") OR (DNA AND "random access memory" AND storage) OR ((encoding data in DNA) AND stor \* ) OR ((nucleic acids) AND (digital storage)) OR ("molecular logic" AND storage AND digital data) OR ((DNA computation) AND (molecular logic) AND storage) OR (((synthe \* DNA) AND storage) AND digital) OR "renewable DNA computing") NOT (sperm OR blood OR food OR "rainbow trout" OR farming

OR "birth cohort study" OR disease OR GenBank OR database OR NCBI OR "species identif \* " OR "VCFtools" OR G4 OR metall \* OR (density-functional theory) OR "endothelial cell" OR "pollution monitor \* " OR "fluorescent protein" OR clinical OR patient OR (specific protein) OR (varietal identification) OR (strain classification) OR "16s rDNA" OR "functional domain composition" OR "KODAK" OR biodiversity OR epigenetics OR "microRNA" OR exocytosis OR potato OR legume OR ecological OR globulin OR agricultural OR haplotype OR "fatty-acids" OR "cell regularion" OR database OR metabol \* OR breed \* OR immuse OR "genetic diversity" OR "gene therapy" OR glacier OR "nucleosome assembly" OR quercetin OR specimens OR grafting OR "human genetics" OR cutaneous OR "Na<sup>+</sup>-Ca<sup>2+</sup> exchange" OR "age estimation" OR phylogenetic OR "minor-groove" OR (fresh water) OR toxoplasmosis) NOT (chaotic OR chaos OR chaotic) NOT ("paraffin-embedding" OR "nuclear protein" OR "insitu hybridization" OR "DNA damage" OR cytomet \* OR "tumor-cells" OR lymphocytes OR adenoviral OR interleukin OR histone OR calcium OR calreticulin OR "stellate cell" OR "human papillomavirus" OR "cell death" OR "freeze-drying" OR "seed protein" OR "bacterial community" OR "scanning tunneling microscopy" OR mouse OR fish OR cholesterol OR microbiome OR telomere OR "microbial communities" OR cholesterol OR "transcriptional regulator" OR coexpression OR pharmaceutical OR domestication OR forensic OR paleosols OR iron OR genotyping OR "gram-positive" OR Sulphur OR basidiospores OR prolamin OR

"internet of things" OR "shock proteins"))为检索式,对 SCIE 收录的自 1900 年以来的文献进行检索,检索日期为 2020 年 5 月 19 日,检索结果为 803 条。DNA 存储作为一个较新的领域,由于数据量相对较少,在制定检索式后选择了人工清洗,清洗后的结果作为本文使用数据进行了后续分析。

专利计量检索式:TI = ("nucleic acid" OR DNA) AND Tiab = ((store or stored or storage or storing OR memory) AND ("nucleic acid" OR DNA) and (information or data)) and IPC = (C12 or C07 or G06 or G09 or G11 or G16 OR H01 OR H03),检索所有年限的全球专利,检索日期为 2020 年 5 月 15 日,检索结果为 370 项(502 件),由于专利数量不多,因此进行人工清洗和技术分类后进行后续分析。

## 3.2 论文分析结果

### 3.2.1 发文年度变化趋势

截止到 2020 年 5 月 19 日,与 DNA 存储的相关论文总共 794 篇。从图 1 展示的年度论文产出情况看,自 1991 年来发文量不断增加。1994 年前 DNA 存储发文量均低于 10 篇,由于“DNA 计算”的产生,自 1995 年以来将生物领域 DNA 的碱基与计算机领域的 0 和 1 进行进制转化,催生了小部分与“DNA 存储”领域相关的论文发表;依托“DNA 密码学”和“DNA 隐写术”相关研究论文的出现,将数据存入 DNA 中进行加密的研究逐渐增多,但仍处于缓慢发展阶段;随着“DNA 存储”的概念进一步明晰,从 2010 年开始,与 DNA 存储技术相关研究的发展速度不断加快,尤其是 2016 年之后,DNA 存储领域进入快速发展时期。

### 3.2.2 主要国家分析

从主要国家(发文量超过 20 篇)分布来看

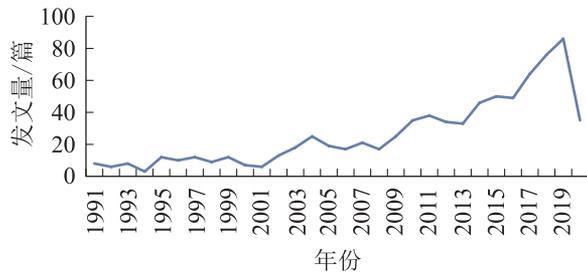


图 1 DNA 存储技术领域发文量年度分布

Fig. 1 Annual Distribution of Published Articles in the Field of DNA Storage Technology

(图 2), DNA 存储技术发文量最多的国家依次是美国、中国、德国、英国、加拿大、法国、日本等, 美国以 283 篇的发文量处于遥遥领先的地位, 占到了 37% 的份额; 中国位居第二, 发文量共计 105 篇, 占到了 14% 的份额。来自这些主要国家的发文量共计 762 篇, 约占总发文量的 96%。

从发文国家/地区的合作网络来看(图 3), 美国处于核心位置, 且与中国合作强度最强, 合作次

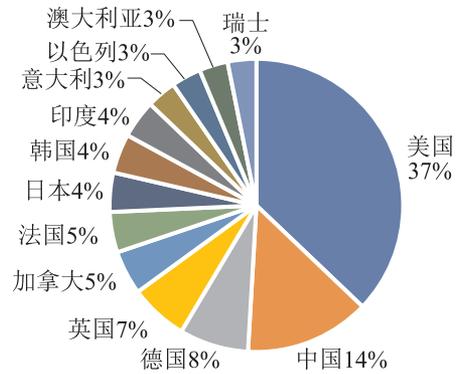


图 2 DNA 存储技术领域主要发文国家

Fig. 2 Major Countries in the Field of DNA Storage Technology

数为 17 次, 与英国、以色列分别合作 13 次, 与德国、瑞士、加拿大合作次数依次为 10 次、9 次和 6 次; 除各国与美国的合作外, 中国与英国合作强度最强, 合作次数为 4 次; 英国与德国合作 6 次; 加拿大与法国合作 3 次。相对来说, 中国在 DNA 存储技术领域的国际合作活跃程度相对较强, 但与美国相比还有较大差距。

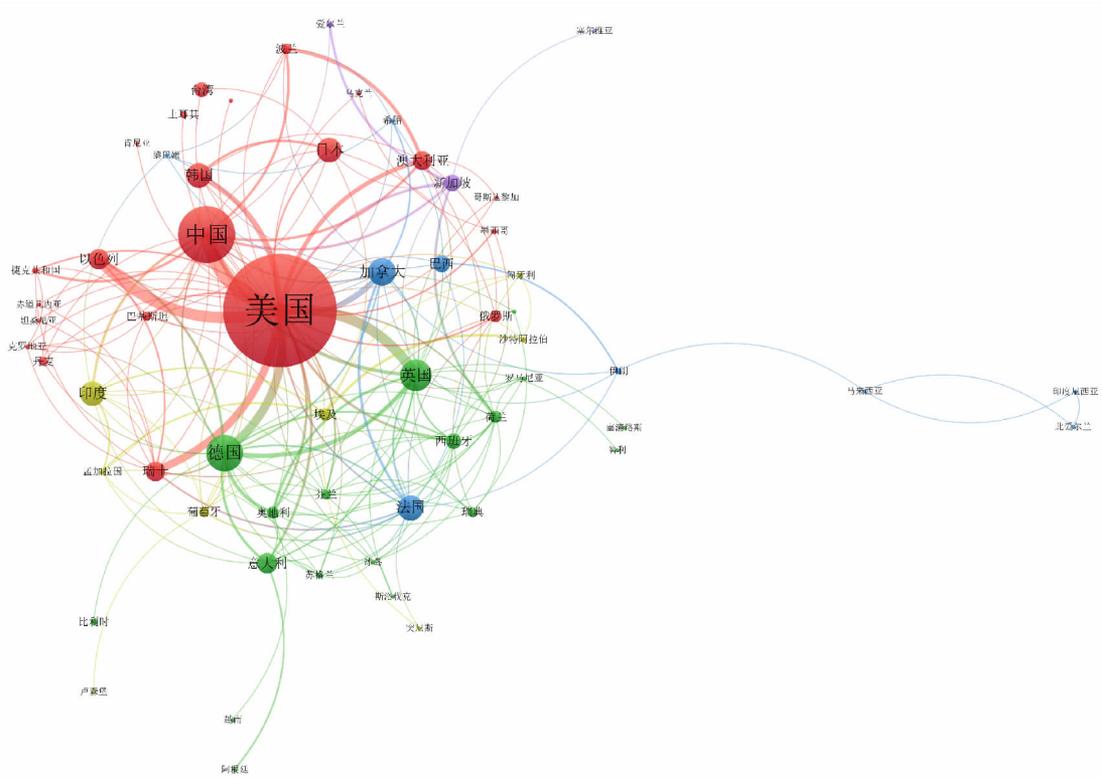


图 3 全球 DNA 存储技术领域发文的国家(地区)合作网络

Fig. 3 National (Regional) Cooperation Network Issued in the Field of Global DNA Storage Technology

### 3.2.3 主要研究机构分析

从开展 DNA 存储的研究机构排名来看(表 1),美国仍处于领先地位,前十位中有 7 家研究机构来自美国,分别为加州大学系统、哈佛大学、麻省理工学院、伊利诺伊大学、华盛顿大学、美国能源部以及微软研究院。除美国外,法国国家科研中心、瑞士苏黎世联邦理工学院、中国科学院以及德国亥姆霍兹联合会也跻身 DNA 存储技术领域研究机构前十强,分别位于第 2 位、第 5 位、第 7 位和第 10 位(其中德国亥姆霍兹联合会与微软研究均为 12 篇,并列第 10 名)。总的来看,无论从发文量还是总被引频次和篇均被引次数,美国加州大学系统、法国国家科研中心和美国哈佛大学均稳居前三。

表 1 DNA 存储技术领域发文量排名前 10 位的研究机构

Tab.1 Top10 Research Institutions in the Field of DNA Storage Technology

排名	研究机构	发文量 /篇	被引频 次总计 /次	篇均被 引次数 /(次/篇)
1	加州大学系统	39	1575	40.4
2	法国国家科研中心	20	905	45.3
3	哈佛大学	16	865	54.1
4	麻省理工学院	16	544	34.0
5	苏黎世联邦理工学院	14	202	14.4
6	伊利诺伊大学	14	308	22.0
7	中国科学院	13	230	17.7
8	华盛顿大学	13	213	16.4
9	美国能源部	13	269	20.7
10	微软研究院	12	188	15.7
10	德国亥姆霍兹联合会	12	386	32.2

从研究机构合作网络总体来看(如图 4 所示),研究机构之间的合作强度并不高,在该领域的发文也多为机构内部间合作,合作较多的包括微软研究院与华盛顿大学合作 6 次,南洋理工大学与新加坡国立大学合作 4 次,伊利诺伊大学与南洋理工大学合作 3 次,麻省理工学院与哈佛大

学合作 2 次。

### 3.2.4 主要研究方向

图 5 对 DNA 存储技术领域的 10 个主要研究方向进行了统计,其中计算机科学方向发文量最多,有 162 篇;其次为生物化学与分子生物学方向,有 149 篇,这两个研究方向的发文量总和占总发文量的大约 40%。科学技术其他主题(132 篇)、生物技术与应用微生物(111 篇)以及化学(107 篇)等也是 DNA 存储技术领域所涉及的主要研究方向。

### 3.2.5 发文作者分布

全球共有 2900 余名研究学者在 DNA 存储研究领域发表了相关文章,该领域发文量排名前 10 位的发文作者(表 2)中有 7 位来自美国;另外 3 位作者分别来自瑞士的苏黎世联邦理工学院、韩国东明大学以及法国特拉斯堡大学(分别为 Robert Grass、Suk-Hwan Lee 以及 Jean-Francois Lutz)。Robert Grass 和他的研究团队在 2015 年把包括 1291 年《瑞士联邦宪章》和阿基米德的方法论在内的 83Kb 的文本编码进 DNA,进行加速老化实验以测量 DNA 衰变动力学,即使在 70°C 的二氧化硅中处理了 DNA 一周后,也可以无错误地恢复原始信息;Suk-Hwan Lee 的研究主要集中于 DNA 水印和 DNA 隐写来保护数据安全;Jean-Francois Lutz 在 2015 年发表研究论文通过化学方法达到聚合物数据存储,以实现与 DNA 存储相类似的功能。

### 3.2.6 高被引论文

表 3 列出了 DNA 存储技术领域研究排名前 5 位的高被引论文。排名前 5 位的高被引论文中,有 4 篇发表在综合性期刊 *Science* 和 *Nature* 上,上,有一篇发表到生物学领域的期刊 *Genome Biology* 上。DNA 存储技术领域排名前 5 位的高被

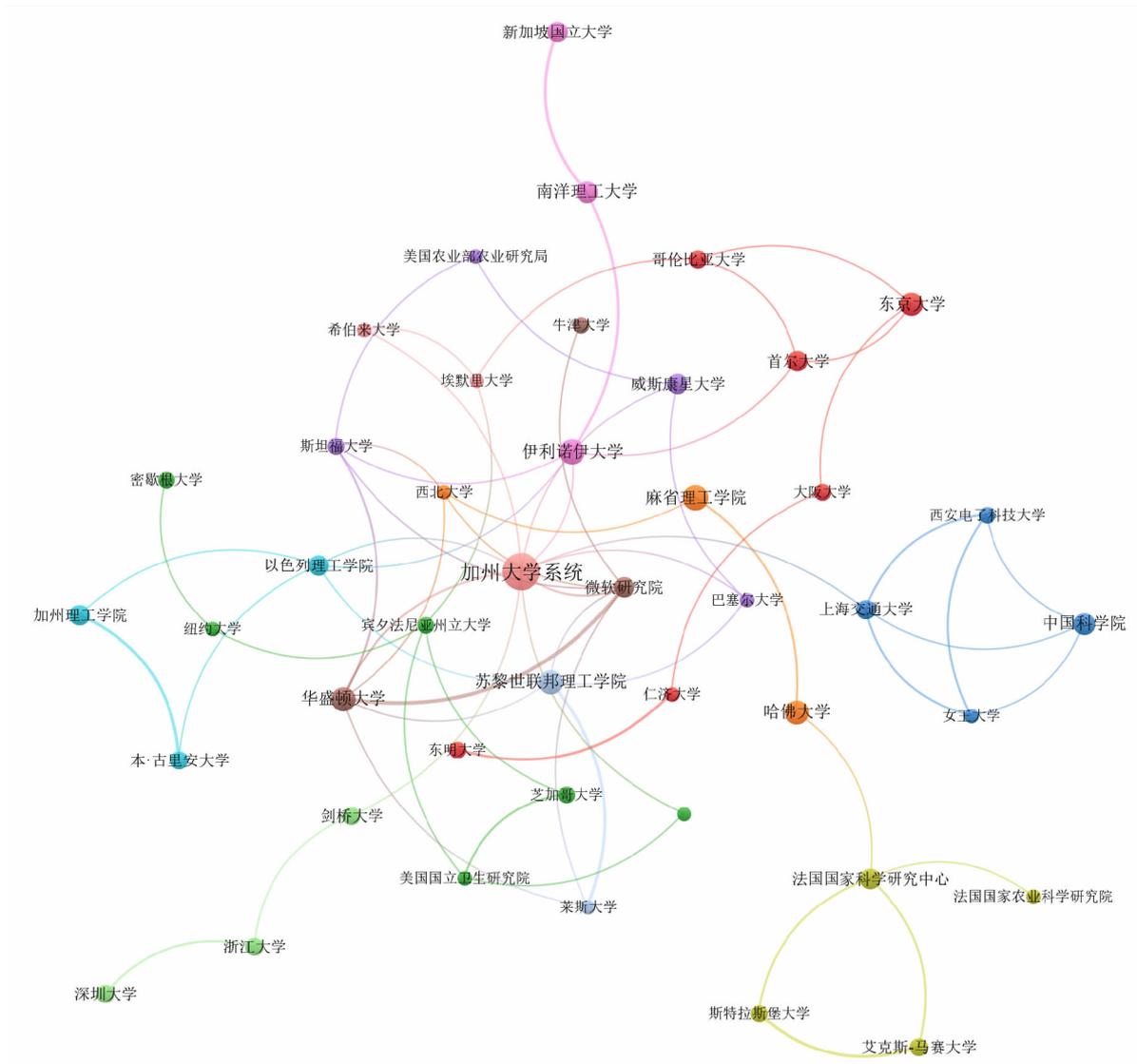


图 4 DNA 存储技术领域发文的研究机构合作网络

Fig. 4 Cooperative Network of Research Institutions in the Field of DNA Storage Technology

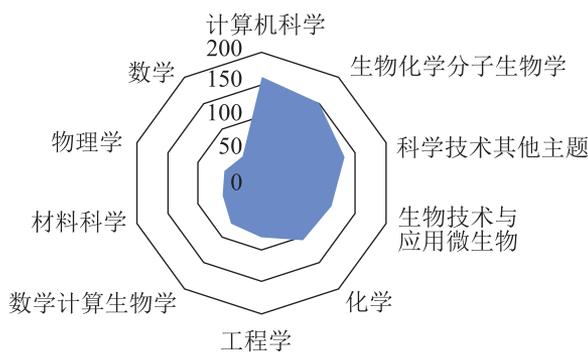


图 5 DNA 存储技术领域的 10 个主要研究方向

Fig. 5 10 Main Research Directions in the Field of DNA Storage Technology

表 2 DNA 存储技术领域研究论文发文量排名前 10 位的发文作者分布

Tab. 2 Distribution of Top10 Authors of Research Papers in the Field of DNA Storage Technology

序号	作者	所属研究机构	机构所属国家	发文量/篇
1	Luis Ceze	华盛顿大学	美国	12
2	Karin Strauss	微软研究院	美国	12
3	Olgica Milenkovic	伊利诺伊大学	美国	10
4	Robert Grass	苏黎世联邦理工学院	瑞士	7
5	Randolph Lopez	华盛顿大学	美国	7
6	Timothy K Lu	麻省理工学院	美国	7
7	Georg Seelig	华盛顿大学	美国	7
8	Reinhard Heckel	莱斯大学	美国	6
9	Suk-Hwan Lee	东明大学	韩国	6
10	Jean-Francois Lutz	斯特拉斯堡大学	法国	6

表3 DNA存储技术领域研究排名前5位的高被引论文

Tab. 3 Top5 Highly Cited Papers in the Field of DNA Storage Technology

序号	题名	来源期刊	发文时间	第一作者
1	Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome	<i>Science</i>	2010	Daniel G Gibson
2	Next-Generation Digital Information Storage in DNA	<i>Science</i>	2012	George M Church
3	The case for cloud computing in genome informatics	<i>Genome Biology</i>	2010	Lincoln D Stein
4	Towards practical, high-capacity, low-maintenance information storage in synthesized DNA	<i>Nature</i>	2013	Nick Goldman
5	Hiding messages in DNA microdots	<i>Nature</i>	1999	CT Clelland

被引论文中,最早为1999年Clelland等<sup>[9]</sup>发表在*Nature*上一种基于DNA的双重隐写技术。

### 3.3 专利分析结果

#### 3.3.1 DNA存储技术专利申请趋势

截至2020年5月15日,公开的所有DNA存储技术相关专利共计249项(380件),专利申请趋势如图6所示。1994—2000年,DNA存储技术专利申请量较少,处于技术萌芽期;2001—2005年,专利数量出现一个小高峰,该阶段DNA存储技术研发受到重视,并得以发展;2006—2012年,专利数量回落,说明DNA存储技术专利研发遇到了技术瓶颈;2013年之后,DNA存储技术有了新的突破,其专利数量呈现快速增加趋势,2018年专利申请量达到48项。由于专利申请到公开最长有18个月迟滞,截至检索日,2018年和2019年还有部分专利申请尚未公开。

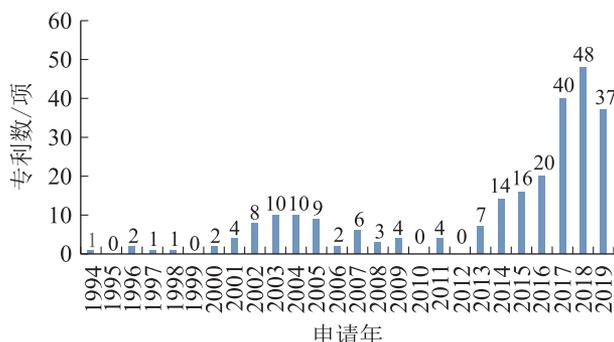


图6 DNA存储技术全球专利申请趋势

Fig. 6 Global Patent Application Trend of DNA Storage Technology

#### 3.3.2 重要国家分布

将专利家族展开对各国/地区受理专利数量进行统计分析(图7),美国是DNA存储技术专利申请人最重视的技术保护市场地,其次是中国和欧洲地区,再次是韩国、日本等。从专利申请人来源国家的分布情况来看(图8),美国是DNA存储技术专利主要来源国家,专利占比58%,远超排名第二的中国(19%),其次是法国、日本、韩国等。

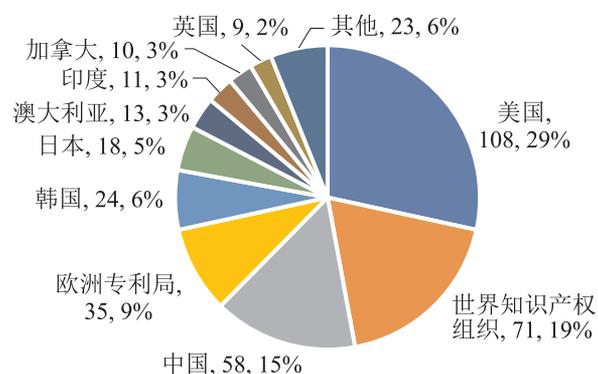


图7 DNA存储技术专利受理国家/地区分布

Fig. 7 Distribution of Countries/Regions Patent Records on DNA Storage Technology

对比分析美国和中国的DNA存储技术专利申请趋势(图9),美国相关专利申请趋势与全球专利申请总体趋势基本一致(图6),是全球DNA存储技术专利研发的主导力量;中国自2004年申请了第一件相关技术专利,2004—2013年保持缓慢的发展,2014年之后专利数量明显增长,成为

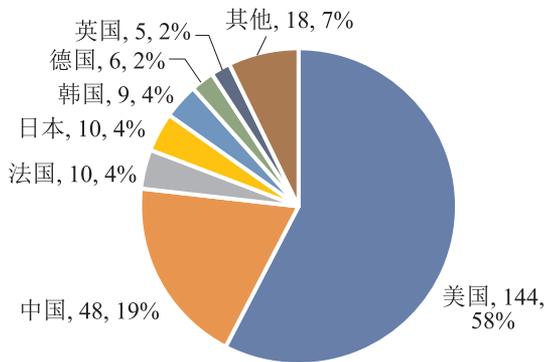


图8 DNA 存储技术专利来源国家/地区分布

Fig. 8 Distribution of Application Countries/Regions Patent Families on DNA Storage Technology

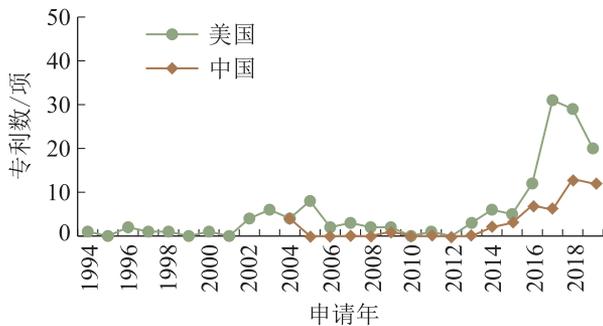


图9 DNA 存储技术专利中美申请趋势

Fig. 9 Patent Application Trends of America and China for DNA Storage Technology

DNA 存储技术专利研发的重要来源国之一。

从专利技术流向的角度来看(表4),美国 DNA 存储技术专利不仅布局了世界知识产权组

织(World Intellectual Property Organization, WIPO)和欧洲专利局(European Patent Office, EPO),还流向了包括韩国、中国、日本、澳大利亚、印度、加拿大、英国在内的全球重要市场,形成较为完善的海外专利布局;中国 DNA 存储技术专利主要集中在本国受理,还有少量布局 WIPO、EPO 和美国市场;法国、德国、日本、韩国、印度和英国等国家虽然专利数量不多,但也布局了海外市场。

### 3.3.3 重要专利申请人分析

在 DNA 存储技术专利申请量排名前 15 位的机构中,企业占 9 席,高校和研究所占 6 席(表 5)。从国别来看,美国机构占 12 席,中国机构占 3 席(天津大学、中国科学院、苏州泓迅生物科技公司),美国机构在 DNA 存储技术专利研发上占主导优势。

专利申请量排名第一位的是美国 Catalog 公司,该公司于 2016 年由麻省理工学院的两名研究生创立,至今申请了 13 项相关专利,保持较高的活跃度。申请量排名第二的是美国微软公司,微软自 2015 年开始申请相关专利,目前有 12 项专利。排名其后的是美国 Thomson Licensing 公司、美国 Iridia 公司、美国哈佛大学等。美国 Nanogen 公司最早进行 DNA 存储技术专利研发,在 1994 至 2000 年期间申请了 4 项相关专利,而后专利申

表4 DNA 存储技术专利重要国家技术流向(单位:件)

Tab. 4 Technology Flow in Important Countries of DNA Storage Technology Patent Records

来源国 \ 受理国/专利局	美国	世界知识产权组织	中国	欧洲专利局	韩国	日本	澳大利亚	印度	加拿大	英国
美国	89	47	12	19	13	11	11	6	7	2
中国	1	7	40	1	0	0	0	0	0	0
法国	3	7	1	10	1	0	0	0	0	0
德国	2	1	2	1	1	0	0	1	1	0
日本	5	2	1	1	0	4	0	0	0	0
韩国	2	1	1	0	7	0	0	0	0	0
印度	1	2	0	0	0	1	0	3	1	1
英国	0	1	0	0	0	0	0	0	0	6

请量为零,其研发中心可能已经转移。美国哈佛大学、中国科学院、美国 Clemson 大学都在 2003 和 2004 年左右开始了相关专利申请,其他机构大多在 2014 年以后开始申请相关专利。

表 5 DNA 存储技术重要专利申请人

Tab. 5 Important Patent Applicants for DNA Storage Technology

序号	专利申请人	专利数 /项	活跃时间 /年
1	美国 Catalog 公司	13	2016—2019
2	美国微软公司	12	2015—2019
3	美国 Thomson Licensing 公司	10	2014—2017
4	美国 Iridia 公司	10	2017—2019
5	美国哈佛大学	10	2003—2018
6	美国 Twist Bioscience 公司	9	2016—2019
7	天津大学	8	2018—2019
8	美国 Roswell 生物技术公司	5	2017—2018
9	中国科学院	5	2004—2019
10	美国 Nanogen 公司	4	1994—2000
11	美国 Clemson 大学	4	2003—2008
12	美国华盛顿大学	4	2015—2019
13	美国麻省理工学院	4	2014—2019
14	美国 Molecular Assemblies 公司	4	2018—2019
15	苏州泓迅生物科技公司	4	2015—2019

从专利的技术流向来看(表 6),美国机构特别是 Catalog、微软、Thomson Licensing、Iridia、哈佛大学、Twist Bioscience、Roswell 等头部机构都非常重视全球市场,进行了比较全面的海外布局。天津大学、中国科学院相关专利主要在中国申请,苏州泓迅生物科技公司申请了一项 WIPO 专利。

分析重要申请人的专利主要保护范围(表 6),Catalog 研究基于核酸的数据存储系统以及用于核酸存储的组合物和方法;微软关注通过迭代 DNA 编辑进行存储、DNA 制造、存储和访问系统以及 DNA 载体介质;Thomson Licensing 研究将数字信息存储在 DNA 的设备方法和系统;Iridia 研发带电聚合物存储信息系统;哈佛大学开发利用

核酸存储信息的方法、用 CRISPR-CAS 系统进行分子记录的方法和系统、DNA 存储信息进行编码和解码方法;Twist Bioscience 关注基于 DNA 的数字信息存储器和用于核酸合成的柔性基底;天津大学致力于 DNA 信息存储编码方法开发;Roswell 开发了基于酶的 DNA 数据存储方法和系统以及无扩增 DNA 数据存储的方法、装置和系统;中国科学院主要保护微流控芯片 DNA 分子存储器和 DNA 数据文件的读取方法及计算机可读存储介质;Nanogen 关注 DNA 光存储装置和向 DNA 光存储设备写入和读取数据的方法;Clemson 大学开发 DNA 的存储器件及其读写方法;华盛顿大学研究基于核酸的电可读只读存储器和基于核酸的数据集成存储系统;麻省理工学院开发了用于信息存储的体外 DNA 写入方法;Molecular Assemblies 研发均聚物编码的核酸存储器;泓迅致力于人工合成 DNA 存储介质的信息存储读取方法和图像重构方法及装置。

### 3.3.4 主要国际专利分类

从专利分类号的角度看(表 7),DNA 存储技术专利主要集中在 G06 部(计算;推算;计数)和 C12 部(生物化学;微生物学;酶学;突变或遗传工程),属于计算机科学与生物科学的交叉技术。从 IPC 小组来看,属于 G06F(电数字数据处理)小组的专利最多,其次是 C12Q(包含酶、核酸或微生物的测定或检验方法)、G06N(基于特定计算模型的计算机系统)、C12N(微生物或酶;变异或遗传工程)、G11C(静态存储器)、G16B(生物信息学),再次是 H03M(一般编码、译码或代码转换)、B01J(化学或物理方法)、G01N(借助于测定材料的化学或物理性质来测试或分析材料)、G11B(基于记录载体和换能器之间的相对运动而实现的信息存储)。

表 6 DNA 存储技术重要专利申请人技术解析

Tab.6 Technical Analysis of Important Patent Applicants for DNA Storage Technology

序号	申请人	技术流向	主要保护范围
1	美国 Catalog 公司	US, WIPO, EPO, AU, IN, KR, GB	基于核酸的数据存储系统; 用于核酸存储的组合物和方法
2	美国微软公司	WIPO, US, EPO, CN, IN	迭代 DNA 编辑进行存储; DNA 制造, 存储和访问系统; DNA 载体介质
3	美国 Thomson Licensing 公司	EPO, WIPO, US, CN	将数字信息存储在 DNA 的设备方法和系统
4	美国 Iridia 公司	US, WIPO, EPO, CN, AU, JP, KR, AU, CA, IL, IN, TW, MX	带电聚合物存储信息的新系统;
5	美国哈佛大学	US, CN, WIPO, EPO, KR	利用核酸存储信息的方法; 用 CRISPR-CAS 系统进行分子记录的方法和系统; DNA 存储信息进行编码和解码方法
6	美国 Twist Bioscience 公司	US, CN, WIPO, KR, EPO, CA, SG	基于 DNA 的数字信息存储器; 用于核酸合成的柔性基底
7	天津大学	CN	DNA 信息存储编码方法
8	美国 Roswell 生物技术公司	US, WIPO, CN, CA, EPO, KR	基于酶的 DNA 数据存储方法和系统; 无扩增 DNA 数据存储的方法、装置和系统;
9	中国科学院	CN	微流控芯片 DNA 分子存储器; DNA 数据文件的读取方法及计算机可读存储介质
10	美国 Nanogen 公司	US, EPO, WIPO	DNA 光存储装置; 向 DNA 光存储设备写入和读取数据的方法
11	美国 Clemson 大学	US, WIPO	基于 DNA 的存储器件及其读写方法
12	美国华盛顿大学	US, WIPO	基于核酸的电可读只读存储器; 基于核酸的数据集成存储系统
13	美国麻省理工学院	US, WIPO	用于信息存储的体外 DNA 写入方法
14	美国 Molecular Assemblies 公司	US, EPO, KR, WIPO	均聚物编码的核酸存储器
15	苏州泓迅生物科技有限公司	CN, WIPO	人工合成 DNA 存储介质的信息存储读取方法; 图像重构方法及装置

表 7 DNA 存储技术专利 IPC 分布

Tab.7 IPC Distribution of DNA Storage Technology Patent Records

IPC 小组	专利数/项	分类号解释
G06F	89	电数字数据处理
C12Q	76	包含酶、核酸或微生物的测定或检验方法
G06N	63	基于特定计算模型的计算机系统
C12N	39	微生物或酶; 变异或遗传工程
G11C	39	静态存储器
G16B	35	生物信息学, 例如特别适用于计算分子生物学中的遗传或蛋白质相关数据处理的信息与通信技术
H03M	24	一般编码、译码或代码转换;
B01J	15	化学或物理方法, 例如, 催化作用或胶体化学
G01N	15	借助于测定材料的化学或物理性质来测试或分析材料
G11B	15	基于记录载体和换能器之间的相对运动而实现的信息存储

## 4 总结与建议

当前, 人类面临数据存储难题, 增长过快的信息洪流即将冲垮传统硬盘的承载力。2019 年, 《科学美国人》将 DNA 存储技术列为十大新兴技术之一, 同时与以金属为材料的微型透镜并称为榜单上最具颠覆性的科技创新技术。DNA 作为生物学数据库, 在信息数据存储方面也具有极大潜力, 早在上世纪 80 年代后期就已有科学家证明了 DNA 作为数据存储介质的能力和优势, 近十年来, 该领域的研究在存储的数据量和存储密度的最大化方面取得了重大进步, 随着 DNA 合成技术、测序技术的不断进步, 快速无错、随机读取 DNA 中存储的信息成为可能, 同时 DNA 编码成

本的下降,学术界、私营部门和公共部门也将通过跨部门合作推动该技术的商业化应用,其展现出来的优势将彻底改变未来数据访问和计算领域。

DNA 存储技术领域因其优越性已引起多个国家和地区决策层的重视,美国是进行相关规划布局最多的国家,其涵盖了从数据“写入”到“读取”的多个技术过程;欧盟虽未明确出台文件政策进行 DNA 存储技术布局,但也对相关研发进行了资助;日本、澳大利亚等国在合成生物学领域重视度也进一步提升;我国也已着手相关领域的研发布局。从计量分析看,DNA 存储技术在论文发表和专利申请方面都呈增长趋势。在论文方面,美国以 283 篇的论文发文量处于遥遥领先的地位,占到了 37% 的份额,中国发文量位居全球第二位;从发文国家的合作网络来看,中国在 DNA 存储技术领域的国际合作活跃程度相对较强,但与美国相比还有较大差距;从机构分析来看,中国科学院在发文量和专利申请量的排名上都进入了全球的前 10 位;在机构合作方面,我国机构合作主要集中在国内,而与国外机构主要集中在和美国加州大学系统的合作中。在专利方面,美国 DNA 存储技术领域的专利申请量上的优势也显示出其在 DNA 存储技术专利技术的主导地位,我国 2014 年以后专利申请数量也增长明显,成为该领域专利数量排名第二的来源国家;从重要专利申请机构来看,DNA 存储技术领域排名前 5 的均为美国机构,前 15 名中我国只占 3 个;从专利技术流向来看,美国更加重视全球市场,而我国则主要集中于国内;从专利主要保护范围来看,各机构对 DNA 存储技术的多方面系统进行保护比较重视;从专利分类号来看,DNA 存储技术专利表明其既涵盖计算机领域也包含生

物科学领域,而电数字数据处理小组是最多的一类。

对我国而言,大力发展 DNA 存储技术对开发替代性的数据存储介质、保障能源安全、维护生态环境可持续发展等具有重要战略意义,虽然我国在合成生物学领域已有所部署和关注(例如,由南方科技大学承担的国家重点研发计划“合成生物学”重点专项,使用合成 DNA 进行数据存储的技术研发项目已于 2019 年 12 月 9 日正式启动),但与其他国家尤其是美国相比还存在较大差距。另一方面,目前 DNA 存储技术还受到数据覆盖和重写、随机读写等方面的制约,尤其是 DNA 合成的成本和合成速率仍然是限制 DNA 存储技术领域发展的主要瓶颈,为此,就我国推动 DNA 存储领域技术发展提出以下几点建议:

**首先,加大支持和资助力度,强化相关技术领域研发及战略布局。**利用工程酶合成 DNA 片段的技术或成为“第二代”合成技术,结合微列阵平台的发展,将从根本上降低成本,促进数据存储应用领域的新技术开发;开发适合 DNA 存储设备的高级编码方案和操作系统,构建存储数据的随机访问检索机制,使进一步轻松读取 DNA 中存储的数据成为现实;通过应用 DNA 修饰技术,大幅增加 DNA 中数据存储的密度。随着核心技术的不断改进,DNA 存储技术领域有望在更广泛的计算机科学领域和其他应用领域带来创新。我国在 DNA 合成和高通量测序方面拥有全球领先的技术和平台,为进一步发展 DNA 存储技术奠定了良好的基础,未来应抓住新兴存储技术的发展机遇,前瞻谋划和布局,加大政府支持与资助力度,并建立多层次的资本市场,拓宽融资渠道;密切跟踪全球重要团队的研究进展,重视关键技术的突破,为产业技术进步积累原创资源,促进新

兴产业高质量发展。

其次,促进多学科研究和公私协同合作,加速成果应用转化。生物学、工程学、计算机科学、化学等多学科交叉融合推动了以DNA数据存储为代表的生物存储与生物计算系统等新兴技术的发展,未来还将推动更多前沿颠覆性创新和高成长性产业发展。我国应在高校、科研机构等打造综合交叉学科群,尝试制定跨学科课程体系,培育新兴学科和特色学科,构建融合创新的科研平台与育人平台,促进多学科研究人员之间的交流和国际协作,通过思想碰撞推动创新发展。此外,该技术具有巨大的市场应用前景,应鼓励企业协同参与,对重点企业进行精准扶持引导,不断提高自身的创新能力和科技水平,采取公私合作模式(PPP),促进企业承接公共机构的研发成果,加快重要科技成果的市场化进程。

最后,加强生物技术安全监管重点关注两用技术的生物安全问题。与其他两用性生物技术一样,DNA存储技术也存在生物安全风险:将恶意代码写入DNA片段,在进行数据解码和分析时,相关的计算系统、软件和算法都存在网络生物安全风险;DNA存储在军事领域的应用还会对国家安全造成潜在危害;且由于DNA介质本身的特性,现有的数据管理模式无法适用也会增加数据失窃和恶意篡改的风险。为此应尽早制定应对新兴技术威胁的防御措施与计划,将两用技术相关安全问题纳入生物安全法立法工作的考虑范畴,制定相应法规和市场准则,构建相应的国家安全预警系统,开发新型监测和监管设备,提高科学监管和防控能力。

#### 参考文献

[1] ZHIRNOV V, ZADEGAN R M, SANDHU G S, et al. Nucleic acid memory [J]. Nature Materials,

2016,15(4):366-370.

[2] BENENSON Y. Biocomputers: from test tubes to live cells [J]. Molecular BioSystems, 2009, 5(7): 675-685.

[3] EXTANCE A. Could The Molecule Known For Storing Genetic Information Also Store The World's Data? [J]. Nature, 2016, 537(7618): 22-24.

[4] ORLANDO L, GINOLHAC A, ZHANG G J, et al. Recalibrating Equus Evolution using the Genome Sequence of an Early Middle Pleistocene Horse [J]. Nature, 2013, 499(7456): 74-78.

[5] WIENER N. Machines Smarter than Men? Interview with Dr. Norbert Wiener. Noted Scientist [J]. US News & World Report, 1964, 84-86.

[6] NEIMAN M S. On the Molecular Memory Systems and the Directed Mutations [J]. Radiotekhnika, 1965, 6: 1-8.

[7] DAWKINS R. The Blind Watchmaker: Why the Evidence of Evolution Reveals a Universe Without Design [M]. WW Norton & Company. 1996.

[8] DAVIS J. Microvenus [J]. Art Journal, 1996, 55(1): 70-74.

[9] CLELLAND C T, RISCA V, BANCROFT C. Hiding Messages in DNA Microdots [J]. Nature, 1999. 399(6736): 533-534.

[10] LEIER A, RICHTER C, BANZHAF W, et al. Cryptography with DNA Binary Strands [J]. BioSystems, 2000, 57(1): 13-22.

[11] BANCROFT C, BOWLER T, BLOOM B, et al. Long-term Storage of Information in DNA [J]. Science, 2001, 293(5536): 1763-1765.

[12] WONG P C, WONG K, FOOTE H. Organic Data Memory Using the DNA Approach [J]. Communications of the ACM, 2003, 46(1): 95-98.

- [13] YACHIE N, SEKIYAMA K, SUGAHARA J, et al. Alignment-Based Approach for Durable Data Storage into Living Organisms[J]. *Biotechnology progress*, 2007, 23(2): 501-505.
- [14] PORTNEY N G, WU Y, QUEZADA L K, et al. Length-based Encoding of Binary Data in DNA [J]. *Langmuir*, 2008, 24(5): 1613-1616.
- [15] AILENBERG M, ROTSTEIN O D. An improved Huffman Coding Method for Archiving Text, Images, and Music Characters in DNA [J]. *Biotechniques*, 2009, 47(3): 747-754.
- [16] GUSTAFSSON C. For Anyone Who Ever Said there's no Such Thing as a Poetic Gene [J]. *Nature*, 2009, 458(7239): 703-703.
- [17] GIBSON D G, GLASS J I, LARTIGUE C, et al. Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome [J]. *Science*, 2010, 329(5987): 52-56.
- [18] ARITA M, OHASHI Y. Secret Signatures Inside Genomic DNA [J]. *Biotechnology Progress*, 2004, 20(5): 1605-1607.
- [19] CHURCH G M, GAO Y, KOSURI S. Next-generation Digital Information Storage in DNA [J]. *Science*, 2012, 337(6102): 1628-1628.
- [20] GOLDMAN N, BERTONE P, CHEN S, et al. Towards Practical, High-capacity, Low-maintenance Information Storage in Synthesized DNA [J]. *Nature*, 2013, 494(7435): 77-80.
- [21] SHIPMAN S L, NIVALA J, MACKLIS J D, et al. Molecular Recordings by Directed CRISPR Spacer Acquisition [J]. *Science*, 2016, 353(6298): aaf1175.
- [22] 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: 345-349.
- [23] 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): 1-12.
- [24] LEE H, WIEGAND D J, GRISWOLD K, et al. Photon-directed Multiplexed Enzymatic DNA Synthesis for Molecular Digital Data Storage [J]. *BioRxiv*. 2020.
- [25] FARZADFARD F, LU T K. Genomically Encoded Analog Memory with Precise in Vivo DNA Writing in Living Cell Populations [J]. *Science*, 2014, 346(6211): 1256272.
- [26] ROQUET N, SOLEIMANY A P, FERRIS A C, et al. Synthetic Recombinase-based State Machines in Living Cells [J]. *Science*, 2016, 353: aad8559.
- [27] FARZADFARD F, GHARAEI N, HIGASHIKUNI Y, et al. Single-Nucleotide-Resolution Computing and Memory in Living Cells [J]. *Molecular Cell*. 2019, 75(4): 769-780.
- [28] BORNHOLT J, LOPEZ R, CARMEAN D M, et al. A DNA-based Archival Storage System [C]// *Proceedings of the Twenty-First International Conference on Architectural Support for Programming Languages and Operating Systems*, 2016, 637-649.
- [29] ORGANICK L, ANG S D, CHEN Y, et al. Random Access in Large-scale DNA Data Storage [J]. *Nature Biotechnology*, 2018, 36(3): 242-248.
- [30] TAKAHASHI C N, NGUYEN B H, STRAUSS K, et al. Demonstration of End-to-End Automation of DNA Data Storage [J]. *Scientific Reports*, 2019, 9

- (1):1-5.
- [31] 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.
- [32] KOCH J, GANTENBEIN S, MASANIA K, et al. A DNA-of-things Storage Architecture to Create Materials with Embedded Memory [J]. *Nature Biotechnology*, 2020, 38(1):1-5.
- [33] ERLICH Y, ZIELINSKI D. DNA Fountain Enables a Robust and Efficient Storage Architecture [J]. *Science*, 2017, 355(6328):950-954.
- [34] NFSA. Technological World First for NFSA [EB/OL]. (2020-09-26). <https://www.nfsa.gov.au/latest/nfsa-dna>.
- [35] DARPA. Molecular Informatics [EB/OL]. (2017-04-03). <https://www.darpa.mil/program/molecular-informatics>.
- [36] NSF. New NSF Awards Support the Creation of Bio-based Semiconductors [EB/OL]. (2018-07-16). [https://www.nsf.gov/news/news\\_summ.jsp?cntn\\_id=295968&org=NSF](https://www.nsf.gov/news/news_summ.jsp?cntn_id=295968&org=NSF).
- [37] IARPA. Molecular Information Storage (MIST) [EB/OL]. (2018-01-15). <https://www.iarpa.gov/index.php/research-programs/mist?id=1077>.
- [38] SRC. 2018 Semiconductor Synthetic Biology Roadmap [EB/OL]. (2018-10-02). [https://jakebeal.github.io/Publications/SSB-Roadmap2018\\_1stEdition.pdf](https://jakebeal.github.io/Publications/SSB-Roadmap2018_1stEdition.pdf).
- [39] DARPA. Creating Technology Breakthroughs and New Capabilities for National Security [EB/OL]. (2019-08-31). <https://www.darpa.mil/attachments/DARPA-2019-framework.pdf>.
- [40] NSF. Semiconductor Synthetic Biology for Information Storage and Retrieval (SemiSynBio-II) [EB/OL]. (2020-02-14). <https://www.nsf.gov/pubs/2020/nsf20518/nsf20518.htm#elig>.
- [41] DARPA. SBIR: HR001120S0019-02 Rapid, Flexible Manufacturing of DNA Molecules for Synthetic Biology and Therapeutic Applications [EB/OL]. (2020-04-09). <https://www.darpa.mil/work-with-us/for-small-businesses/HR001120S0019-02>.
- [42] Office of the Director of National Intelligence. IARPA Announces Launch of the Molecular Information Storage Program [EB/OL]. (2020-01-15) <https://www.dni.gov/index.php/newsroom/press-releases/item/2086-iarpa-announces-launch-of-the-molecular-information-storage-program>.
- [43] Asia-Pacific Biotech News. Synthetic Biology in Japan [EB/OL]. (2014-05-31). <https://www.asiabiotech.com/18/1805/18050031x.html>.
- [44] 中华人民共和国科学技术部. 日本开发出新的DNA解析技术 [EB/OL]. (2016-10-20). [http://www.most.gov.cn/gnwkjdt/201610/t20161019\\_128284.htm](http://www.most.gov.cn/gnwkjdt/201610/t20161019_128284.htm). Ministry of Science and Technology of the People's Republic of China. Japan has Developed a New DNA Analysis Technology [EB/OL]. (2016-10-20). [http://www.most.gov.cn/gnwkjdt/201610/t20161019\\_128284.htm](http://www.most.gov.cn/gnwkjdt/201610/t20161019_128284.htm).
- [45] 中国科学院深圳先进技术研究院. 深圳先进院等发起成立国际合成生物设施联盟 [EB/OL]. (2019-05-10). [http://www.siat.ac.cn/yw2016/201905/t20190510\\_5292826.html](http://www.siat.ac.cn/yw2016/201905/t20190510_5292826.html). Shenzhen Institute of Advanced Technology Chinese Academy of Sciences. Shenzhen Advanced

- Institute Initiated the Establishment of the International Union of Synthetic Biological Facilities [EB/OL]. (2019-05-10). [http://www.siat.ac.cn/yw2016/201905/t20190510\\_5292826.html](http://www.siat.ac.cn/yw2016/201905/t20190510_5292826.html).
- [46] CSIRO. The Big Six; CSIRO's Plans for Our Future [EB/OL]. <https://www.csiro.au/en/News/News-releases/2016/The-big-six-CSIROs-plans-for-our-future>.
- [47] CSIRO. Synthetic Biology [EB/OL]. (2020-04-30). <https://www.csiro.au/en/Research/LWF/Areas/FSPs/Synthetic-Biology>.
- [48] 南方科技大学新闻网. 我校作为首席科学家牵头单位获批国家重点研发计划项目 [EB/OL]. (2019-07-15). <https://newshub.sustech.edu.cn/zh/html/201907/24169.html>.  
News SUSTech. SUSTech Scientist Approved as Chief Scientist for National Key R&D Project [EB/OL]. (2019-07-15). <https://newshub.sustech.edu.cn/html/201907/15015.html>.
- [49] 中国科学院深圳先进技术研究院. 深圳先进院2020年度牵头获批7项国家科技部重点研发计划项目 [EB/OL]. (2020-12-09). [http://www.siat.cas.cn/xshd2016/202012/t20201209\\_5813871.html](http://www.siat.cas.cn/xshd2016/202012/t20201209_5813871.html).  
Shenzhen Institute of Advanced Technology Chinese Academy of Sciences. Shenzhen Advanced Institute Took the Lead in Approving 7 Key R&D Projects of the Ministry of Science and Technology in 2020 [EB/OL]. (2020-12-09). [http://www.siat.cas.cn/xshd2016/202012/t20201209\\_5813871.html](http://www.siat.cas.cn/xshd2016/202012/t20201209_5813871.html).

#### 作者贡献说明

宋 琪:文章框架设计,材料搜集,检索式制定及初稿撰写;

丁陈君:政策规划和总结建议撰写,文章修改;

吴晓燕:专利分析;

陈 方:文章修改。

## 美国成立国家人工智能计划办公室

2021年1月12日,美国白宫科学技术政策办公室(OSTP)宣布成立国家人工智能计划办公室(National Artificial Intelligence Initiative Office)。它将成为联邦政府协调和沟通美国各界AI研究和决策的中心枢纽,督导和实施美国国家AI战略,进一步促进美国的AI研发,确保美国未来数年内AI这一关键领域的领导地位。

该办公室依据最近通过的《2020年国家人工智能计划法案》设立。该法案显示了两党对美国政府在AI领域长期努力的大力支持,并将美国政府现有的许多AI政策和举措编入法律并加以扩展。例如,将美国AI计划确立的5项关键任务纳入法律,扩大2018年成立的人工智能专责委员会并使其成为常设机构,承认2020年成立的国家AI研究所的合法地位,规定要对2019年发布的国家AI研发战略规划进行定期更新,将白宫2019年指导的关键AI技术标准活动扩展至包括AI风险评估框架等。

张娟(中国科学院成都文献情报中心)编译自

<https://trumpwhitehouse.archives.gov/briefings-statements/white-house-launches-national-artificial-intelligence-initiative-office/>