

特约评述

DOI: 10.12211/2096-8280.2021-040

从全球专利分析看DNA合成与信息存储技术发展趋势

陈大明¹, 张学博¹, 刘晓¹, 马悦^{1,2}, 熊燕^{1,2}⁽¹⁾ 中国科学院上海生命科学信息中心, 中国科学院上海营养与健康研究所, 上海 200031; ⁽²⁾ 中国科学院大学, 北京 100049)

摘要: 随着大数据技术的发展以及数据的快速增长, DNA作为信息存储的介质, 受到诸多研究机构和企业的高度重视。DNA存储技术发展空间和市场潜力巨大。有预测认为, 到2024年将有约30%的数字业务开始尝试用DNA存储信息。在巨大的市场需求面前, 与之相匹配的专利布局成为在竞争中赢得主动的先决条件。本文从专利的角度, 梳理分析了DNA合成与存储技术的发展趋势。要实现DNA的合成及存储, 寡核苷酸或多核苷酸的合成是“写”的根本, 目前, 全球寡核苷酸或多核苷酸合成技术已经经历了三代发展历程, 第一代和第二代均采用亚磷酰胺化学合成法, 第三代合成以酶促合成为原理。根据全球专利分析显示, 每年所公开的专利数量快速增长, 一批新创立企业加入到第二代合成技术的开发中。随着利用末端脱氧核糖核苷转移酶等聚合酶的酶促合成技术的开发, 一批新专利权人着手布局第三代技术的专利。同时, DNA存储所需的核酸组装、序列设计、信息存储等相关技术也发展迅速, 这些技术相对应的专利权人呈现出明显的跨界融合特征, 微软、英特尔、华为等信息技术企业纷纷加入到DNA信息存储技术专利的竞争与合作之中。可以预见的是, 未来的DNA存储技术必然涉及化学、材料、生物、信息、机械、电子等多领域的技术融合, 而技术的进一步融合或将推动DNA存储领域呈现出类似“摩尔定律”技术定期升级的发展路径。

关键词: DNA合成; DNA信息存储; 专利; 知识图谱

中图分类号: Q-1; Q819 **文献标志码:** A

A global patent analysis: trends in DNA synthesis and information storage

CHEN Daming¹, ZHANG Xuebo¹, LIU Xiao¹, MA Yue^{1,2}, XIONG Yan^{1,2}⁽¹⁾ Shanghai Information Center for Life Sciences, Shanghai Institute of Nutrition and Health, Chinese Academy of Sciences, Shanghai 200031, China; ⁽²⁾ University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: With the rapid growth of digital data production, there is a strong motivation for developing new data storage media. DNA, as a medium for data storage, has attracted the attention of many research institutions and enterprises. DNA based storage technology is expeditiously evolving and has great market potential. It is predicted that, by 2024, about 30% of digital businesses will begin to try to use DNA for data storage. To meet the huge market demand, a matching patent scheme has become extraordinarily important for winning the initiative in the competition. This article analyzes the

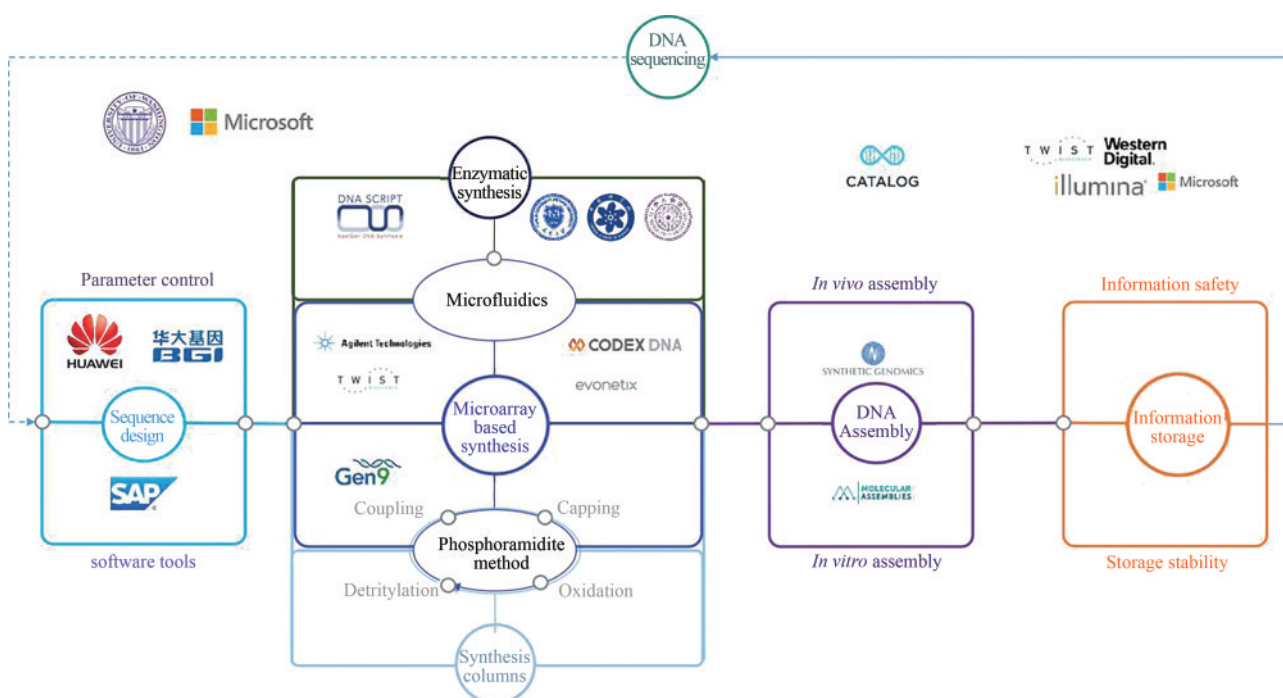
收稿日期: 2021-04-01 修回日期: 2021-05-06

基金项目: 国家重点研发计划“合成生物学”重点专项(2020YFA0908600); 国家自然科学基金(NFSC)-中国科学院(CAS)联合项目“合成生物学发展战略研究(2021-2035)”(XK2019SMA002)

引用本文: 陈大明, 张学博, 刘晓, 马悦, 熊燕. 从全球专利分析看DNA合成与信息存储技术发展趋势[J]. 合成生物学, 2021, 2(3): 399-411

Citation: CHEN Daming, ZHANG Xuebo, LIU Xiao, MA Yue, XIONG Yan. A global patent analysis: trends in DNA synthesis and information storage [J]. Synthetic Biology Journal, 2021, 2(3): 399-411

development trend of DNA synthesis and storage technology from the perspective of patent analysis. We have searched and obtained 1833 patents related to DNA synthesis and storage on a global scale (excluding gene sequencing patents required for DNA storage technology, also excluding DNA synthesis patents dedicated to diagnosis, treatment and other applications) by comprehensively using keywords, international patent classification, patentee, inventor search and other methods. Based on individual reading and comparison, we have utilized patent value analysis, citation analysis, cluster analysis, technical efficacy analysis and other methods to select representative patents that are expected to provide references for further investigations, patent layout and operation decisions in this field. To achieve the goal of using DNA to store digital information, the synthesis of oligonucleotides or polynucleotides is the basis of "writing". So far, the global oligonucleotide or polynucleotide synthesis technology has experienced three generations of development. The first and second generations use phosphoramidite chemical synthesis method, while the third generation is based on the principle of enzymatic synthesis. Global patent analysis has revealed that the number of patents published each year greatly increases while the synthesis technology of oligonucleotides or polynucleotides evolves to phosphoramidite chemical with combination of microarray-based chip technology. Many established companies have joined the development of the second generation of synthetic technology. The emergence of the third generation synthesis technology, which is based on the use of polymerases such as terminal deoxyribonucleoside transferase, is also reflected by the number of patents. Meanwhile, techniques required for DNA based storage, such as nucleic acid assembly, sequence design, and information storage, are also rapidly developing. The patentees corresponding to these technologies have shown obvious cross-industry integration features, since companies such as Microsoft, Intel, and Huawei have successively joined the competition and cooperation of DNA storage patents. It is foreseeable that in the future DNA based storage technology will inevitably involve the convergence of technologies in multiple fields such as Chemistry, Materials, Biology, Informatics, Mechanics, and Electronics. The further integration of these technologies would promote a regularly upgrading development path similar to "Moore's Law" in the field of DNA storage. Based on high-throughput, high-efficiency, high-fidelity and low-cost DNA synthesis, using comprehensive information encoding and decoding, integrating "edit"- "write"- "read"- "dissolve" functions, DNA storage system in the future will become a truly "usable" solution.



Keywords: DNA synthesis; DNA based information storage; patent; knowledge map

随着信息量的指数式增长, 开发新型的数据存储技术已成为各界关注的焦点。磁芯存储器在20世纪中期曾用于早期计算机, 但由于随着其体积小性能会变得不稳定, 导致后来为半导体存储品所取代。如今, 硅基集成电路已经极为常用, 但其数据存储密度仍然无法满足高速增长需求。为此, 也有研发者试图以纸为基板, 在其上集成超小型半导体元件, 但其相对于硅基集成电路的改进主要体现在成本和环保优势方面, 仍然无法满足万物互联时代的高密度、小体积的存储要求。在人类已研发的存储介质中, DNA作为存储介质具有高密度、高稳定性、高保密性、小体积、易拷贝、可并行访问、强兼容性的优点。

DNA合成和存储技术涉及生物学、化学、信息科学、机械技术、电子技术、软件技术等领域的集成, 系统揭示其发展脉络及技术谱系, 有利于更有效地攻克信息存储难题。专利文献集中体现了发明人的智慧, 对其进行系统检索、鉴别、统计、分析和研究, 可以较为完整地厘清全球DNA合成和存储技术的发展现状, 分析该领域的主要专利申请人(专利权人)的技术布局, 识别已有的技术路线、未来的发展潜力等。本文作者在前期合成生物学专利分析^[1]基础上, 聚焦DNA合成和存储技术, 综合利用关键词、国际专利分类、专利权人、发明人检索等方法, 检索并筛选出全球范围内DNA合成和存储相关的1833件专利(不包括DNA存储技术所需的基因测序专利, 也不包括专用于诊断、治疗等其他应用的DNA合成专利), 在逐一阅读和比对的基础上, 综合运用专利价值分析、引证分析、聚类分析、技术功效分析等方法从中筛选出具有代表性的专利, 以期为该领域的研究开发、专利布局和运营等决策提供参考。

1 DNA合成与存储技术的发展阶段

20世纪50年代, 利用磷酸二酯法实现了寡聚二核苷酸的合成^[2], 在此基础上不断完善了寡核苷酸合成方法^[3-6]。20世纪80年代开发的基于亚磷酸胺的DNA合成法^[7]为DNA合成仪的创制奠定了基础。早期的DNA合成, 主要用于分子生物学

研究, 应用范围相对较窄。21世纪以来, 随着合成生物学的发展^[8], DNA合成受到关注。不过, 在21世纪前十年, DNA合成的应用中, 面向DNA存储的专利布局有限。究其原因, 早期的柱式合成法不仅成本高^[9], 而且合成通量较低、合成的错误率也较高^[10]。为降低DNA合成的成本, 研发人员试图采用基于微阵列的芯片来合成DNA^[11]。在此之前, 微阵列主要应用于生命科学和医学的分析仪器中^[12-13], 这也为后来的第二代合成仪的开发提供了启示。并且以此为基础, 开发了光刻合成、电化学脱保护合成、喷墨打印合成这三种芯片式原位合成技术^[14-16]。虽然这些合成方法在化学原理上仍是亚磷酸胺化学合成法, 但由于借鉴了半导体领域的工艺, 使合成通量实现大幅提高。其中, 喷墨打印技术又因其高通量、高效率、低成本成为最受关注的焦点, 并极大推动了DNA合成的发展, 相关专利数量也大幅提升。2013年, 美国半导体研究联盟启动“半导体合成生物学”(semiconductor synthetic biology, SemiSynBio)计划^[17], 使半导体与合成生物学的交叉融合进一步加深。2018年美国发布“半导体合成生物学路线图”^[18], 进一步推动半导体与合成生物学的融合“会聚”。在此背景下, DNA合成与存储的专利数量快速增长(图1), 一方面是由于半导体技术在DNA合成中的应用, 另一方面是投资者看好DNA存储可以作为半导体存储器的互补技术。此时, 亚磷酸胺化学合成法的进一步改进是重点, 而末端脱氧核糖核苷转移酶(terminal deoxynucleotidyl transferase, TdT)等聚合酶的酶促合成法^[19]的研发, 在实现其合成效率可较亚磷酸胺化学合成法高一个数量级的同时, 也促使相关专利布局渐次展开。

在这一历程中, 哈佛大学威斯生物启发工程研究所的Church团队^[20]做了很多开创性的工作, 其于2012年首次利用DNA存储了多媒体文件(专利申请号US15/970 257)。该团队于2016年构建了基于基因组编辑的分子记录器, 使人的细胞实现按时间顺序获得数字和存储数字信息^[21]。2017年, 该团队又进一步将黑白图像、电影短片的信息存储至活细胞的基因组^[22]。在哈佛大学团队的开创性工作发表后不久, 欧洲生物信息研究所(EBI)

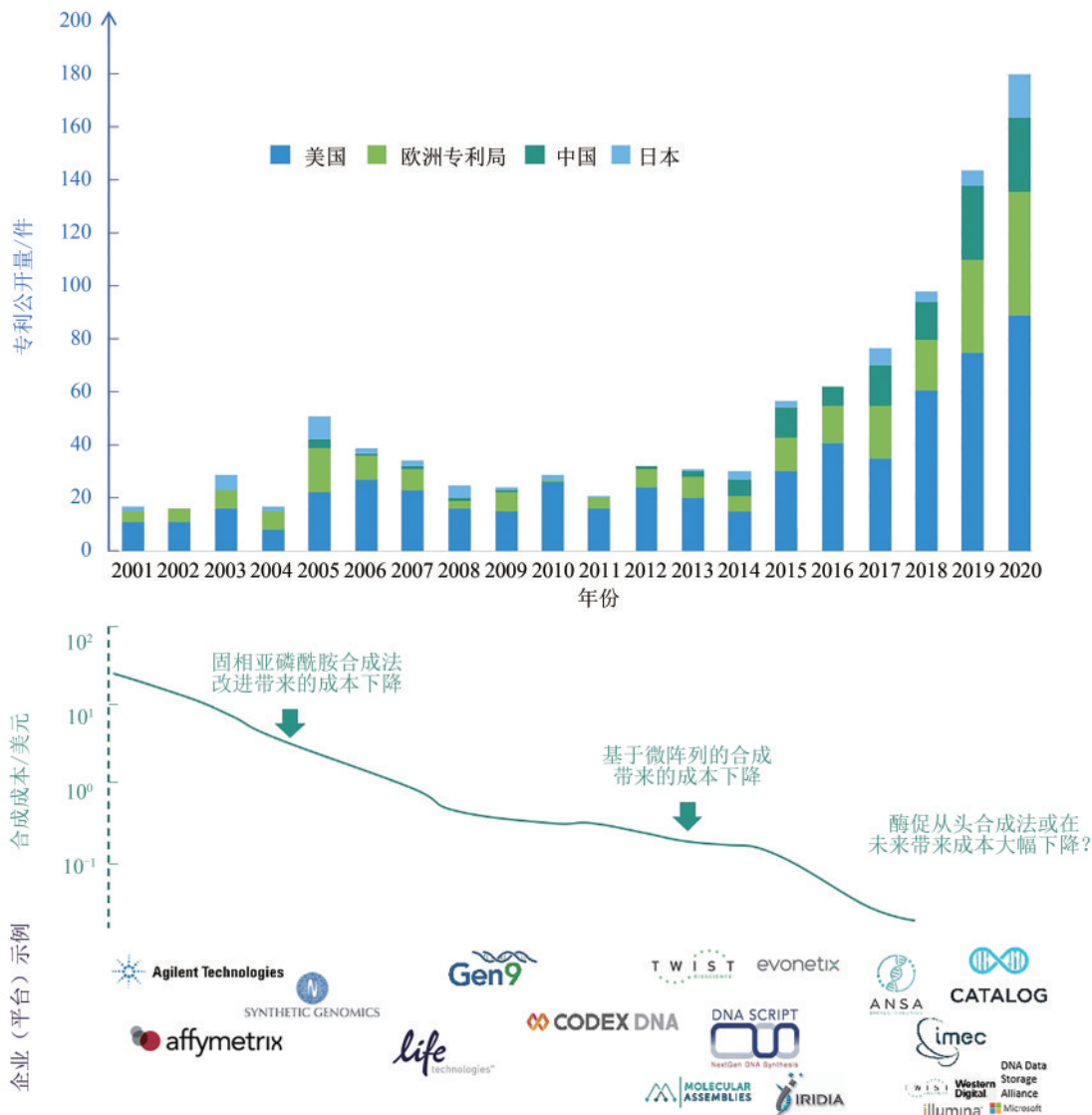


图1 2001—2020年公开的DNA合成与存储专利数量、代表性专利权人及合成成本的变化

[专利的检索日期为2021年3月1日，图中对应年份为专利公开年，所示的国家（地区）的公开量为该区域对应的知识产权局公开的专利量；图中所示的企业（或机构平台），以对应阶段加入DNA合成与存储技术开发的部分企业为代表]

Fig. 1 Patent publication numbers and representative patentees of DNA synthesis and storage, and the cost changes for gene synthesis during 2001—2020

[①The patent search date is March 1, 2021; The corresponding year in the figure is the patent publication year, and the publication amount of the country (region) shown is the amount of patents published by the corresponding Intellectual Property Office in the corresponding region; ②The enterprises (or institutional platforms) shown in the figure are represented by those who have joined the development of DNA synthesis and storage technologies at the corresponding stage]

的团队于2013年利用DNA存储了多媒体文件的同时，还引入了纠错机制，实现了完整的解码^[23]（专利申请号EP13728990.6）。此后，诸多的高校和研究机构也加入了基于DNA的信息存储研究行列，例如哥伦比亚大学、纽约基因组中心的团队利用喷泉码作为容错纠错算法，建立二进制和碱基的映射关系，弥补了因分段序列拼接需要设置

重叠部分的局限，降低了冗余度，提升了存储的密度^[24]。

DNA存储的巨大优势，不仅吸引了研究机构 and 一批创业者加入，也使微软、英特尔、华为等龙头企业将目光投向其中。从企业专利权人分析，美国昂飞公司（Affymetrix）作为生物芯片的早期开发者之一，较早布局了一些相关专利，而安捷

伦 (Agilent)、合成基因组公司 (Synthetic Genomics) 等专利权人也是该领域的积极参与者。其中, 安捷伦是最早利用喷墨打印来合成寡核苷酸的开发者, Twist 生物科学 (Twist Bioscience) 公司则进一步拓展了该技术, 使芯片上可合成的基因座数量增加到数千个, 引领了第二代基因合成仪的发展。在酶促合成方面, 成立于 2013 年的美国分子组装 (Molecular Assemblies) 和成立于 2014 年的法国 DNA Script 公司, 都是酶促合成的代表性专利权人。近年来, 微软等信息技术企业也开始研发 DNA 存储技术, 例如, 微软和华盛顿大学不仅合作申请了很多专利, 还共同开发了用于 DNA 存储数据的全自动系统。2020 年, Twist 生物科学、因美纳 (Illumina)、西部数据 (Western Digital)、微软四家公司也通过成立联盟, 共同推进 DNA 存储的发展。

2 DNA 合成与存储技术的专利布局 and 知识图谱

2.1 专利布局

DNA 作为数据存储的介质, 其合成技术是其中的重要技术之一, 而寡核苷酸或多核苷酸的合成又是长片段 DNA 合成和组装的基础。21 世纪初, 安捷伦等专利权人开发的技术为喷墨打印合成 DNA 奠定了基础, 其所开发的原位合成技术,

涉及将固体支持物与核苷酸之间的官能团化、“脱保护、偶联、盖帽和氧化”循环的展开、寡核苷酸阵列的切割等方面。近十年来, 随着技术的不断发展, 专利布局所保护的客体已从寡核苷酸合成转到多核苷酸合成, 酶促合成的相关专利也逐渐增多 (表 1)。在这过程中, 也涉及一批表面图案化的官能团配套试剂的研发, 例如活化剂包括但不限于 *N*-(3-三乙氧基硅烷基丙基)-4-羟基丁酰胺 (HAPS), 11-乙酰氧基十一烷基三乙氧基硅烷、*N*-癸基三乙氧基硅烷、(3-氨基丙基)三甲氧基硅烷、(3-氨基丙基)三乙氧基硅烷、3-缩水甘油氧基丙基三甲氧基硅烷 (GOPS)、3-碘丙基三甲氧基硅烷、丁基醛基三甲氧基硅烷、(3-氨基丙基)-二乙氧基甲基硅烷、(3-氨基丙基)-二甲基乙氧基硅烷、(3-氨基丙基)-三甲氧基硅烷、(3-缩水甘油氧基丙基)-二甲基乙氧基硅烷、缩水甘油氧基三甲氧基硅烷、(3-巯基丙基)-三甲氧基硅烷、3,4-环氧环己基乙基三甲氧基硅烷、烯丙基三氯硅烷、7-辛-1-烯基三氯硅烷或双(3-三甲氧基硅丙基)胺等; 钝化剂包括但不限于全氟辛基三氯硅烷、三氟辛基三乙氧基硅烷 (FOS)、叔丁基-[5-氟-4-(4,4,5,5-四甲基-1,3,2-二氧戊环-2-基)吡啶-1-基]-二甲基硅烷、全氟辛基三氯硅烷、全氟辛基二甲基氯硅烷、全氟癸基三乙氧基硅烷、全氟辛基三乙氧基硅烷、全氟辛基三甲氧基硅烷、辛基氯硅烷、二甲基氯辛基硅烷、甲基二氯辛基硅烷、三氯辛基硅烷、三甲基辛基硅烷、三乙基辛基硅烷、十八烷基三氯硅烷等。

表 1 寡核苷酸或多核苷酸合成的代表性专利

Tab. 1 Representative patents of oligonucleotide or polynucleotide synthesis

代表性专利 ^① (申请号)	申请年	技术内容	当前申请(专利权)人
US09/779389	2001	减少合成过程中臭氧相关的降解	昂飞
US10/355433	2003	合成期间的黏度控制	安捷伦
US10/652063	2003	功能化载体(通过三芳基甲基连接基团与固体支持物结合)	安捷伦、科罗拉多州立大学
US10/652054	2003	利用环外胺三芳基甲基保护基的前体, 促进核苷酸间键的形成	安捷伦、科罗拉多州立大学
US10/652064	2003	多核苷酸的前体	安捷伦、科罗拉多州立大学
US11/118599	2005	基材和合成试剂	安捷伦、科罗拉多州立大学
EP2005251944	2005	原位合成多核苷酸的循环, 包括单体附着、官能团化产生步骤	安捷伦
US11/284495	2005	寡核苷酸阵列的切割	安捷伦
US11/203328	2005	控制从阵列释放的寡核苷酸序列的方法(利用硅烷醇作为可切割的接头, 所有释放的单链寡核苷酸的 3' 和 5' 末端都不适合作为酶促反应的底物)	安捷伦

续表

代表性专利 ^① (申请号)	申请年	技术内容	当前申请(专利权)人
US11/469405	2006	用于新寡核苷酸的脱三苯甲基化的装置	安捷伦
US11/899828	2007	α -效应亲核试剂的保护(避免酸诱导的脱嘌呤)	安捷伦
US11/897898	2007	硫醚取代的芳基碳酸酯保护基团	安捷伦、科罗拉多州立大学
US13/670220	2012	用于寡核苷酸合成或纯化的试剂	安捷伦
PCT/IL2014/050938	2014	液滴包含基因合成试剂,在容器内的位置通过电压源控制来确定,通过依次确定液滴位置来执行基因合成过程	Yeda 研发有限公司
US15/304701	2015	酶促添加核苷酸-纯化-酶促扩增的循环	DNA Script 公司
EP2015807404	2015	用乙酰氧基-甲基保护非天然核苷酸,将保护的非天然核苷酸掺入固体支持物的核苷酸序列中,在非天然核苷酸中除去乙酰氧基甲基	安捷伦
US15/742237	2016	在酶促合成核酸序列的过程中,利用受保护的核苷酸与通用模板结合使用以指导合成	赛默飞世尔
US15/051511	2016	偶联寡核苷酸的方法:使用试剂活化第一寡核苷酸的末端,将其与第二寡核苷酸的末端缀合,以产生酰胺键连接	安捷伦
US15/814826	2017	组合两个液滴以促进两种不同寡核苷酸之间的连接,实现高保真度的合成	Gen-9 公司
CN201780061546.X	2017	多核苷酸合成装置中,在固体表面增加面积的纹理(内部区域包含多个凹陷或突起),以增加合成产量	Twist 生物科学公司
US16/099487	2017	在酶促合成中,使用电极来调控反应区和反应位点的 pH,从而调控聚合酶的活性	哈佛大学
SG10202009688Q	2017	含有特定序列的 polX 家族的 DNA 聚合酶,能用于合成没有模板链的核酸分子	DNA Script 公司、巴斯德研究所、法国国家科学研究中心
PCT/FR2017/050215	2017	用于实施酶促合成的设备,包括合成装置和计算机系统,其中合成装置由反应器、试剂供应管、温度调控装置等组成	DNA Script 公司
US16/165952	2018	基板上每个结构与加热单元(包括可寻址电极)接触,通过加热电极使分布于表面的溶剂处于气相,促进多核苷酸与溶剂的缩合	Twist 生物科学公司
US16/479141	2018	通过掺入预定义序列的核苷酸、第二链的核苷酸,来延伸第一链(通过聚合酶的作用)	牛津纳米孔科技公司
US16/636875	2018	利用含有特定序列的 DNA 聚合酶,提高酶促合成的效率	DNA Script 公司
CN201880095407.3	2018	用于在寡核苷酸的合成期间使核苷亚磷酸胺双重偶联的方法,核苷残基的游离羟基与受保护的核苷亚磷酸胺的第一样品偶联,随后暴露于氧化剂,再与受保护的核苷亚磷酸胺的第二样品实现第二偶联,并进一步暴露于氧化剂	安捷伦
CN201980048339.X	2019	在连接酶的作用下,利用预定序列的第一核苷酸与预定序列的第二核苷酸杂交,在裂解位点处裂解双链多核苷酸	牛津纳米孔科技公司
CN201910834783.6	2019	在末端脱氧核糖核苷转移酶的氨基酸序列中特定位点引入突变,提高了其与 3' 羟基末端的核苷酸的偶合效率,从而改进酶促反应合成核酸的方法	中国科学院天津工业生物技术研究所
US16/239453	2019	装置的可寻址电极可控制多核苷酸合成(脱保护、延伸或裂解等)	Twist 生物科学公司
US16/726073	2019	有效洗涤此前合成步骤的残留试剂、溶剂或副产物,以降低多核苷酸合成的错误率	Twist 生物科学公司
PCT/EP2019/087048	2019	利用引发剂、阻滞剂在同一反应容器中合成多种寡核苷酸的方法,并对其加以寡核苷酸的测定	DNA Script 公司
US17/033017	2020	使用通用引物,在聚合酶的作用下实现与模板链互补的合成;错配的异源双链可被核酸内切酶识别并切割,从而减少合成错误	Gen-9 公司
US16/925785	2020	设计具有特定的氨基酸序列的末端脱氧核糖核苷转移酶	DNA Script 公司
US16/854719	2020	多核苷酸合成装置的表面结构含二氧化硅,在表面上设多个凹槽或柱,每个凹槽或柱的宽度为 6.8~500 nm,间距长度约为宽度的 2 倍,深度约为长度的 60%~125%,表面上的多个基因座直径为 0.5~100 μm ,每个簇包含 50~500 个基因座并且具有 0.5~2 mm 的横截面	Twist 生物科学公司
US16/920204	2020	利用单向错配核酸内切酶活性的分子,使核苷酸错配的多个双链核酸分子断裂,去除错配的核苷酸	Codex DNA 公司
PCT/US2020/037104	2020	通过调节金属辅因子的氧化态,调节末端脱氧核糖核苷转移酶(TdT)等聚合酶促进的多核苷酸合成	微软

续表

代表性专利 ^① (申请号)	申请年	技术内容	当前申请(专利权)人
PCT/GB2020/050247	2020	含有特定序列的末端脱氧核糖核苷转移酶	Nuclera Nucleics 公司
PCT/EP2020/053417	2020	多模板的无模板酶促合成	DNA Script 公司
US16/832990	2020	使用纳米流体装置合成寡核苷酸,在多个纳米孔通道中有与多个电极接触的电解质溶液,可通过电压控制从电极处的电解质溶液产生酸	Palogen 公司
US16/921712	2020	用于从头合成高度精确的核酸序列的装置,该装置具有主通道和微通道,其中微通道具有高的表面积与体积比;在微通道中可添加特异的延伸反应试剂的液滴,留出足够的时间进行延伸反应;该设备可降低合成的错误率	Twist 生物科学公司
PCT/GB2020/050558	2020	固体支持物上,在2'端受控的局部脱保护基础上合成寡核苷酸	Nuclera Nucleics 公司
US17/021616	2020	用于合成DNA或RNA的自动化模块化系统和方法	Synthego 公司

①专利族以其中的任一专利号代表。

在寡核苷酸或多核苷酸的合成过程中,除固相亚磷酸胺合成法相关的技术外,还需要喷墨打印相关的设备。要将碱基单体作为“墨水”逐个喷射在芯片上,涉及到喷嘴、试剂吸注、精密定位、图像识别、微流体控制等诸多方面,而液滴的生成、分配和控制更是涉及很多精密技术,因而涉及开发和改进精密加工技术的专利也是必不可少的(表2)。

DNA 组装主要分为体内组装和体外组装,其中体外诊断可在反应试管中、多孔平板中、平板表面上、柱中、微流体管等微流控装置中、毛细

管中等不同场景中进行。近年来,同源重组等体内组装方法也有不少专利布局(表3)。

在将数字化信息转化为可存储的基因序列的同时,如何准确访问多核苷酸序列数据存储系统中的数据、如何在多核苷酸序列的日志中创建时间记录、如何在稳定的环境中长期保存多核苷酸序列、如何保障多核苷酸序列中存储的信息的安全等问题,都需要考虑和应对。微软、哈佛大学等企业和大学在这些领域的专利已有所布局,许多专利也体现出其前瞻性(表4)。

表2 喷墨打印及微流控的代表性专利

Tab. 2 Representative patents of inkjet printing and microfluidics

代表性专利 ^① (申请号)	申请年	技术内容	当前申请(专利权)人
US09/897340	2001	第一硅烷用于降低表面能,第二硅烷用于分子的官能化(用于形成固相合成的初始单体)	安捷伦
US15/860445	2018	将电磁辐射(EMR)施加到表面预定区域,以移除预定区域、能够结合核苷的反应性基团,从而定义不同的寡核苷酸延伸的基因座	Twist 生物科学公司
US11/117884	2005	喷墨打印头溶剂、可寻址核苷酸阵列	安捷伦
US10/932886	2004	喷射器将液滴分配至基板	安捷伦
US15/929030	2018	用于DNA合成装置的芯片,通过芯片控制每个电极	英特尔公司
US11/365770	2006	使用基板旋转进行阵列制备,每种核酸占据支持物的单独已知区域	昂飞
PCT/US2018/041397	2018	合成装置中,利用光学寻址系统选择性地将光传递到孔,以介导或控制孔中的反应	德雷珀实验室(The Charles Stark Draper Laboratory)
HK19122646	2019	用于核酸延伸的柔性材料,包括尼龙、硝化纤维素、聚丙烯、聚碳酸酯、聚乙烯、聚氨酯、聚苯乙烯、缩醛、丙烯酸、丙烯腈、丁二烯苯乙烯、聚对苯二甲酸乙二醇酯、聚甲基丙烯酸甲酯、聚氯乙烯、透明PVC箔、苯乙烯聚合物、含氟聚合物、聚醚砜或聚酰亚胺,在预定位置的表面上沉积核苷	Twist 生物科学公司
EP2004803181	2004	设有用于DNA并行合成的矩阵	合成基因组股份有限公司

①专利族以其中的任一专利号代表。

表3 DNA组装的代表性专利

Tab. 3 Representative patents of DNA assembly

代表性专利 ^① (申请号)	申请年	技术内容	当前申请 (专利权)人
US11/804996	2007	分析靶核酸序列中是否存在可能会破坏多重寡核苷酸组装反应的预定干扰序列,再利用聚合酶来实现核苷酸的组装	Gen-9公司
US14/379005	2013	液滴中的组装	哈佛大学
EP2015818477	2015	定点的DNA切割	Gen-9公司
US14/752773	2015	用于分析确定连续、较大的核酸序列的计算机实现过程	10X基因组学有限公司
EP2015852140	2015	检测错误组装的核酸	Gen-9公司
EP2015747649	2015	末端具有序列互补区的两个双链核酸片段共价连接	生命技术公司
PCT/US2017/050164	2017	利用RNA引导的内切核酸酶,其能够靶向并切割DNA序列,产生特征DNA片段,退火后实现DNA片段的共价连接	麻省理工学院
US16/231117	2018	在高保真度多核苷酸组装中操纵液滴	Gen-9公司
CN201811377712.X	2018	利用酿酒酵母体内组装系统组装和储存长序列,同时设强纠错系统	天津大学
US16/736258	2020	利用细胞提取物、核酸外切酶、单链结合蛋白的蛋白质组合物等来实现高效的核酸组装	赛默飞世尔
US16/879705	2020	利用核酸内切酶、核酸外切酶、聚合酶和连接酶的组合来组装核酸	Twist生物科学公司
PCT/EP2020/060535	2020	在多个反应室中将两个或更多个预定核酸片段多重组装	赛默飞世尔

①专利族以其中的任一专利号代表。

表4 DNA存储的代表性专利

Tab. 4 Representative patents of DNA storage

代表性专利 ^① (申请号)	申请年	技术内容	当前申请 (专利权)人
US15/565372	2016	集核酸合成、核酸测序和计算机设备于一体的存储系统	华盛顿大学
US15/394484	2016	利用控制器对封装有DNA数据的液滴进行控制	英特尔公司
US15/625998	2017	在多核苷酸序列的日志中创建时间记录	微软
US15/427808	2017	提高存储在多核苷酸序列数据存储系统中数据的随机访问的准确性	微软
US15/431897	2017	向多核苷酸添加取代原始校准区域的序列来修饰校准区域	华盛顿大学、微软
PCT/US2017/029751	2017	将真实信息单链核酸、假单链核酸结合形成核酸混合物,从而保障基于核酸的安全通信	哈佛大学
US15/427344	2017	比较引物靶序列和有效载荷序列,使引物靶序列的区域、有效载荷序列的区域有大于阈值的相似性	微软
CN201880026589.9	2018	信息的生物加密或生物解密	Twist生物科学公司
US16/017714	2018	数据存储介质包括干燥产物,保护DNA免受降解	微软
PCT/US2018/056900	2018	确定核苷酸序列,鉴定序列中核苷酸之间的转换或边界,将预定值分配给所鉴定的转变或边界,以产生对应序列中的编码值	哈佛大学
EP2018779622	2018	编码信息的方法中,根据核酸的多态特征生成加密密钥,将加密信息编码成合成DNA	苏黎世联邦理工学院
PCT/GB2019/053669	2019	多核苷酸链移动通过纳米反应器时,对其进行选择性修饰	牛津纳米孔科技公司
PCT/US2019/058414	2019	数据存储介质包括二维(2D)支撑结构(金属箔,玻璃或塑料),人工合成的、编码数字信息的DNA分子置于其上,而整个结构置于非反应性材料(例如二氧化硅)或金属薄层的保护层中,以保护DNA免受降解;同时,设有可寻址的特定数字信息	微软
IL277102	2020	索引信息被添加到多个DNA片段	欧洲分子生物学实验室
PCT/US2020/032384	2020	用优化的数据结构和功能,搜索和提取DNA中存储的数据	Catalog公司

①专利族以其中的任一专利号代表。

在DNA信息存储的过程中,还有更多技术因素需要考虑,例如,序列中鸟嘌呤(G)和胞嘧啶(C)的比例需要保持在合理区间;同时,碱

基重复率也是存储中需要考虑的参数。因而,不少专利也将GC含量和碱基检出算法嵌入其中(表5)。

表5 其他的代表性专利

Tab. 5 Other representative patents

代表性专利 ^① (申请号)	申请年	技术内容	当前申请(专利权)人
PCT/US2016/031674	2016	在核酸扩增产物的有效除去非靶核酸序列(例如引物序列)	Twist生物科学公司
US15/156134	2016	无细胞的核酸克隆	Twist生物科学公司
US15/259420	2016	在DNA合成中,对恶意序列进行检测	SAP公司
US16/098471	2016	对存储中的由GC含量和DNA序列的碱基重复率的参数加以控制,使DNA序列中GC含量为45%~60%,且不超过2个连续的单碱基存在	华大基因
US15/607364	2017	碱基检出算法用于分析原始数据,并可识别碱基在多核苷酸链上的给定位置	微软
CN201810986270.2	2018	在将数字信息转换成核酸序列时,在保障碱基的存储密度及编码效率的前提下,控制GC含量的比例,提高存储的效率	华为技术有限公司
US16/759282	2018	用于预测合成和组装多核苷酸的可行性的软件工具,通过分析GC含量等参数对多核苷酸进行评分	Twist生物科学公司

①专利族以其中的任一专利号代表。

2.2 知识图谱

在合成生物学专利分析的基础上,本文系统梳理了DNA合成与存储技术的专利文献,绘制了相关技术的知识图谱(图2)。在DNA合成与存储技术专利的知识谱系中,设计存储数字信息的核酸序列是基础,在此基础上可采用第一代、第二代或第三代合成技术进行DNA合成。其中,第一代和第二代合成技术均采用亚磷酰胺化学合成法,而第二代中所用的半导体技术或可在未来应用于酶促合成。从各种技术和方法的演进过程可以看到,DNA合成技术在不断迭代中发展,而又保持一定的连续性。总体上,目前的技术迭代沿着合成通量的增加、合成原理的升级两个维度进行,这两个维度的组合必将形成一系列技术,这也是专利布局和运营的起点。

同时,这些技术与组装技术组合,构成长片段DNA的合成和柱状技术。由于核酸组装本身也可有不同的技术组合(例如,利用各种工具酶的方法,可以与基因组编辑的技术组合),因而这种排列组合就为长片段DNA合成的不断升级提供了动力。在此基础上,再重点考虑“写”的速度、成本,“读”的便捷性,以及信息存储的安全性、稳定性等技术要求。“写”与“读”(通常利用先进的测序

技术)的结合,构成了未来DNA存储系统的基本架构。这种架构建立后,随着各个环节或工艺的不断优化,未来或许可以期待DNA存储也呈现出类似“摩尔定律”的定期升级换代。

3 总结与展望

随着人类社会进入海量数据时代,信息存储的重要性已经不言而喻。在巨大的需求面前,DNA存储的优势进一步展现。近十年来,寡核苷酸的合成已经拓展至多核苷酸,在提高寡核苷酸合成的长度和效率、降低合成成本的过程中,全球的企业、大学和研究机构有一系列的专利布局,该领域的专利竞争格局也呈现出交叉融合的态势。

3.1 从专利布局看DNA合成与存储技术的发展

对近20年来的DNA合成与存储的相关技术专利进行分析,可以发现合成通量的提升、从化学方法向酶促合成的生成,是推动合成能力提升的动力。以DNA为介质,其天然性突破了传统的存储技术在寿命限制和低数据密度两方面的局限(DNA数据存储的保存周期长,且不会损坏;1 kg

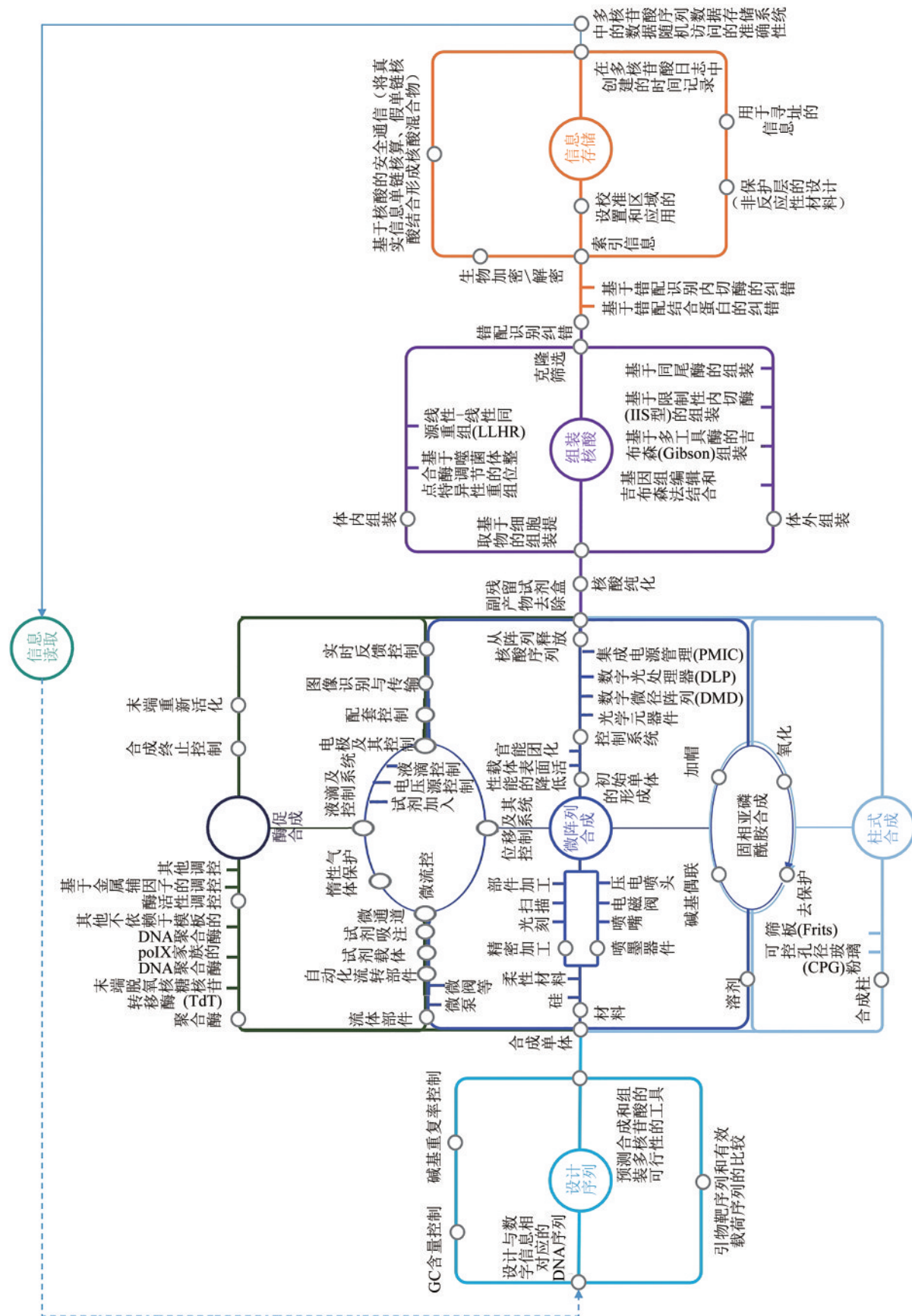


图2 DNA合成与存储技术的知识图谱
Fig. 2 Knowledge map of DNA synthesis and storage

的DNA可以存储海量数据)^[25]，因而DNA合成技术的升级已经使得“写”的速率有所突破，而在“读”方面的基因测序技术本身也在不断升级，并伴随着成本的不断下降，使得DNA存储或呈现出类似数十年前集成电路早期发展的态势。

从技术的角度看，DNA合成仪结合了化学、物理、材料、半导体、生物、流体力学、信息科学等领域，这种学科间的交叉融合促进了技术的不断改进，将进一步带来合成通量、合成长度、合成精度的提升，以及合成成本的降低。在未来，喷墨打印或微流控技术与酶促合成进一步结合，将在通量、可扩展性和成本方面突破现有方法的限制，实现DNA合成的全面升级。

从专利分析来看，尽管基于DNA的信息存储仍处于早期，但仍然可以预见其未来专利将解决的技术功效包括更高的读写速率、更高效的编解码、更低的存储成本、更高的存储稳定性等方面。例如，在存储稳定性方面，长期稳定的存储需要排除热辐射、电离辐射、化学诱变等方面的因素。目前微软在该功效的方向已有一定的技术布局，但未来发展若综合考虑不同的应用场景，或可使用微纳尺度的真空室作为容器。又如，在编码方面，需要同时考虑存储效率、合成难度、加密安全等要求，当前的专利布局已经考虑到尽可能地避免连续的单碱基重复，同时也需要避免高鸟嘌呤和胞嘧啶占比（GC含量）。为了探索更好的编码，也有研发团队在考虑二进制、三进制、四进制模型中的不同编码策略（例如CN201910909594.0）来实现理想的技术功效。随着信息存储的进一步发展，未来或许也可探索多进制模型的混合使用的可能性和可行性。

以高通量、高效率、高保真、低成本的DNA合成为基础，综合信息编解码、基因测序等技术的改进，集“‘编’-‘写’-‘读’-‘解’”于一体的DNA存储系统或将成为真正“用得上”的解决方案。咨询公司高德纳（Gartner）在2020年的预测中认为，随着未来数据的海量增长，到2024年将有约30%的数字业务开始尝试用DNA进行信息存储^[26]。从现有专利权人的视角来看，合成或存储的质量和准确性、单个碱基的平均合成消耗或

成本、合成通量这三大指标是关键。从精度来看，质量和准确性是所有客户的基本要求，而当前的DNA合成和存储技术还需要大量的克隆筛选和纠错，才能产生准确的序列。只有将0.1%左右的碱基对误差率降至更低，才能巩固降低成本、提高生产通量的优势：从合成长度来看，长片段的DNA合成伴随着相对较高的错误率；从合成成本来看，较高的错误率必然导致纠错等各种成本上升。以优化和提升这些技术功效为导向，仍有很大的专利布局空间。随着DNA的信息存储专利将越来越多地涉及到编码、解码、信息寻址、信息安全等的部分，集基因合成、基因测序、计算机控制于一体的DNA存储设备或存储系统的开发也越来越深入，微软等信息技术企业开始大规模地在该领域进行研发。

3.2 从竞争格局看DNA合成与存储专利的布局与运营

在第一代的柱式合成技术领域，美国的ABI、Biolytic、Digilab、BioAutomation和Synthomics等公司，德国的K&A Laborgeraete和PolyGen，以及俄罗斯、日本和中国的企业都开发了相关的产品。总体上看，与第二代合成仪相比，柱式合成技术路径上的专利布局在近年来已经逐渐减少。

在第二代合成技术的开发路径中，昂飞公司侧重于利用物理掩膜法进行光刻合成。考虑到掩膜开发成本较高，罗氏的下属企业也曾尝试无掩膜技术的开发，利用数字化控制、光敏保护基团介导来实现原位合成，但这一技术路径的专利较少。与光刻合成法不同，电化学介导的脱保护合成法利用芯片表面上的微电极阵列，可控制合成链末端的酸敏基团水解来实现脱保护，继而促进合成反应。美国CustomArray公司（原Combimatrix公司）在电化学介导的脱保护合成法开发中，推出过可合成上万种单链的产品，但还存在寡核苷酸合成错误率高、产量低、准确率低、不稳定等缺点，因而后续的开发者不多。喷墨打印是当前DNA合成中的主流方法，早期的专利布局以安捷伦公司的专利最具代表性，而近十年又以Twist生物科学公司的专利布局最为典型。总体上看，Twist生物科学公

公司的专利布局已成体系，而 Evonetix 公司、Synthonics 公司等也有一定数量的专利。然而，要实现高通量、高效率、低成本、低错误率的 DNA 合成和存储，还需要加大技术研发和专利布局。

本文基于现有专利布局所展示的知识图谱，为我国研究机构和企业专利布局提供了参考。总体上看，与美国研究机构和企业相比，中国专利权人在 DNA 合成和存储专利的布局起步相对较晚，高价值专利数量也相对较小。该领域的发展仍处于技术生命周期中的早期发展阶段，因而迎头赶上的空间仍然存在，尤其是未来集基因合成、基因测序、计算机控制于一体的 DNA 存储系统的开发，与信息编解码等技术的交织，将带来相关技术策略、技术功效的“排列组合”，有非常大的挖掘空间。对这些领域，需要系统性地解析其潜力，针对性地提出适合各研究机构或企业的专利战略。

参 考 文 献

- [1] 陈大明, 刘晓, 范月蕾, 等. 从全球专利分析看合成生物学技术发展趋势[J]. 合成生物学, 2020, 1(3): 372-384.
CHEN D M, LIU X, FAN Y L, et al. A global patent analysis: trends in DNA synthesis and information storage[J]. 2020, 1(3): 372-384.
- [2] MICHELSON A M, TODD A R. Nucleotides part XXXII. Synthesis of a dithymidine dinucleotide containing a 3', 5'-internucleotidic linkage[J]. Journal of the Chemical Society, 1955(8): 2632-2638.
- [3] SCHEUERBRANDT G, DUFFIELD A M, NUSSBAUM A L. Stepwise synthesis of deoxyribo-oligonucleotides[J]. Biochemical & Biophysical Research Communications, 1963, 11(2): 152-155.
- [4] KHORANA H G, BÜCHI H, CARUTHERS M H, et al. Progress in the total synthesis of the gene for ala-tRNA[J]. Cold Spring Harbor Symposia on Quantitative Biology, 1968, 33: 35.
- [5] AGARWAL K L, KUMAR A, KHORANA H G. CIX. Total synthesis of the structural gene for an alanine transfer ribonucleic acid from yeast. Synthesis of a dodecadeoxynucleotide and a decadeoxynucleotide corresponding to the nucleotide sequence 46 to 65[J]. Journal of Molecular Biology, 1972, 72(2): 351-371.
- [6] BÜCHI H, KHORANA H G. CV. Total synthesis of the structural gene for an alanine transfer ribonucleic acid from yeast. Chemical synthesis of an icosadeoxyribonucleotide corresponding to the nucleotide sequence 31 to 50[J]. Journal of Molecular Biology, 1972, 72(2): 251-288.
- [7] BEAUCAGE S L, CARUTHERS M H. Deoxynucleoside phosphoramidites-A new class of key intermediates for deoxypoly-nucleotide synthesis[J]. Cheminform, 1981, 22(20): 1859-1862.
- [8] 赵国屏. 合成生物学——革命性的新兴交叉学科,“会聚”研究范式的典型[J]. 中国科学:生命科学, 2015, 45(10): 905-908.
ZHAO G P. Synthetic biology: a revolutionary new interdisciplinary discipline and a typical "convergence" research paradigm[J]. SCIENCE CHINA Life Sciences, 2015, 45(10): 905-908.
- [9] SAINI N, RAMAKRISHNAN S, ELANGO R, et al. Migrating bubble during break-induced replication drives conservative DNA synthesis.[J]. Nature, 2013, 502(7471): 389-392.
- [10] KOSURI S, CHURCH G M. Large-scale *de novo* DNA synthesis: technologies and applications[J]. Nature Methods, 2014, 11(5): 499-507.
- [11] SAAEM I, MA K S, MARCHI A N, et al. *In situ* synthesis of DNA microarray on functionalized cyclic olefin copolymer substrate. [J]. ACS Applied Materials & Interfaces, 2010, 2(2): 491-497.
- [12] KIM C C, JOYCE E A, CHAN K, et al. Improved analytical methods for microarray-based genome-composition analysis[J]. Genome Biology, 2002, 3(11): 0065.
- [13] GUNDERSON K L, STEEMERS F J, LEE G, et al. A genome-wide scalable SNP genotyping assay using microarray technology.[J]. Nature Genetics, 2005, 37: 549-554.
- [14] PAUL S S, TRABELSI H, YASEEN Y, et al. Microbial cell factories engineering for production of biomolecules[M]. Mehshana, Gujarat, India: Academic Press, 2021.
- [15] LI S, ZHAO G, JIN W. Enabling technologies in synthetic biology—DNA synthesis, assembly and editing[J]. Chinese Journal of Biotechnology, 2017, 33(3): 343-360.
- [16] MEISER L C, KOCH J, ANTKOWIAK P L, et al. DNA synthesis for true random number generation[J]. Nature Communications, 2020, 11: 5869.
- [17] SELBERG J, MARCELLA G, ROLANDI M, et al. The potential for convergence between Synthetic Biology and Bioelectronics [J]. Cell Systems, 2018, 7(3): 231-244.
- [18] ZHIRNOV V V, RASIC D. 2018 semiconductor synthetic biology roadmap[R]. Durham: Semiconductor Research Corporation, 2018.
- [19] SONG X P, BOUILLON C, LESCRIER E, et al. Dipeptides as leaving group in the enzyme-catalyzed DNA synthesis[J]. Chemistry & Biodiversity, 2012, 9(12): 2685-2700.
- [20] CHURCH G M, GAO Y, KOSURI S. Next-generation digital in-

- formation storage in DNA[J]. *Science*, 2012, 337(6102): 1628.
- [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(7663): 345-349.
- [23] 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.
- [24] ERLICH Y, ZIELINSKI D. DNA Fountain enables a robust and efficient storage architecture[J]. *Science*, 2017, 355(6328): 950-954.
- [25] Potomac Institute for Policy Studies. The future of DNA data storage[R]. Arlington: Potomac Institute for Policy Studies, 2018.
- [26] PLUMMER D, HILL J, SALLAM R, et al. Top Strategic Predictions for 2020 and Beyond[R]. Stamford: Gartner Inc, 2020.



通讯作者: 熊燕(1967—),女,博士,研究员。研究方向为科技战略情报研究。

E-mail: yxiong@sibs.ac.cn



第一作者: 陈大明(1982—),男,硕士,研究员。研究方向为科技情报、专利分析等。

E-mail: chendaming@sibs.ac.cn