序

DOI: 10.12211/2096-8280.2021-033

DNA数字信息存储：造梦、追梦与圆梦

钱珑，沈玥，元英进，欧阳颀，杨焕明

生物体具有精妙绝伦的信息系统。2010年，美国J. Craig Venter研究院将化学合成的约1 Mb的基因组DNA导入受体细菌，成功启动了世界首个“人造生命”辛西娅。这条基因组携带有46位科学家的姓名和一个专属邮箱地址，诉说着人类作为造物主，设计生命、书写遗传密码的浪漫主义情怀。如今，虽然我们对基因组的奥秘仍一知半解，将人工信息写入DNA分子却已成为触手可及的技术现实。DNA信息存储从狭义上讲，是以线性碱基序列的形式，合成并保存编码任意数字信息的DNA分子；从广义上讲，意味着数字信息与生命信息的物理融合。2018年底，美国国家标准技术研究所、国际半导体研究联盟、美国情报高级研究计划局等联合发布《半导体合成生物学路线图》；2021年5月，我国科技部发布了“十四五”国家重点研发计划“生物与信息融合”（BT与IT融合）重点专项项目申报指南。这代表着世界两大经济体对于以DNA存储为代表的未来颠覆性融合技术的顶层认可。

DNA数字信息存储从艺术尝试走向技术现实，依赖的是DNA合成与DNA测序技术的跨越式发展。当前，已可以实现近GB规模任意格式文件在DNA分子中的稳定存储，并可以在几天内对数据进行恢复。如果说初期的DNA存储仅仅是合成与测序技术的附属品，那么近年来，随着全球信息量的爆发式增长和传统信息存储资源告急，DNA存储的“破局性”价值得到认可，一步跨入了工程化阶段。这座工程学大厦的根基是DNA单碱基分辨率的读写技术，经历几十年更迭的DNA扩增、组装技术，和用于DNA分子封装的创新材料科学技术。DNA存储的另一支柱是数字信息的编码理论；各种通信领域的成熟算法应用于DNA存储的信息压缩与纠错，形成了百家争鸣的局面。

必须注意的是，DNA存储并不是已有技术的生搬硬套。生命系统的信息组织形式与电子信息系统的最大区别在于其并行本质。反应体系中的DNA分子在容纳海量数据的同时，存在着广泛的冗余和串扰效应。DNA信息的写入、复制、分发、寻址、读取等操作均是基于扩散的并行反应。因此，传统的信息操作方式都需要适配生化体系特点；这些研究反过来也将启发我们对于小尺度电子信息系统的认知。本刊集结了国内一线工作者，对DNA存储的技术和理论链条中的各个环节进行了逐一论述与研究成果的分享。中国科学院深圳先进技术研究院戴俊彪团队系统总结了从DNA短片段合成到长片段组装的一系列生化方法及其优缺点，并针对DNA存储的需求，提出了全流程联合优化策略；而东南大学陆祖宏团队则聚焦二代高通量DNA合成技术，对不同工业实现方案进行了详细的横向比较。天津大学齐浩团队针对DNA寡核苷酸库不均一所造成数据的缺失问题，介绍了合成、保存和扩增等过程中新的生化技术手段；南方科技大学蒋兴宇团队则着重综述了DNA存储现有的数据加密和修改策略，由于寡核苷酸库存储与传统存储介质的读写方式不同，加密和修改需要对信息本身进行预设的修饰。在编码方面，天津大学陈为刚团队报道了自主开发的适用于染色体长片段存储和二代大规模并行测序读取的数据编解码方法，巧妙利用长片段DNA载体无需索引和二代测序极低错误率的特点，实现了较高的编码密度。此外，我国学者还特别关注技术标准评价体系的制定，深圳华大生命科学研究院沈玥团队报道了自主搭建的DNA存储的编解码算法系统性评估平台，用以对各种算法的文件适配性、存储稳定性和数据安全性进行全面、定量的对比评估。这些工作彰显了DNA存储的研究热度及其广泛多样的技术领域。通过系统梳理全球公开专利，中科院上海生命科学信息中心熊燕团队详细绘制了过

收稿日期：2021-03-18 修回日期：2021-03-18
引用本文：钱珑，沈玥，元英进，欧阳颀，杨焕明. DNA数字信息存储：造梦、追梦与圆梦[J]. 合成生物学，2021，2(3)：303-304

去二十年来DNA合成与信息存储的技术发展历程，并依此对该领域的技术深化与交织进行了趋势预测。

除了当前研究热点所在的寡核苷酸库存储，DNA存储还存在其他极具潜力的实现方式。在本刊中，天津大学元英进团队针对DNA存储不同实现方式的特点，精彩地将其类比于硬盘、光盘与磁带三大模式，它们分别对应着从大到小的数据规模和从易到难的操作流程。今年初，元英进团队在《国家科学评论》发表基于酵母人工染色体的DNA信息存储工作，为DNA存储的“光盘”模式书写上浓重一笔，在本刊中，上海交通大学樊春海团队对此最新成果进行了特别评述。另一值得注意的是近两年崭露头角的“磁带”模式，即利用基因编辑工具在基因组特定区段上动态写入指定信息，磁带模式可以实时记录发生在细胞内外的生化事件，是与生命系统联结最紧密的人工信息系统。北京大学钱珑团队在综述中集中展望了包括基因组动态写入在内的DNA存储的前沿研究与应用方向。这些研究的兴起指向了超越信息存储的下一个梦想：DNA将成为信息系统与生命系统的桥梁，介导以生命为载体的存算一体结构，通过工程生物学手段，使得传感器、处理器等概念在极低能耗的微小生命单元中得以实现，最终造就基于生命的人工信息系统。

历经造梦、追梦，DNA存储正向着圆梦迈进。但它并不会止步于此，而将不断创造出新的生物技术目标。这是科学与工程的交替式前进，是不同学科交叉融合、协同向前的生动案例，并且极有可能成为生命系统对生产力的直接转化的第一个成熟案例。从DNA存储到生物智造、细胞治疗，再到脑机接口，生命系统与信息世界正在分子、细胞、机体和高级意识的多个层次上进行全面融合。站在生物与信息融合技术的元年，我们或可从现在的技术雏形中窥见未来的生活方式。

中图分类号：Q819　　**文献标志码：**A

钱珑：北京大学前沿交叉学科研究院定量生物学中心助理研究员。研究领域主要包括：(1)生物网络的进化原理；(2)合成生物学元件的自动化挖掘与设计；(3)DNA存储。

沈玥：研究员，深圳华大生命科学研究院合成生物学首席科学家，国家重点研发计划项目首席科学家，广东省高通量基因组测序与合成编辑应用重点实验室副主任。研究领域主要包括：(1)DNA合成使能技术与装备；(2)合成基因组学及其下游应用；(3)DNA数据存储。

元英进：天津大学教授，教育部“珠峰计划”合成生物学前沿科学中心主任，系统生物工程教育部重点实验室主任。研究领域主要包括：(1)合成生物学；(2)生物化工。

欧阳颀：中国科学院院士，北京大学前沿交叉学科研究院定量生物学中心副主任，国家农业转基因生物安全委员会成员。研究领域主要包括：(1) 生物调控网络的动力学研究；(2) 正向与逆向工程方法在合成生物学中的应用；(3) 生物微流系统在定量生物学研究中的应用。

杨焕明：中国科学院院士，中国科学院大学及中国医学科学院教授，华大基因联合创始人。发展中国家科学院(TWAS)院士及美国国家科学院、德国国家科学院、丹麦皇家科学院等外籍院士。研究领域主要包括：(1)基因组学；(2)合成生物学。