

作者: 元英进等 来源: 《国家科学评论》 发布时间: 2021/4/2 16:11:25

选择字号: 小 中 大

### 天津大学元英进团队验证DNA数据存储新方法

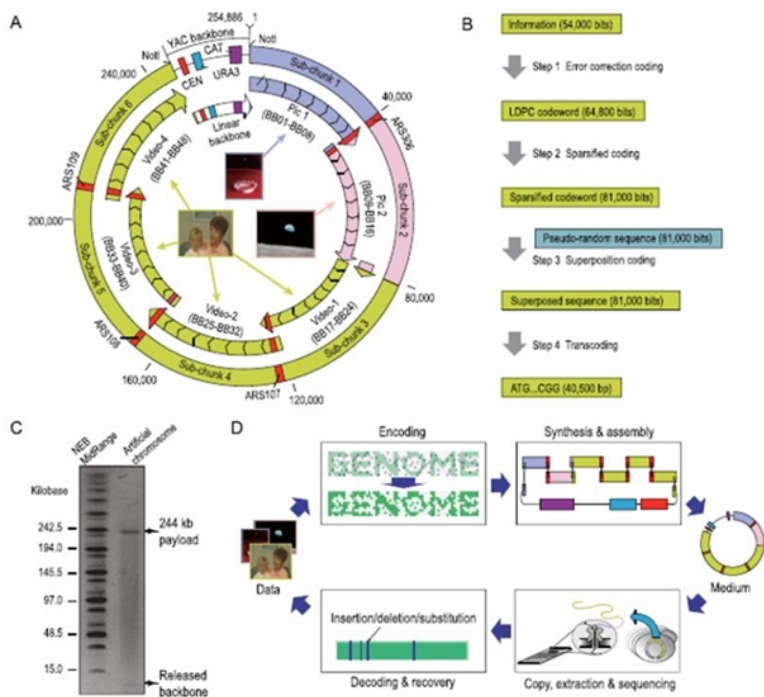


图 (A) 人工染色体示意图; (B) 编码流程; (C) 染色体长度验证 (D) 数据存储工作流程

天津大学元英进教授团队从头编码设计合成了一条长度为254,886 bp、专用于数据存储的酵母人工染色体,借助无线通信中前沿的纠错编码将两张经典图片和一段视频存储于高效组装的人造染色体,利用酵母繁殖实现了数据稳定复制,用便携式的三代纳米孔测序器件实现了数据快速读出与无错恢复。

从医学图像与天文观测,从视频监控到社交网络,全球数字化趋势使得数据量快速增长。据国际数据公司估计,到2025年全球数据总量将达到惊人的175 ZB (1 ZB ≈ 109 TB)。面对快速增长的海量数据,基于磁、光、电等的传统介质的存储技术面临功耗、体积以及使用寿命等限制。近年来,随着合成生物学的快速发展,DNA信息存储由于其高信息密度与低能耗处理等特点,成为应对数据存储发展挑战的新机遇。2021年1月,美国半导体产业协会(SIA)发布的《半导体10年计划》,已将DNA数据存储列为未来海量数据存储的重要选项。

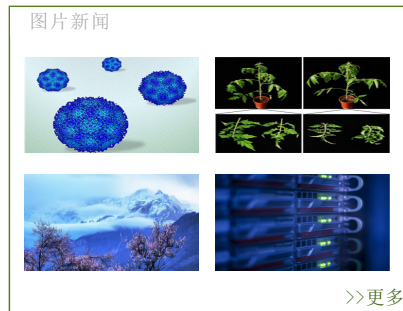
借助团队在酵母人工基因组化学合成领域的积累,天津大学元英进教授带领一个跨学科团队设计合成了一条存储数字信息的酵母人工染色体,存储了两张图片及一段视频,并实现了数据的稳定复制与快速可靠读出。该工作首次将单菌内用于数据存储的DNA碱基数量提升到了百kbp级,容纳数据量为37.8 KB。日前,该研究以“An artificial chromosome for data storage”为题在线发表于National Science Review。天津大学微电子学院青年教师陈为刚副教授、化工学院博士研究生韩明哲以及周见庭副研究员为论文共同第一作者,元英进教授为论文通讯作者。

研究者从染色体的编码设计、组装与稳定复制、数据可靠恢复等方面展示了这种数据存储模式的潜力。

染色体的编码设计中,借助叠加伪随机序列应对三代测序的插入/删除(insertion/deletion)错误,采用现代通信中已广泛验证的低密度奇偶校验(Low-Density Parity-Check, LDPC)码纠正替代错误,实现了纳米孔测序高达10%错误率的数据可靠恢复。

染色体组装与稳定复制方面,设计插入一定数量的酵母自主复制序列(autonomously replicating sequence, ARS),提升的染色体稳定性,支撑高效组装和稳定复制。实验验证染色体稳定复制100代,依然能可靠读出数据。

- 相关新闻 相关论文
- 1 二战留下DNA印记
  - 2 研究揭示组蛋白修饰与DNA复制和修复偶联机制
  - 3 新基因工具可按时序编辑DNA序列
  - 4 科学家发现最古老DNA序列
  - 5 最古老DNA提取纪录诞生!来自冻土中的猛犸象
  - 6 科学家揭示猪肉性状DNA甲基化调控机制
  - 7 《权力的游戏》中的冰原狼是现代狼的远亲
  - 8 科学家探明DNA甲基化如何调控猪肉性状



- 一周新闻排行
- 1 最新!中国科学院院士增选初步候选人名单公布
  - 2 教育部公示2020年学位授权审核结果
  - 3 地球近十万年来最大规模星球碰撞发生在中国依兰
  - 4 科协发布30个重大科学、工程及产业技术问题
  - 5 新一轮“本科扩招潮”来了吗
  - 6 基金委地球科学部公布4项目专业评审组名单
  - 7 美国CDC:德尔塔毒株一传九,需改变抗疫策略
  - 8 科研经费管理放权!国务院有关政策福利来了
  - 9 振而不“兴”西部高校需要哪些“强干预”
  - 10 这种“活化石”野桃靠实力称雄青藏高原

- 编辑部推荐博文
- 人工智能的七宗罪
  - 2021年夏季青藏高原考察:乃夏村采集
  - 从研究生到人才:培一程、陪一程、送一程
  - 杂技术学之一:顶技
  - 博士生身份界定:学生vs.科研工作者
  - 世界顶尖科学家谈创新活动:感想与启示
- 更多>>

数据读出过程中，利用三代纳米孔测序在大约10分钟获得足够的原始读段后，结合研究者设计的生物信息学与纠错译码混合流程，原始图片及视频可以从高错误率约10.79%的原始读段中可靠恢复，所需测序覆盖度仅为16.8x。

该存储模式中，数据逻辑密度（包含载体）为1.19 bit/bp，与目前文献中指标最高的四进制编码DNA喷泉方案相当（Erlich and Zielinski, 2017）。文章进一步指出了该存储模式与传统光盘存储的相似性，也即一次写入，多次读出，低成本可靠复制和便携式读取。作者还提及了降低合成成本、构建多条人工染色体存储更多数据的可行性。（来源：科学网）

相关论文链接：<https://doi.org/10.1093/nsr/nwab028>

打印 发E-mail给:

关于我们 | 网站声明 | 服务条款 | 联系方式 | 中国科学报社 京ICP备07017567号-12 京公网安备 11010802032783

Copyright © 2007-2021 中国科学报社 All Rights Reserved

地址：北京市海淀区中关村南一条乙三号

电话：010-62580783