

DNA 能当“硬盘”用？

DNA,携带生命核心“代码”的神秘大分子。它有着经典而美丽的双螺旋结构,犹如无限绵延的莫比乌斯带,编织出生命体演化的壮阔历史。生命体规模巨大的遗传信息就在此进程中代代传承,精准而安稳,从信息科学的角度看,确实不可思议——单个人体细胞的平均直径仅有5到200微米,却轻松包罗一个人的全部遗传信息,30亿对碱基。

DNA可以保存多久?最新答案是200万年。近年科学家曾从格陵兰岛冻土中成功提取200万年前的DNA序列,其中信息仍历历可辨。

既然DNA携带的信息可跨越数百万年之久,那么今天我们这个时代,能否借用这条生命史的莫比乌斯带保存人间讯息,使之成为接续人类文明的一条纽带?

大胆假设,自然要小心求证。我们先来看看,DNA为什么这么能“装”?

关键在于两方面。其一就是双螺旋结构。“这样的结构使DNA分子在甚小空间内也可以紧凑排列,承载极高密度的信息。”北京大学软件研究所副研究员张成解释道。

此外,DNA四种碱基(腺嘌呤A、胞嘧啶C、鸟嘌呤G和胸腺嘧啶T)的排列组合变化无穷,也就意味着它们可以编码几乎无限量信息,使得DNA理论上具备比任何现有存储设备更高的存储容量。

DNA究竟有多能“装”?理论上,仅1克DNA就能够存储约1000万小时的高清视频数据,想想看,1千克DNA又能装下多少?

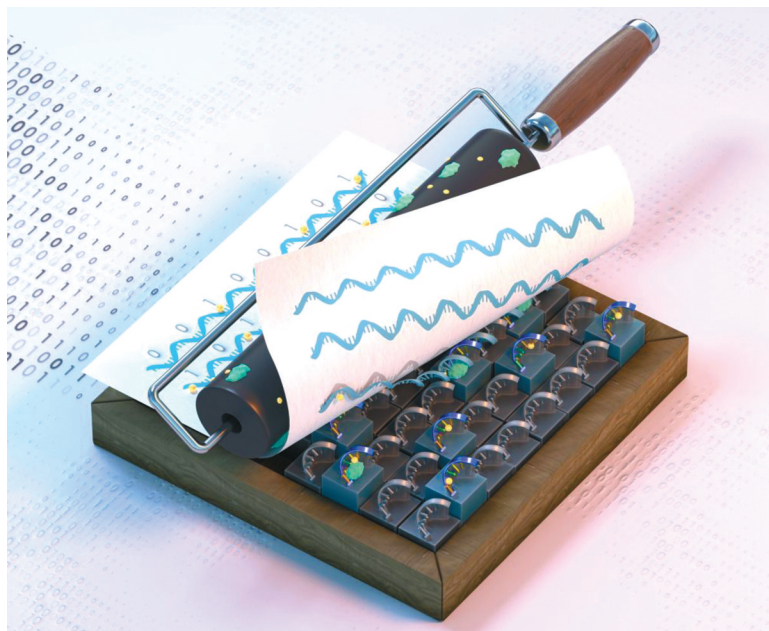
2023年,我国产生的数据总量达32.85泽字节(ZB),相当于1000多万座中国国家图书馆的数字资源总量。折合成我们更熟悉的单位GB,我国每天产生的数据量达到惊人的900亿GB。据国际机构估计,刚过去的2024年,全球生成数据有159.2ZB之多。如果仅凭数据中心来存储,我们要建多少座数据中心才能储存这些数据!

寻找体积更小、容量更大的存储介质,已经是迫在眉睫之事。DNA,能够接过接力棒吗?

以信息时代的后见之明回望人类演化,我们的文明史,其实也是一部数据史。

在全球80亿人平均每人每天产生高达1.5GB数据的今天,“数据怎么存、存在哪里好”的问题,日益凸显。要一块多大的“硬盘”,才能装下我们不断延长的文明史?答案可能在生命最基础的单元之中。对,就是DNA。DNA做的“硬盘”,你想拥有吗?

你的DNA 其实能当“硬盘”用



并行分子印刷DNA存储示意图



北京大学DNA存储研究团队在做试验。

把数据刻进DNA

把海量数据存入DNA,也就三步。

第一步,将数据写入DNA,也就是将今日信息世界通行的二进制数据转化为DNA序列,0和1不同的排列组合实现的多样编码,与DNA序列四种碱基遵循一定之规的排列,二者之间要想转化无碍,就要制定严谨的编码规则,比如,每2个二进制位对应1种核苷酸组合。

接下来,构建存储信息的DNA分子。也就是按照前一步获得的编码布置碱基,合成DNA链。只不过,目前常用的化学合

成法效率还嫌低了些,科学家正在积极探求新方法,利用DNA聚合酶催化的酶促合成法,就为许多科学家所看好。还有科学家提出并行写入DNA信息。近期,北京大学团队利用“分子活字印刷”方法,实现了27.5万比特的平行DNA存储。

最后,就是存储与读取了。DNA存储对环境的要求,不外低温、干燥、避光,低温(一般为零下20摄氏度甚至更低)的干燥环境能减缓DNA分子的降解速度,避光则是为了防止光照引发DNA分子损伤,确保存储数据不会“变质”。

给DNA“硬盘” 一点耐心

前面提到,DNA存储不仅存得久,而且存得牢,若能有效避免水分与紫外线的影, DNA可以在数十年内保持稳定不变。

更妙的是,DNA存储还是一种“低碳之选”。DNA存储额外耗能极低,理想保存环境与日常环境相差不大,要想在能源有限条件下实现数据的长期保存,DNA存储堪称首选。

但是,在今天谈DNA存储,还是一种憧憬,我们也需要明白,有几样问题若不解决,DNA存储还是不能成为我们的下一代主流存储。

——成本高昂。目前合成2MB的DNA数据需要7000美元,读取数据需要再加2000美元,如果以DNA形式存储1GB大小的电影,编码大约需要花费358万美元。

——读写速度太慢。以目前的技术合成DNA,每添加一个碱基大约需要0.1秒钟。别觉得这时间不长,以此速度,保存一首高音质的MP3歌曲都需要差不多1个月。

——读取准确率还不够高。写入过程可能出现不正确的情况,读取时也可能因测序误差导致还原数据不准确。分子数据不比完全虚拟的0与1,自身难免有瑕疵,若不找到合适的处理办法,DNA存储的精度会大打折扣。

革命性的技术,往往起于看似幼稚的尝试,人类文明史无数转折点都佐证了这一点。如果整部人类文明史承传赓续的方式,都可以来一个大进步,那么我们又有什么理由不对今天的探索多些耐心呢?在传递地球生命的莫比乌斯带上,刻下万物灵长探索与创造的证明,会是智慧生命演化史上激动人心的时刻。

据半月谈
图据新华社客户端