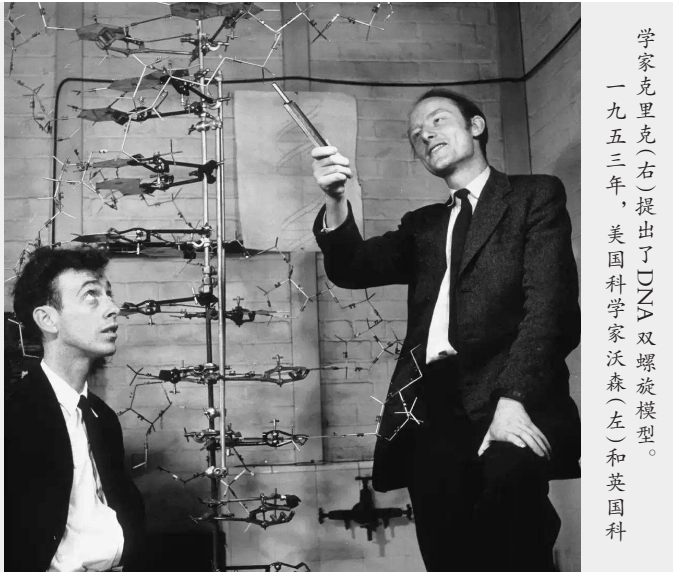




1953年2月28日,第一个DNA双螺旋结构的分子模型诞生,引发生物学研究的重大革命,开启了分子生物学时代。

今年是DNA双螺旋结构发现70周年。过去70年间,人类逐步了解了书写“生命天书”的基本“字母”和“书写”规则,更在不断深入“阅读”其内容的基础上,开启了全新的“创作”之路。DNA双螺旋结构的发现,指引着人类从认识基因走向合成生命,在造福人类自身的同时,也将给未来世界带来更多可能。



科学家沃森(左)和英国科
学家克里克(右)提出了DNA双螺旋模型。
一九五三年,美国科学家沃森(左)和英国科

DNA双螺旋发现70年:从认识基因走向合成生命

■吴家睿

“龙生龙,凤生凤,老鼠生儿会打洞”。生物体与非生物体之根本差异在于,生物体能够保存和利用信息,并把自身的信息一代又一代地传递下去。从古至今,人们一直试图阐明生物体这种承载和传递信息的能力。

这个梦想在1953年终于得到了实现——美国科学家沃森和英国科学家克

里克提出了DNA双螺旋模型,借此在分子水平上清晰地解释了生物体是如何保存和传递信息的。在此基础上,分子生物学诞生,现代生命科学的“帷幕”也就此正式拉开。

过去70年间,人类认识到生命最根本的基石是四个“碱基字母”及其相应的“书写规则”,并由此打开“生命之书”的阅读之门,进入了比世界上任何一个图书馆藏书都要丰富的“生命之书”殿堂。

“生命之书”之阅读篇

研究者从DNA双螺旋模型中认识到,生命的底层逻辑是“文字”,即用“A”“T”“G”“C”4种碱基作为基本单元连接成长长的多核苷酸链,这样的两条多核苷酸链相互缠绕而形成DNA双螺旋。

这4种碱基正是记录生命遗传信息的“字母”,其中每3个碱基组成一个类似单词的遗传密码子,对应一个特定的氨基酸。蛋白质是执行生命活动的主要功能元件,由多个氨基酸作为基本单元连接而成。从4种碱基中选择3种来构成一个密码子,总数为64,所以生物体拥有64个密码子“单词”。

由于生物体用来合成蛋白质的天然氨基酸只有20种,因此除了两种氨基酸(甲硫氨酸和色氨酸)分别由一个密码子决定以外,其他18种氨基酸分别对应2到6个密码子。生物体内的基因正是一段由多种密码子连接起来的“碱基组成和排序。用来指导一种蛋白质的氨基酸组成和排序。

由此可见,大自然在创造生命时采用了“写字”的方式,只需4个碱基作为基本的“字母”,就可以创作出无数的“作品”。因此,研究者的主要工作通常就是阅读“生命之书”中的一段段碱基序列组成的“句子”,并揭示这些句子的含义和可能的生物学功能。

认识编码基因

从“一字一义”到“复杂句式”

研究者最初认为,一个编码基因用于指导一种蛋白质的合成。但随着“阅读”工作的深入,研究者发现,虽然这种“一基因一蛋白”的观点在大肠杆菌等简单生物体上基本成立,但在动植物等复杂生物体上,情况就明显不一样了——这类生物体拥有大量的“断裂基因”,即在一个编码蛋白质的碱基序列之间插有若干段不编码蛋白的碱基序列,其基因内编码蛋白质的碱基序列称为“外显子”,不编码的称为“内含子”。换句话说,“断裂基因”好比一段完整的“语句”被拆开,在中间加入了一些无意义的文字。

研究者还注意到,生命形式越高等,断裂基因就越多。比如,在用于制造面包和啤酒的酵母细胞里,只有4%的基因拥有内含子;而在小鼠或人的基因组内,绝大部分基因都拥有内含子。

为什么大自然在创作“生命之书”时会发展出这样的策略?这看上去与人类的文字创作历史有着某种相似之处。在文学发展早期,创作者以描述为主,每句话都比较明确。但随着文学的发展,创作者希望在一个句子里表达出不同的内容和感觉,从而发展出了复杂句式。比如,爱尔兰作家乔伊斯在其著名的“意识流”小说《尤利西斯》的许多句子中,就掺杂了没有意义的文字。

对拥有这种策略的复杂生物体而言,一个断裂基因内多个外显子通常会受到不同的剪切和拼接,即可变剪

接”——如果把不同的外显子用可变剪接方式进行连接,一个基因就能制造出多种蛋白质。目前已经知道,人类拥有的基因总数大约是两万多个,而能够制造出的蛋白质种类显然要远远大于这个数目。

发现表观遗传

生命“涂抹”的色彩也可传承

“生命之书”拥有众多编码各种蛋白质的基因,生物体必须精准地控制这些基因,以便在需要时制造出特定的蛋白质。为此,生命在基因到蛋白质合成的过程中增加了一个中间步骤,称为“转录”——把基因的碱基序列之信息转录到一条称为“信使RNA”(mRNA)的碱基序列之上,然后用mRNA作为模板指导蛋白质合成。也就是说,生命通过转录方式选择性地“阅读”基因。

近年来的研究发现,DNA上的许多碱基序列通常被化学基团修饰,其中最常见的是甲基化修饰。而DNA甲基化修饰的主要功能正是促进或抑制某个基因的转录活动。由于DNA碱基序列上的化学修饰通常受到机体内外环境的调节,这就成了生命响应环境变化来控制基因转录的重要手段。

最重要的是,这些控制基因转录活动的化学修饰往往可以通过细胞分裂的过程,传递给子代细胞。显然,在这一过程中,DNA碱基序列本身并没有发生改变。人们把这种在DNA碱基序列上的化学修饰信息传递现象称为表观遗传,并由此产生了一门新学科:表观遗传学。

如果把DNA碱基序列的各种化学修饰比喻为不同的颜色,那么“生命之书”就不再是最初人们认识到的单色印刷本,而是一本五颜六色的彩色图书——不仅用碱基序列写出的“文字”可以被复印和传递,用化学基团涂抹在这本“书”里的各种“颜色”也可以被复印和传递。

经典遗传学观点认为,生命只采用多核苷酸链上的碱基序列保存和传递所有可以遗传的信息,且只有DNA碱基序列上发生的改变才能够遗传给下一代。而环境等外部因素导致的个体性状之改变,不论好坏,都不会遗传给后代。

比如,个体因膳食不平衡导致的肥胖是不会遗传给后代的。也就是说,个体在身体内建了一道无形的“墙”,机体在环境作用下后天所获得的性状信息,无法隔“墙”传递到其生殖细胞的DNA序列上,因此后代是不会受到亲代生存环境之影响。而表观遗传现象的发现不仅打破了碱基序列是生命遗传活动唯一载体的教条,而且一系列研究证明,表观遗传信息在某些情况下可以从亲代遗传给子代。

由此可见,表观遗传活动把生命的开放性提升到了一个新的高度,使得从外部环境获得的信息通过表观遗传修饰与机体内的DNA碱基序列上的信息进行整合,不仅能够影响个体当下的生理和病理活动,还能够传递给下一代。



人类认识自然的一个主要目的是要改造自然,为人类服务。同样,人类也不会停留在对生命的“阅读”阶段,而是努力地发展各种新技术去改造生命。为了改写“生命之书”,研究者发展出了两种基本工具:识别“字母”——测定DNA上4种碱基的排列顺序;编辑“文字”——剪切和连接DNA上的碱基序列。

要想识别“生命之书”里的碱基“文字”并对相关内容进行改写,首先需要的就是DNA测序技术。早在1977年就诞生了第一代DNA测序技术。不过,当今生命科学研究的主流是高通量的第二代测序技术。它不仅大大降低了测序成本,还显著提高了测序速度,用一代测序技术完成一个人类基因组30亿个碱基的测序需要三年左右的时间,而使用二代测序技术则可能在一周内完成。不久前,研究者又开发出能够检测单个多核苷酸分子的第三代测序技术。

人们往往用“剪刀加浆糊”形容文字编辑工作。要对“生命之书”里的碱基序列进行编辑,同样需要“剪刀加浆糊”。研究者开发出来的“基因剪刀”是各式各样的核酸酶,其中最常用的是

“生命之书”之编写篇

“限制性内切酶”。这类酶能够识别DNA上特定的碱基序列,从而找到准确的剪切位点,并实现DNA链内的定点切割。

2012年,美国科学家通过改造细菌的核酸酶系统发展出一种全新的“CRISPR-Cas9”技术,它被誉为“基因魔剪”,是目前进行基因编辑最强大的工具,没有之一。碱基序列被剪切之后,需要用“浆糊”重新连接,为此,研究者有针对性地发展出了若干DNA连接酶,以便把DNA链间缺口连接起来。

基因工程诞生

生命“移花接木”带来人类福祉

上世纪70年代,随着DNA限制性内切酶的发现,研究者开始了“生命之书”的编写工作,并把这类在分子水平上对DNA碱基序列进行的操作的技术称为“基因工程”——通常是将外源基因

转入到受体细胞,从而使其特性发生改变或产生新的性状。在这个过程中,首先是利用限制性内切酶把一种生物体(供体)DNA上的特定基因切下来,将其与运载工具(如质粒或病毒)上的DNA在体外人工连接,形成新的重组DNA,然后转送到另一种生物体(受体)中进行扩增和表达。

基因工程诞生以来,为人类的福祉做出了巨大贡献。生产治疗糖尿病的胰岛素就是个典型范例。

2022年是胰岛素用于临床治疗的第100周年。第一代医用胰岛素主要是源于牛或猪的胰腺提取物。这种生产方式不仅昂贵,而且产量很低,远不能满足临床需求。1981年,美国研究者利用基因工程技术,将人的胰岛素基因导入大肠杆菌,通过大肠杆菌大量生产重组人胰岛素。从此第一代胰岛素产品完全被这个第二代产品所取代。为了进一步提高胰岛素的疗效和安全性,研究者又对胰岛素基因进行精细改造,获得了第三代产品——重组胰岛素类似物。

基因工程在农作物领域同样发挥了巨大作用。1983年,科学家培育出了第一个转基因作物——转基因烟草;到2002年,世界上大约有550万到600万公顷种植转基因烟草。1996年,美国的农场主开始种植转基因大豆——在这种大豆里转入了植物“矮牵牛”的一种抗性基因,从而可以抵抗除草剂。2021年,美国转基因大豆种植面积占美国大豆总种植面积的95%,总产量为1.2亿吨。

基因编辑迭代

为治愈遗传病点亮希望之光

基于DNA限制性内切酶的基因工程技术极大改善了人类的生产生活,但该技术应用在基因编辑的应用中也表现出一些明显的不足,比如实验流程比较复杂,结果获取需时较长,编写能力不够



人工合成生命始终是研究者在探索生命奥秘过程中一个不灭的梦想。

早在1828年,德国化学家维勒就在实验室中利用氰酸铵合成了尿素,首次证明人类可以在实验室里利用简单的无机分子合成源自生物体内的有机化合物。上世纪60年代,中国科学家在世界上首次人工合成了第一个具有生物活性的蛋白质——由51个氨基酸组成的牛胰岛素;1981年又在世界上首次人工合成了具有生物活性的多核苷酸链——由76个核苷酸组成的酵母丙氨酸转移核糖核酸(酵母tRNA^{Ala})。

随着科学技术的发展,近年来研究者在实验室合成生命的能力也在显著提高,他们甚至试图创作出自然界中不存在的“生命之书”。

试管里合成生命

人类基因组计划将出“编写版”

2010年5月,美国生物学家文特尔发布了世界上首个“人造生命”——科

“生命之书”之创作篇

式实施,必将对人类认识和控制自身带来更为深远的影响和意义。

设计和改造生命

按照人的意愿进行全新创作

研究者的脚步并没有停留在按照自然界已有的“生命之书”进行复制,而是开始按照人的意愿来创作全新的“生命之书”。

例如,美国生物学家文特尔在其化学全合成的首个“人造生命”上进行了设计和改造,把他认为冗余的或非必需的基因“句子”从该基因组上删除,化学合成了一个比原基因组碱基序列少了一半的“微型细菌基因组”。而含有这个微型基因组的细胞,其自我繁殖的速度比具有全基因组序列的细胞快了近五倍。

自然状态下的“生命之书”通常拥

精准等。

为此,研究者一直在努力开发更好的基因编辑技术。2012年,美国科学家道德娜和萨尔庞捷在细菌的基因编辑系统基础上发展出了一种理想的基因编辑技术“CRISPR-Cas9”。这种技术的基本原理是:利用一段设计好的RNA序列“sgRNA”引导DNA内切酶“Cas9”至特定的DNA序列上进行剪切。该技术经过不断迭代,目前不仅可以在细胞的基因组内插入长达3.6万个碱基的外源DNA片段,而且可以对细胞内的DNA乃至RNA进行单个碱基的修改。

这类新型基因编辑技术的出现,显著提升了人类抗击疾病的能力,尤其为治疗基因碱基序列异常的遗传性疾病提供了有力武器。据统计,目前已知的单基因遗传病超过9000种,对人类的健康造成了巨大危害。

例如,血红蛋白基因异常会导致“地中海贫血”。目前,全球有近3.5亿地中海贫血基因的携带者,我国长江以南各省是该病的高发区,在部分高发地区,这种基因的携带者在人群中超过10%。过去,由于没有好的治疗药物或方法,重度地中海贫血患者只能定期进行输血;现在则有望利用CRISPR技术来治疗地中海贫血。目前,全球进入临床研究阶段的地中海贫血基因编辑治疗产品有6个,其中5个是采用CRISPR技术。

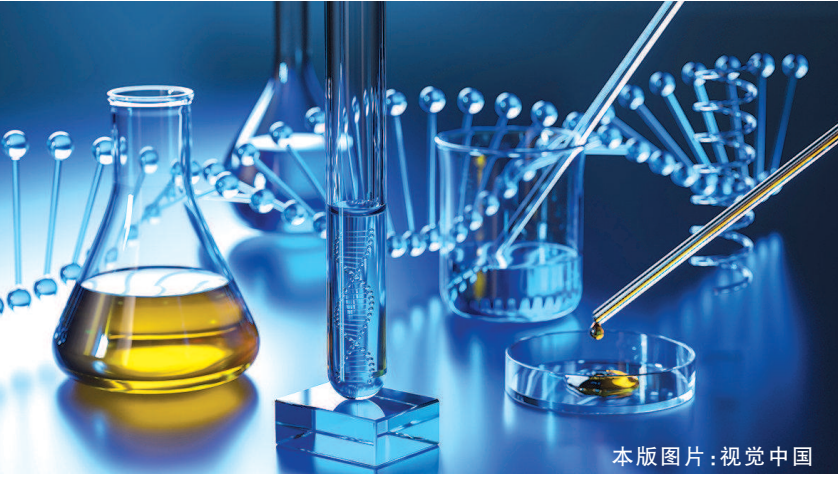
值得注意的是,CRISPR技术在改良农作物品种方面同样有着巨大优势。它不仅可以利用外源基因来改造作物,还能够把作物自身的基因按照设计好的目标直接进行“改写”。2013年,中国科学家高影霞在世界上首次报道了利用CRISPR技术编辑水稻和小麦DNA序列的研究工作,并在2014年报道了使用CRISPR技术修改小麦的基因序列,从而使编辑后的小麦能够抵抗小麦白粉病的侵袭。

可以说,在CRISPR等一系列基因编辑技术的推动下,传统的那种不可控的作物自然育种未来将转变为高度可控的作物设计育种。

有64个“单词”——遗传密码子,用来指导20种天然氨基酸合成蛋白质,其中有一些密码子是冗余的,称为“同义密码子”。2019年,英国科学家发布了一个“人造”大肠杆菌,它含有一个人工设计并化学合成的人造大肠杆菌基因组,其中全新设计并合成的人工基因“句子”大约包含400万个碱基对;在这些人工设计的基因里,只保留了61个遗传密码子,删除了3个同义密码子,这些改造过的基因序列仍然可以制造出正常的蛋白质。

研究者不仅试着修改基因“句子”和密码“单词”,还在碱基“字母”上做起了文章。2014年,美国研究者宣布他们在大肠杆菌的DNA序列里加入了两种非天然碱基——dNaM和dTPT3,这两种碱基也能够 在体内通过复制的方式传递信息。而在2017年的论文里,这些研究者进一步证明,这两个插入到天然基因序列里的非天然碱基,同样能够在蛋白质合成中引导特定的非天然氨基酸。

(作者为中国科学院生物化学与细胞生物研究所研究员)



本版图片:视觉中国