

生物信息学下一个十年在哪里？

■本报记者 韩扬眉

“10年来,数学、计算机与生命科学交叉领域发生了显著变化。如今,随着大模型、人工智能等技术的发展,各个学科面临巨大挑战,到了必须要‘交叉’的时候了。这无论对现代科学的发展,还是服务国家战略需求,都非常重要。”中国科学院院士、中国科学院数学与系统科学研究院研究员马志明为青年科研人员鼓劲打气。

近年来,ChatGPT等深度学习大模型的诞生,给几乎每个学科领域带来了新机遇。基于数据、信息而生的生物信息学无疑将迎来新变化,它的下一个10年在哪里?

近日,第十届“数学、计算机与生命科学交叉研究”青年学者论坛在北京举办。会上,多位院士专家、优秀中青年学者与学生共同交流生物信息学的前沿进展,展望生物信息学未来的发展。

“暗箱”待开

在人类生物医学发展史上,遗传密码的破解使得生物医学进入了大数据时代,生物信息学也驶入了“快车道”。

“这是划时代的变化,生物医学已进入用数据刻画的时代,数据带来了海量的信息。”中国科学院院士陈润生说,上世纪90年代,他参加了人类基因组和水稻基因组的分析研究,那时,生物信息学还比较冷门。

而事实上,生物信息学给生命科学研究带来了一场变革。

陈润生举例说,基于大数据的大规模测序,我们能够更好地抵御未来烈性传染病的侵袭;基于大数据,医生有了准确的靶向判断,严重的肿瘤疾病也可防可治。“大数据带来大信息牵引的精准医学,已经成为新一轮国家科技竞争的战略制高点。”

“但大数据也带来了新的挑战,那就是我们并不了解它的生物医学本质。”陈润生指出,以生物组学大数据为例,它是复杂的,也是不同尺度、多维度、异质化的,比如病历是用自然语言写出来的,生理生化指标是波形的,胸片是影像的,而组学数据是字符串形式的。那么,这些不同标记如何协调统一共同分析,是大数据分析的难题之一。另一个挑战是,数据指标是动态和非线性的。

“随着越来越多表征疾病数据的产生,我们

需要理解疾病是怎么发生的、导致疾病的因素是什么。这就好比‘暗箱’,想要打开它,就需要用人工智能对生物医学数据进行深度挖掘。”陈润生说。

大模型的出现,或许是生物信息学新变革的开始。

陈润生表示,大模型的工作方式必然会深入到各个环节,但不需重造,更多的是利用,未来大模型和生物医学是更高层次的结合。“推动大模型向真实的大脑学习,不仅是训练它的学习量,更多的是看清脑的复杂结构组成。”

大科学引领大发现

打开“暗箱”的工作,单靠一个学科或领域并不现实。正如中国科学院院士贺福初所说,生物信息学的未来,更多是需要数学和系统科学等学科的深度参与,这可能将生物信息学带往新的方向。

贺福初呼吁,未来要进一步探索生命科学理论或基本规律。

“大科学时代产生了大数据,能够形成规律性的总结吗?”贺福初表示,只有在数学等其他科学的基础之上,才能从大数据中提炼信息,发现知识和重要原理。

事实上,这在2000多年前就已有印证。古希腊毕达哥拉斯学派相继发现,算术的本质是“绝对的不连续量”,音乐的本质是“相对的不连续量”,几何的本质是“静止的连续量”,天文学的本质是“运动的连续量”。他们因此认为,数学是通向理性世界的必备工具。

后来的科学发展也表明,数学作为整个自然科学的基础,奠定了很多学科“大发现”的基石。

自从生命科学进入大数据时代,关于生命的基因组数据增长速度是摩尔定律的4倍。然而,在贺福初看来,基因组无法解释生命的生老病死。

“时空各异,基因组相同;生老病死,基因组不变。”贺福初表示,诸多生命现象无法从基因组层面得以阐明,必须通过蛋白质组在时间、空间上的动态变化,发掘生命现象的本源与本质。

基于此,由贺福初领衔的“人体蛋白质组导航计划”,使命是解读人体蛋白质组的构成原理与变化规律,实现蛋白质组学驱动的医学范式转变,共

创智慧医学新时代。

“科学大发现时代离不开学科交叉。”贺福初强调,生命科学将借助数学、计算科学与人工智能进入大发现时代。

下一个十年需要更多年轻人

前辈的指导与鼓励给年轻人带来更多信心。

论坛共同主席、中国科学院数学与系统科学研究院研究员张世华告诉《中国科学报》,10年前,他是一名助理研究员,那时,每个领域的学者都不少,但在交叉学科里缺少交流的机会和平台。于是,他与中国科学院遗传与发育生物学研究所研究员王秀杰一起倡导、发起、组织论坛,加强从事数学、计算机与生命科学交叉研究的青年学者之间的联系,交流生命科学与计算生物学研究领域的最新成果。

当年很多初出茅庐的年轻人和学生,如今都已成为数学、计算机与生命科学交叉领域里的中流砥柱。他们希望在生物信息学的下一个10年迎来更多的年轻力量。

北京大学生物医学前沿创新中心研究员张泽民曾供职于高新技术企业,他对产业和学术都有深刻的感受。在他看来,“生物信息学领域技术总在变,但它的科学思维和目标一直没变,所以无论技术如何变化,我们关注的焦点在于如何把数据信息变成知识,从数据和新技术中挖掘更多机会”。

在人工智能大模型时代,生物数据几乎6个月翻一番,生物信息学会有哪些变化?

同济大学生命科学与技术学院教授张勇表示,生物信息学可能更多从实验学科向理论学科转变,从理科向工科转变。

“如何把大模型应用于生命科学领域,除了要了解数据外,更重要的是要提出一个好问题,这比关注模型本身更重要。”北京大学生命科学学院研究员高歌说。

陈润生指出,伴随大数据在生物、医学中的深度应用,以及物质资源向数据资源的不断迁移,未来10年对生物信息的需求会越来越迫切和广泛。为此,加强生物信息学的学科建设、鼓励源头创新、培养人才队伍非常重要。同时,也期待生物信息学领域研究人员在科技创新和国家经济社会发展中作出更多更大的贡献。

发现·进展

中国热带农业科学院橡胶研究所

实现橡胶树树干轮廓信息自动化提取

本报讯(记者张晴丹)近日,中国热带农业科学院橡胶研究所机械装备与自动化团队在橡胶树树干轮廓信息提取方面获得新进展。该团队成功提取了橡胶树树干轮廓曲线及其表面特征信息,不仅能提高割胶装备树干轮廓仿形精度,还为橡胶树材积率估算和其他果树环割装备的研发提供了理论依据和技术支撑。相关研究成果发表于《森林》。

自动化割胶机械的推广与应用是天然橡胶产业发展的必然趋势,能有效减少胶工资源短缺、割胶环境恶劣带来的影响。目前,自动化割胶机械主要采用机械式仿形方法,仿形精度易受树干轮廓形状和树干表面粗糙度影响,如何实现树干轮廓高精度仿形已成为自动化割胶领域的研究重点,而获取精准的树干轮廓信息是提高仿形精度的关键之一。

研究团队基于激光测距技术,自主研发了橡胶树树干轮廓自动化采集系统,实现了树干轮廓点云数据及前水线、后水线、剖面、裂缝、凹坑等表面特征信息的自动化采集,轮廓特征识别率为94.67%。上述结果表明,自动化采集系统和方法具有可行性,且样条曲线适用于复杂轮廓的拟合,可作为橡胶树树干轮廓信息提取的有效方法。

研究团队还成功提取了橡胶树树干轮廓曲线,重复提取误差为0.04%,结果表明轮廓信息提取方向和树干位置不会影响轮廓曲线提取精度,在轮廓提取过程中无须进行位置调节,验证了树干轮廓信息提取方法的准确性和稳定性,可保证橡胶树树干轮廓信息提取的精度和效率。

相关论文信息:

<https://doi.org/10.3390/fl4061122>

中国科学院海洋研究所

研究揭示生命起源的关键

本报讯(记者廖洋 通讯员王敏)近日,中国科学院海洋研究所研究员孙卫东课题组在生命起源领域取得重要进展。课题组利用高温高压实验证明了氮气可以快速参与蛇纹石化过程并生成大量氨气,说明了地球早期在岩浆海后期,蛇纹石化导致地球大气由“二氧化碳+氮气”转变为“氨气+甲烷”,在闪电作用下合成大量氨基酸,在超临界水+二氧化碳层形成氨基酸浓汤,这是生命起源的关键。相关研究成果发表于《科学通报》。

传统认为冥古代地球大气主要成分是二氧化碳和氮气,缺乏甲烷和氨气,在中性大气里氨基酸合成效率将大大降低,氨基酸能否在原始大气中大量合成存在争议。针对冥古代地表氨基酸合成缺乏关键原料氨这一问题,孙卫东课题组进行了蛇纹石

化合成氨的高温高压水热实验,研究“橄榄岩-水-氮气(=二氧化碳)”体系在冥古代地表温压条件下反应,发现大量证明氮气参与蛇纹石化合成氨过程,进而为生命起源过程中氨的来源提供启示。

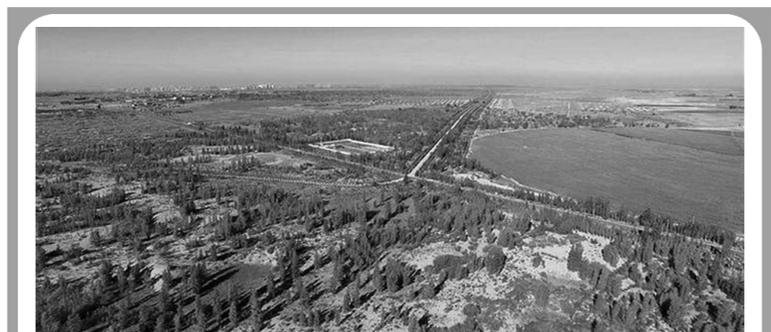
研究证明,在冥古代地表可以广泛发生蛇纹石化合成氨过程,保守估算,该过程每年可以为地表系统提供超过10¹⁵克数量的氨。基于上述结果,研究人员提出,在冥古代地表广泛的蛇纹石化作用下,氨气、甲烷和氨大量生成,在闪电作用下形成氨基酸浓汤,为生命起源提供了适宜环境,这对于理解早期大气演化和前生物合成反应具有重要意义。

相关论文信息:

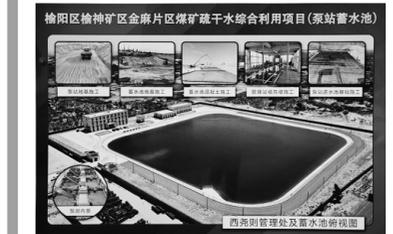
<https://doi.org/10.1016/j.scib.2023.04.038>

榆阳区生态产品特点与价值的实现途径

■王飞 汪月琴 穆兴民



榆阳区沙漠造林20年的林地。



榆阳区神矿区金麻片煤矿疏干水综合利用项目(泵站蓄水池)。



榆阳区老城区文化街区。

作者供图

大部分地段煤层累计可采厚度在15米以上,开采后地下释放很大空间。采空区有极大的地质风险、水文和生态风险,且随机性强,隐患长,潜在损失很大。

四是碳汇潜力及其市场化优势问题。榆阳区林地面积大,全区林木保存面积482.2万亩,但大多数植被高度和冠幅有限,10~20年林地的覆盖率仅0.4~0.5,成熟林的生物量常在10~30吨/公顷,生物量积累不高,而且达到稳定时间短,故在碳交易市场缺少优势。但该区可造林面积大,林草植被生态效益显著,市场潜力很大。

五是利用矿井疏干水带来的可持续问题。榆阳区水资源短缺,水资源总量约为9.135亿立方米,目前取用地下水(约53%)比例很大;每年2.1亿立方米疏干水和1.77亿立方米的可外供用水量对榆阳区工农业生产、生态治理和水系补水非常重要。但疏干引发的资源、生态和环境效应复杂,已经导致地下水水位明显下降,区域浅层地下水环境开始恶化;存在浅层水丧失问题,周边县市地下水和新建灌溉系统、果园和

大棚等可持续性面临挑战。

六是丰富而零散的生态产品需要整合的问题。榆阳区旅游资源非常丰富,主要包括地文景观、水域风光、生物景观、遗址遗迹、建筑设施、旅游商品和人文活动等。目前全域旅游规划布局完善,设计宏大细致,具有远见和魄力,但对促进生态产品价值实现的具体设计还不够,对生态、农业、自然教育缺少挖掘,需要科学整合各类资源,提升旅游搭配多样性,继而提高吸引力。

七是煤矿开发后的“地下空房子”问题。榆阳区境内煤矿开发后留下大量空间,这不但会持续引起地下沉降,还会给区域水资源、水循环和生态系统稳定性带来危害。这是实现榆阳区生态安全和产品价值必须考量的重要因素,需要正视觉发展过程带来的挑战,化危为机,超越创新,通过思考谁来做、怎么做、花多少、值多少等一系列科学问题,给出创新性的解答。

八是生态产品价值和生态空间价值厘清的问题。生态空间是指具有自然属性,以提供生态服务或生态产品为主体功能的国土空间;生态

产品是生态空间各要素的综合结果,碳汇和生态产品的价值实现需要与生态空间保护和维持相对应;生态空间通常以保护为责任,缺少价值估算的相关政策,榆阳区应创新性地解答这一问题。

九是自身发展和域外影响同步提高的问题。榆阳区的经济和社会发展迅速,但由于榆阳区矿产资源开发的服务范围很大,生态产品价值的实现需要在当地落实,因此面临如何增加当地和周边地区在全国范围内的持续吸引力,以便更好地支持未来可持续发展的问题。

促进碳汇和生态产品价值实现的三个机遇

首先是利用好国家生态保护的政策红利。中共中央、国务院出台了一系列关于碳达峰、碳中和的文件和政策,生态环境、农林、综合经济领域也都相继出台了相应的落实政策或建议,力促碳汇产品的形成,并通过与企业开展碳交易来实现碳汇功能。榆阳区可依据全国主体功能区规划、建立健全生态保护补偿机制和生态产品价值实现机制等相关系列文件,从制度上进行创新探索,促进当地生态产品价值的实现。

其次是发挥良好的经济和社会管理能力。目前榆阳区经济相对发达,2022年全区生产总值1673.5亿元,可以鼓励和动员经济效益好的企业通过碳汇交易实现“双碳”目标。同时榆阳区曾先后在全国实施了集体产权制度改革,率先在全国开展林业碳汇试点工作,并用于社会治理实践,规划布局并开启“大美榆阳”全域旅游,具有促进碳汇和生态产品价值实现的创新能力和管理智慧。

最后是依靠科技创新支撑发展。在实现碳汇和生态产品价值的实践中,榆阳区可继续借助生态保护与修复理论和技术,实现煤矿采空区综合治理与矿井疏干水综合利用,聚焦环保监测、智慧矿山、现代农业、文化旅游及“智慧榆阳”建设等方面发挥科技创新引领支撑作用,同时在数字榆阳建设中,将碳汇和生态产品有机嵌入全社会管理的环节。

促进碳汇和生态产品价值实现的三个建议

一是把碳汇作为生态产品和生态服务的“载体”来保护,通过碳汇交易促进全域生态保护和生态服务功能。针对榆阳区碳汇对面积依赖大、受水资源限制大和单位固碳能力相对较低的实际情况,建议在考虑榆阳区林地固碳能力、发展碳汇交易的同时,结合生态恢复在减少风力侵蚀危害、降低风速减少地表蒸发、改善区域气候环境方面的作用,强化生态服务功能,并

尽快核算生态服务价值,尤其是对当地居民生命健康造成的影响,从而促进生态产品价值长期实现。

在碳汇和生态产品价值交易方面,建议从三个方面强化:第一,考虑矿产开发对其区域水资源影响和生态破坏程度,建议用矿产开发生态补偿费用,支持人工造林和自然恢复以及对植被的保护和管理,保持碳汇功能稳定和持续提升;第二,基于榆林区蒸发量大、水资源短缺的事实,通过“生态林+光伏”的形式,减少林地蒸发量,并通过面板集流促进降雨渗入深层,提高水资源利用效率,生态产品的相关责任由清洁发电企业承担;第三,探索和开发新型林产经济和林下经济模式,如林下适度放养等,逐步改善土壤养分,提高林地蓄水特性与抗旱能力,生态产品和碳汇由受益者承担。

二是创造性用好矿产采后宝贵空间,建立战略性地下水库。由于榆阳区气候干旱、降雨变化大,现有的水库大气蒸发强烈,而且淤积严重,新修水库面临土地占用、优良坝址缺乏等困难,因此建议在地下矿产开发过程中,主动考虑利用未来地下空间资源和生态功能进行规划,通过隔墙、通道和采空区顶面加固等措施,在地下形成相对独立、逐步建设的水库,既可以为未来开采的矿井水的储存提供空间,减少疏干水到达地表后的蒸发损失以及土壤与岩石失水形变带来的风险,同时又可形成数十亿立方米的蓄水空间,为未来榆阳区工业发展、城市用水和生态用水提供应急保障。

利用榆阳区的煤矿具有建设地下水库的较好建库条件和可利用邻近的黄河水来补充地下水的条件,未来可以形成一个100亿~120亿立方米左右的超大型地下水库,不但可以减少地面沉降及地面水库占地与险患,还可以防范未来可能出现的极端干旱事件,为全区甚至整个榆林市提供水资源保障。

三是本着“尊重历史、抓住机遇、当代创造、永惠子孙”的态度,抢占先机建设“长城今朝生态文明建设研学基地”,促进生态产品价值长期实现。加强对榆阳区内长城沿线的综合保护并开展研学。榆阳区和榆林市长城分布广、类型多、重要性强,有厚重的史实作为基础;榆林市有生态脆弱区、农牧交错区、文化交融区、革命老区和能源基地,生态、经济和文化类型多样,研学内容丰富;长城保护和生态文明建设有密不可分的关系,统筹长城和长城文化保护、乡村振兴、生态保护等方面工作一体化发展,打造特有的优势品牌,以此将榆阳区打造成陕西榆林市自然、生态和文化标杆,书写陕西高质量发展新篇章的亮丽篇章。

(作者系西北农林科技大学研究员)