

中国农业科学院:人工智能让育种“物美价廉”

自从作物被驯化以来,培育集抗性强、优质、高产等性状为一体的作物品种一直是育种家的梦想。DNA分子结构模型的发现推动了分子生物学的发展,让育种家们能够从基因和分子水平上解码作物的生命秘密,通过调控基因获得特定表型,以期培育出最想要的作物品种。

然而,如何调控作物基因才能培育优良品种?如何不用大规模田间试验就能预测基因变异后的作物生长状况?时至今日,这些问题依然困扰着育种学家们。

近日,中国农业科学院生物技术研究所副研究员汪海与合作者共同开发出从基因组DNA序列预测基因表达调控模式的人工神经网络模型,有望借助人工智能(AI)技术实现定向育种。相关成果发表在《美国科学院院刊》上。

从经验到精准定向

育种,从某种意义上来说,是把来自不同种质资源的优良等位基因聚合起来。

作物育种经历了漫长的改良之路。传统育种是耕作者对作物表型变异的肉眼观察,通过主观判断选出高产优质抗性强的育种材料。后来,职业育种家出现,他们根据对作物遗传规律的认识,通过预先设计杂交育种试验,再从后代中筛选出优良栽培品种。

这些方法曾为作物改良、有效解决粮食安全问题作出了巨大的贡献。但在某种程度上,却都是基于经验和观察,完全根据表型对育种材料进行选育的“经验育种”。科学家曾“无奈”而又

形象地将其形容为“一把尺子一杆秤,用牙咬,用眼瞪”。

“作物表型易受环境、气候等因素影响,依赖于经验育种效率低,且成本高、田间管理难度大。过去几十年甚至上百年来,基本是沿用这种方式,并无大的突破。”华南农业大学生命科学学院教授王海洋告诉《中国科学报》。

直到20世纪50年代,分子生物学与基因工程的诞生,打开了人类认识生命本质的大门。作物育种从经验育种时代进入了分子定向育种时代。这个时期,育种家可在明确基因型的表型效应

的情况下,有的放矢地把符合预期要求的基因型进行组合。

“找到控制作物最佳性状的基因,对其进行标记,在后代中监测追踪,从而有目的地对单一目标性状进行基因改良,大大提高了育种效率和精确度。”王海洋说。

然而,伴随着高通量基因组测序技术的发展,越来越多的作物全基因组密码被解开。在海量的基因组数据面前,控制优良性状的基因是哪些?怎样的基因组合才能产出最优的作物品种?上述分子标记有效利用与定向育种的先决条件,人们却不得而知。

汪海表示,明确哪些分子标记和哪些性状相关联,需要借助机器学习模型或深度学习模型帮助育种家根据基因型预测表型。人工智能技术突破了人的经验,使作物育种更加精准而高效。

深度学习模型帮助预测优势品种

机器学习是借助计算机算法建立模型并解析数据,通过不

断学习数据的自身特征并训练模型,从而实现对目标对象的判断和预测。

汪海告诉记者,传统的基于线性模型的机器学习方法由于不考虑生物学过程背后的分子机制,造成模型不会“举一反三”,在某个基因上学习到的特征不能运用到相似分子机制的基因,而且不能有效预测低频、罕见变异的表型效应。以玉米为例,玉米自然群体中就有超过50%的变异属于低频、罕见变异。

以基因组序列为预测变量的深度学习模型可以克服这一难点。

研究人员以基因家族代替单个基因为单位随机分配训练集和测试集数据,以解决“进化依赖”造成的模型“过拟合”问题。接着进一步利用多种算法对模型进行解析,获得了调控基因表达的关键DNA基序。在此模型基础上,研究人员利用进化上亲缘关系较近的两个物种,成功预测了同源基因的相对表达量,并进一步获得了调控同源基因相对表达量的关键DNA基序。

汪海表示,深度学习模型通过模拟分子生物学过程,可在自然群体中预测直接造成表型的因果变异,而非和因果变异紧密连锁的变异。未来可以针对因果变异进行基因组编辑,直接将有利自然变异引入现有的育种材料。

此外,与传统高投入、大规模的田间试验相比,人工智能神经网络模型可在计算机中对基因组DNA序列进行虚拟诱变,并利用模型预测变异的后果。“从而再挑选符合预期目标的变异序

列进行实验验证,实现低成本定点定向设计育种。”汪海说。

智能化育种4.0时代

“这是作物优良基因挖掘方法的突破,也代表了未来的发展方向。”中国农业大学农学与生物技术学院植物遗传育种学系教授、国家玉米改良中心主任李建生如是说。

以人工神经网络为代表的新一代人工智能技术具有更强大的数据挖掘能力,正推动作物育种走向智能化的“4.0”时代。

中国农业大学作物基因组与生物信息学系教授王向峰撰文以玉米为例,对育种“4.0时代”进行了详细的阐释:依托人工智能、基因组测序、基因编辑等相关技术,实现玉米基因组基因型与表型大数据的快速积累,通过遗传变异等数据的整合,实现作物性状调控基因的快速挖掘与表型的精准预测,通过人工改造基因元器件与人工合成基因回路,使作物具备新的抗逆、高效等生物学性状,并通过在全基因组层面上建立机器学习预测模型,创建智能组合优良等位基因的自然变异、人工变异、数量性状位点的育种设计方案,最终实现智能、高效、定向培育新品种。

在人工智能技术辅助育种方面,美国农业公司已有应用。比如原孟山都公司,通过人工智能筛选,只需对最具开发潜力的品种分子进行田间测试,即可帮助农民增收。此外,借助机器学习和预测建模技术,快速为农民提供数字化解

决方案。

“中国要实现应用还有一段路程要走。”李建生表示,与国外农业公司种业集中度高、规模大相比,中国种业公司多为“作坊式”生产且分布分散,要实现高通量的基因筛选与预测,需要改良适合中国种业发展的模型和方法。

在研究方面,汪海坦承,目前,把深度学习等人工智能技术应用于基因组学领域在国内外都刚刚起步。

在他看来,阻碍人工智能技术在基因组学中广泛应用的因素之一是跨领域人才缺乏。“基因组学领域的人需要学习和掌握人工智能技术方法,并根据基因组学领域问题的特殊性,对人工智能技术进行改造。”

除此之外,训练深度学习模型需要大量的数据。然而在农业领域,作物的基因型和表型数据量却积累不足。

王海洋建议,研究人员在育种后,除了留下优质品种数据,也要保存非理想型品种的全套基因组和表型数据,以便数据建模时进行优劣比较,找出调控优良表型性状的基因。

大数据时代下智能化育种的前提是标准化大数据体系。而农业数据采之不易且不统一,王海洋表示,作物表型数据差异性较大,不同人采集的数据真实可靠性与准确性也难以控制。除此之外,彼此数据不开放共享,使得研究中可比较的数据量少。“有数据是第一步。对数据进行规范化采集处理、存储与管理,并建立开放共享的数据库更重要。”

(中国科学报)

中美科学家揭开甘蓝型油菜“家族血统”之谜

本报讯 柯高阳 甘蓝型油菜是油菜的常见类型,也是人类食用植物油的重要来源。中美科研人员历时4年研究,通过全基因组重测序揭开甘蓝型油菜“家族血统”之谜,有助于通过基因选育提升甘蓝型油菜的产量和品质。相关研究成果已于近日由国际学术期刊《自然·通讯》在线发表。

甘蓝型油菜由白菜与甘蓝通过自然杂交进化而来,因产量高、抗病性强在全球得到广泛种植。但甘蓝型油菜起源于何时、“家族血统”如何,科学界仍在研究。2015年,来自西南大学农学与生物科技学院、中国农科院蔬菜研究所和美国佐治亚大学的3支科研团队合作,共同开展甘蓝型油菜群体遗传学研究。科学家对从各地选取的588份有代表性的甘蓝型油菜材料进行全基因组重测序,同时结合中国农

科院蔬菜研究所的199份白菜和119份甘蓝重测序数据,通过群体基因组分析追根溯源。

分析研究发现,约7000年前,甘蓝型油菜由地中海地区白菜品种里的欧洲芜菁,和苕蓝、花菜、西兰花、中国芥蓝等4种甘蓝的共同“祖先”杂交合成。经过多代繁衍,这些“祖先”先后杂交出适应不同生长条件、具有不同特点的多个甘蓝型油菜品种。

“掌握甘蓝型油菜的‘家族血统’情况,有助于人们通过基因选育方式,有目的地选择有益基因类型、优化脂肪酸组成,从而提升油菜的产量和品质。”研究团队成员、西南大学农学与生物科技学院唐章林教授介绍,在这项研究中取得的甘蓝型油菜全基因组重测序结果,还对开展油菜全基因组选择育种和基因工程改良具有重要意义。

温室气体减排有望缓解降水变化对农业的影响

美国《国家科学院学报》11日在线发表的一项新研究显示,全球变暖将造成降水量变化,影响一些常见农作物的产量。采取有力的温室气体减排措施有望缓解降水变化对农业的影响。

国际热带农业中心、英国利兹大学和智利大学等机构的研究人员对从弱到强四种温室气体排放强度造成的永久性降水变化的发生时间进行了预测。结果显示,澳大利亚西南部、非洲南部、南美洲西南部、墨西哥中部和地中海区域将变得更加干燥,而中国、印度、加拿大、俄罗斯和美国东部将变得更加湿润。

研究人员警告说,到2040年,全球多达14%的种植小麦、玉米、水稻和大豆的土地将变得更加干燥,而多达31%的种植这四种作物的土地将变得

更加湿润。另外,尽管部分地区降水量的增加可能意味着产量增加,但考虑到海平面升高、气温升高和洪水风险增加等因素,这些地区的农作物产量未必能增加。

研究发现,如果采取快速的温室气体减排行动,例如实现《巴黎协定》制定的温室气体减排目标,有望推迟气候变化导致的降水变化出现时间或缩小受影响区域,从而降低农业适应成本。

研究人员还表示,虽然温室气体减排努力无法改变全球农作物主产区的降水量发生变化的总趋势,但与较高的温室气体排放强度相比,较低的温室气体排放强度有望为这些地区做出调整以适应气候变化争取20年到30年的时间。

(新华网)