

中国种业科技创新的智能时代——“玉米育种4.0”

王向峰¹, 才 卓²

(1. 中国农业大学/国家玉米改良中心, 北京 100193; 2. 吉林省农业科学院/国家玉米工程技术研究中心(吉林), 长春 130033)

摘 要: 二十一世纪是人类社会全面步入人工智能的时代。人工智能与生命科学的结合必将为医疗、健康、农业等行业带来革命性变革, 同时也为我国玉米种业的发展提供了千载难逢的历史机遇。回顾育种发展历史, 作物育种技术的变迁大致经历了3个阶段, 即古近代耕作者基于表型观察开展的农作物驯化与农家品种选育; 现代职业育种家借助遗传学与统计学知识, 基于经验开展的新品种改良; 分子育种工作者应用基因工程和分子标记技术定向改造作物性状。伴随人类社会步入互联网、大数据、人工智能“三位一体”的时代, 育种学已步入第4个历史时期。美国科学院院士、玉米遗传育种学家Edwards Buckler教授提出“育种4.0”的理念, 未来育种将在基因编辑与合成生物学、基因组与生物信息学、大数据与人工智能等跨学科、多技术支撑下, 实现作物新品种的智能、高效、定向培育, 最终推动育种学从“艺术”到“科学”到“智能”的革命性转变。

关键词: 玉米; 育种4.0理念; 基因组选择; 人工智能

中图分类号: S513.035

文献标识码: A

Era of Maize Breeding 4.0

WANG Xiang-feng¹, CAI Zhuo²

(1. National Maize Improvement Center, Beijing 100193;

2. National Maize Engineering Research Center(Jilin), Changchun 130033, China)

Abstract: In the first twenty years of 21st century, human society has entered the era of artificial intelligences (AI). Combining AI and life sciences has promoted revolutionary changes in medicines, health and agriculture, and also provide unprecedented opportunities for maize seed industry in China. Crop breeding has experienced three historical phases, in which, ancient farmers cultivated landrace varieties by observing crop trait variations, professional breeders applied designed breeding schemes according to genetics and statistics, and modern molecular breeders applied genetic engineering technology to modify desired traits. Professor Edwards Buckler, a renowned maize geneticist raised the concept of “Breeding 4.0” which emphasizes that nowadays crop breeding will be supported by multiple cross-disciplinary technologies including gene editing and synthetic biology, genomics and bioinformatics, Big Data and AI, together to intelligently and efficiently accelerate breeding cycles, and finally promote the revolutionary change of crop breeding from “art”, to “sciences”, and finally to “intelligence”.

Key words: Maize; Breeding 4.0; Genomic selection; Machine learning

21 世纪的第二个 10 年是全球人工智能技术、基因工程技术、生命科学突飞猛进的 10 年, 同时也为

我国玉米育种领域步入崭新发展阶段提供了千载难逢的机遇。我国玉米科研领域虽然在玉米基因工程技术、玉米功能基因解析、玉米基因组学、玉米群体遗传学等基础研究方面取得了长足进展, 但育种现代技术应用方面与西方发达国家种业公司相比仍有很大差距。未来三十年内, 我国乃至世界将面临全球人口激增、气候持续变化、土地资源退化、环境污染等威胁粮食安全的诸多问题, 给作物育种行业带来新的挑战, 迫使育种科技亟需革命性的改变^[1]。提升我国玉米育种的科技竞争力、建立玉米育种自主创新体系、加强科研创新成果转化, 将传统

录用日期: 2019-01-10

基金项目: 国家自然科学基金项目“应用机器学习的策略预测玉米杂种优势”(31871706)

作者简介: 王向峰, 教授, 国家青年千人计划专家。

E-mail: sysbio@cau.edu.cn

才 卓, 研究员, 曾任国家玉米工程中心主任。

E-mail: caizhuo@126.com

致 谢: 中国农业大学国家玉米改良中心戴景瑞院士、李建生教授在本文撰写过程中给予了大力支持与悉心指导, 并提出了宝贵的修改意见, 特表诚挚感谢!

玉米育种技术与信息科学、生物技术等高科技手段的深度交叉融合,是保证我国玉米种业具有持续国际竞争力的核心。

1 农作物育种的发展历程

近百年来玉米育种模式的发展大体经历了3个时期,即主要依赖表型观察,通过自交加代选育优秀自交系的传统经验育种;以杂种优势群体划分模式为基础,筛选高配合力亲本组合为核心的杂种优势育种;综合了单倍体育种、分子标记育种、转基因育种的现代生物工程育种。2018年初,美国康奈尔大学玉米遗传育种学家、美国科学院院士 Edwards Buckler 教授提出了“育种4.0”的理念,即作物育种技术的发展伴随人类社会的进步已经历了3个标志性阶段,目前正跨入第四个阶段^[2]。

1.1 育种1.0阶段

覆盖过去上万年的作物育种发展历程,主要依赖对作物表型变异的肉眼观察,凭借耕作者的经验和主观判断,筛选具有符合农业生产所需要的产量、

品质与农艺性状的育种材料。该阶段主要是耕作选育具有广泛表型变异的地方农家品种,为现代栽培品种的培育奠定了遗传资源基础。

1.2 育种2.0阶段

始于十九世纪末,伴随着人类对生物遗传规律的认识以及田间统计学、数量遗传学在育种中的应用,由职业育种家通过预先设计杂交育种试验,选育现代栽培品种。

1.3 育种3.0阶段

贯穿二十世纪末至今,伴随现代分子生物学与基因工程的发展,采用分子标记辅助选择及转基因技术的手段,实现单一目标性状的导入与修饰。同时,基于连锁群体的功能基因克隆、自然群体的全基因组关联分析、复杂农艺性状的基因功能解析等研究领域发展迅猛,加深了人类对作物基因组及基因功能的认识。此外,各类高通量测序与芯片平台的发展使全基因组选择辅助育种技术在作物育种中逐步应用起来,现代育种技术达到新的高潮。

1.4 育种4.0阶段

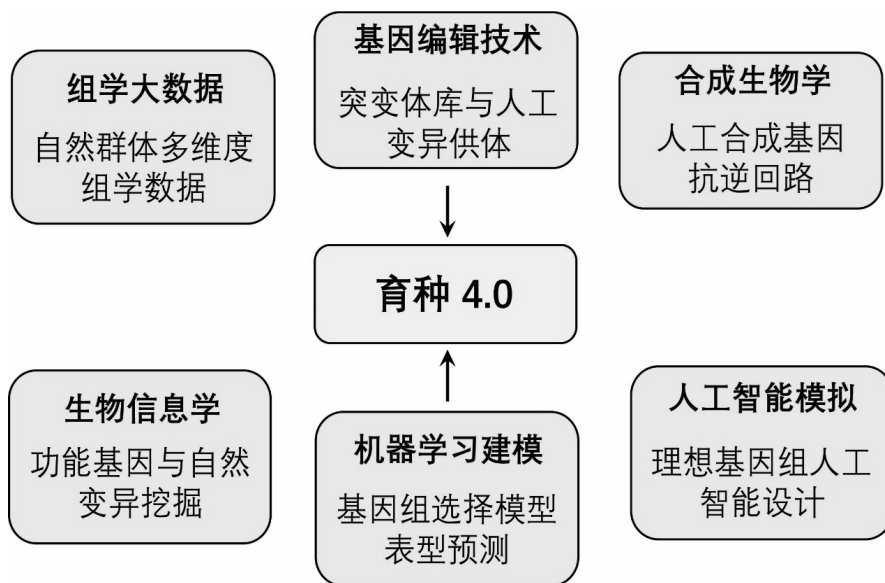


图1 依托多层面生命科学与信息科学推动育种技术向着智能化的方向发展

Fig.1 Directional development of breeding technology to intelligently rely on multi-layered life sciences and information science

是伴随人类社会步入互联网、大数据、人工智能育种4.0“三位一体”的时代提出的革命性育种理念,强调生命科学、信息科学与育种科学的深度融合^[2]。人工智能是计算机科学的一个分支,是指应用计算机算法和程序模拟人类意识与思维方式以及大脑处理信息的过程。人工智能的研究包括机器人、语言识别、图像识别、自然语言处理和专家智能决策系统等领域^[3]。机器学习方法学体系是人工智能的核

心,是指应用计算机算法建立模型解析数据,通过不断学习数据的自身特征训练模型,然后对现实世界中的事物做出判断和预测^[4]。如图1所示,育种4.0阶段的育种家将依托多层面生物技术与信息技术推动育种向着智能化的方向发展,即以基因组测序技术与人工智能图像识别技术为依托,通过基因型与表型数据的自动化获取与解析,实现玉米组学大数据的快速积累;以生物信息学与机器学习技术为依

托,通过遗传变异数据、各类组学数据、杂交育种数据的整合,实现作物性状调控基因的快速挖掘与表型的精准预测;以基因编辑与合成生物学技术为依托,通过人工改造基因元器件与人工合成基因回路,实现作物具备新的抗逆、高效等生物学性状;以作物组学大数据与人工智能技术为依托,通过在全基因组层面上建立机器学习预测模型,建立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,实现智能、高效、定向培育新品种。

育种4.0阶段的核心目标是建立“作物基因组智能设计育种”的跨学科、多交叉技术体系,该体系涵盖生命科学领域的基因组技术、表型组技术、基因编辑技术、生物信息学、系统生物学、合成生物学,以及信息领域的人工智能技术、机器学习技术、物联网技术、图形成像技术,共同支撑作物育种科学向更高的层面发展。到本世纪中叶,作物新品种的培育周期可以缩短至一年或数月,甚至实现在短时期内快速驯化出崭新的农作物品种为人类社会所用。育种4.0时代的作物基因组智能设计育种,将推动育种技术从“艺术”到“科学”到“智能”的革命性转变。

2 玉米基因组智能设计育种的大数据基础

2.1 玉米组学大数据资源

玉米基因组智能设计育种体系的建立必须以玉米基因组、表型组、转录组、蛋白组、代谢组等各类组学大数据为基础,采用机器学习数据挖掘的策略为玉米育种建立各种基因组预测模型,辅助育种家智能决策育种方案^[5]。随着二代测序与芯片等高通量基因型检测费用的大幅下降,我国现代玉米育种中广泛实际应用的亲本自交系材料、杂交后代材料的基因型数据将会呈指数积累。同时,转录组、蛋白组、代谢组等数据的积累也将逐步拓展到种质资源群体的水平,为进一步挖掘优良等位基因及调控元件、解析玉米发育过程与基因调控网络、解析玉米杂种优势分子机理、解析玉米抗虫抗病及胁迫应答等生物学过程提供广泛的数据资源^[6-9]。

表型组学是生命科学领域近年来兴起的一门新型交叉学科^[10]。作物表型组学研究发展迅猛,与人工智能、光学成像、图像识别技术在农业中的应用密不可分,同时这些技术也是未来农业发展精细耕作、智能装备的科学基础。未来玉米表型组的研究将注重开发田间农艺性状的采集装置,利用野外无人机、田间机器人、小区测产系统等农业机械搭载的智能装备完成田间作物表型的自动化获取。玉米表型组

装置配合物联网农情监测系统实现对局部地区气象数据、土壤数据、病虫害数据等环境参数的自动化采集,不仅可以用于基因与环境互作的分析,也可以用于对玉米表型预测精度的校正。

作物生理表型组将成为表型组研究的下一个热点^[11]。玉米等作物在病虫害、环境胁迫、营养匮乏等条件下,体内会发生一系列生理生化活动的变化。这些变化可以通过各类光学成像设备得以监测,如激光三维扫描成像、高光谱成像、多光谱成像、热成像、普通光学成像、叶绿素荧光动力学成像、近红外成像、雷达成像等,这些光学设备可以用来采集植物内部近百种生理指标。深度学习技术可以从生理表型组图像数据中提取特征波段,反演出植物在感病、感虫、胁迫等条件下的生理生化变化状态^[12]。数字化的表型组数据不仅可以对复杂的复合表型进行分解,也消除了肉眼性状调查引入的主观偏差,提高挖掘性状调控基因的精度。

2.2 全基因组选择辅助育种

分子设计育种的概念自上世纪末提出,即利用与性状连锁的DNA变异信息作为“分子标记”,对重要的性状决定基因进行筛选、优化、组合,模拟出满足不同育种目标的理想基因型^[13]。在过去三十年的时间里,分子设计育种在基因层面上辅助育种家选择亲本材料和设计杂交育种组配方案,其发展历程大体经历了以下3个阶段:单基因选择、多基因聚合以及全基因组选择辅助育种。与分子标记辅助育种不同的是,基因组选择育种是考虑基因组中数万甚至数百万的分子标记信息,在训练群体中建立基因组选择模型推导基因型与表型间的相关性,在候选群体中模拟和预测杂交后代可能产生的表型^[14],育种家可以根据杂种一代表型预测的结果选择育种价值较高的亲本材料,设计合理的杂交、回交育种方案。

玉米育种主要利用自交系间的杂种优势现象培育杂交种,是应用基因组选择育种的理想作物^[15]。玉米中应用基因组选择育种的流程分为如下几个步骤(图2):首先,选择来自不同杂种优势群的亲本自交系材料进行基因组重测序或SNP芯片基因分型,亲本群体的基因型数据需要生物信息分析进一步处理,并通过建立系统发育树对亲本材料的群体结构和亲缘关系进行初步解析;其次,选择一定比例的代表性亲本材料通过杂交实验建立F₁训练群体,并对杂种一代的表型进行完善、细致的性状调查,每个杂种一代的基因型则可以通过其双亲的基因型推导出来;再次,在已知基因型与表型的F₁训练群体中构建基因组选择模型,推导基因型与表型的相关性,模型

中可以进一步加入亲本表型、环境变量、群体结构等参数作为固定效应调整模型的预测精度;最后,在未知表型的候选群体中应用基因组选择模型进行预测,以模拟的 F_1 的基因型作为输入,预测杂种一代的开花期、株高、产量等表型。在开展玉米基因组选择育种实验中,代表性亲本自交系材料的选择,以及训练群体与候选群体的划分比例尤为重要。前期研究表明,在群体结构较为一致、亲本材料较为固定的前提下,对玉米开花期的预测精度可以达到0.75左右,对株高的预测精度可以达到0.85左右,对穗重和穗重中亲优势的预测精度可以分别达到0.65和0.80左

右。除了对表型性状的预测,基因组选择模型也可以对遗传力、一般配合力、特殊配合力、收获指数等育种中常用指标进行预测,辅助育种家选择优良亲本自交系。除了农艺性状相关表型的预测,通过全基因组关联分析挖掘的抗虫抗病、耐盐耐旱等胁迫应答相关分子标记也可以用于建立基因组选择模型。玉米自交系材料耐盐研究的结果表明,利用盐胁迫条件下获得的代谢组数据,与基因组数据关联分析挖掘的2 200个标记建立的基因组选择模型,对亲本材料耐盐等级评价的精度可以达到0.75~0.80的水平。

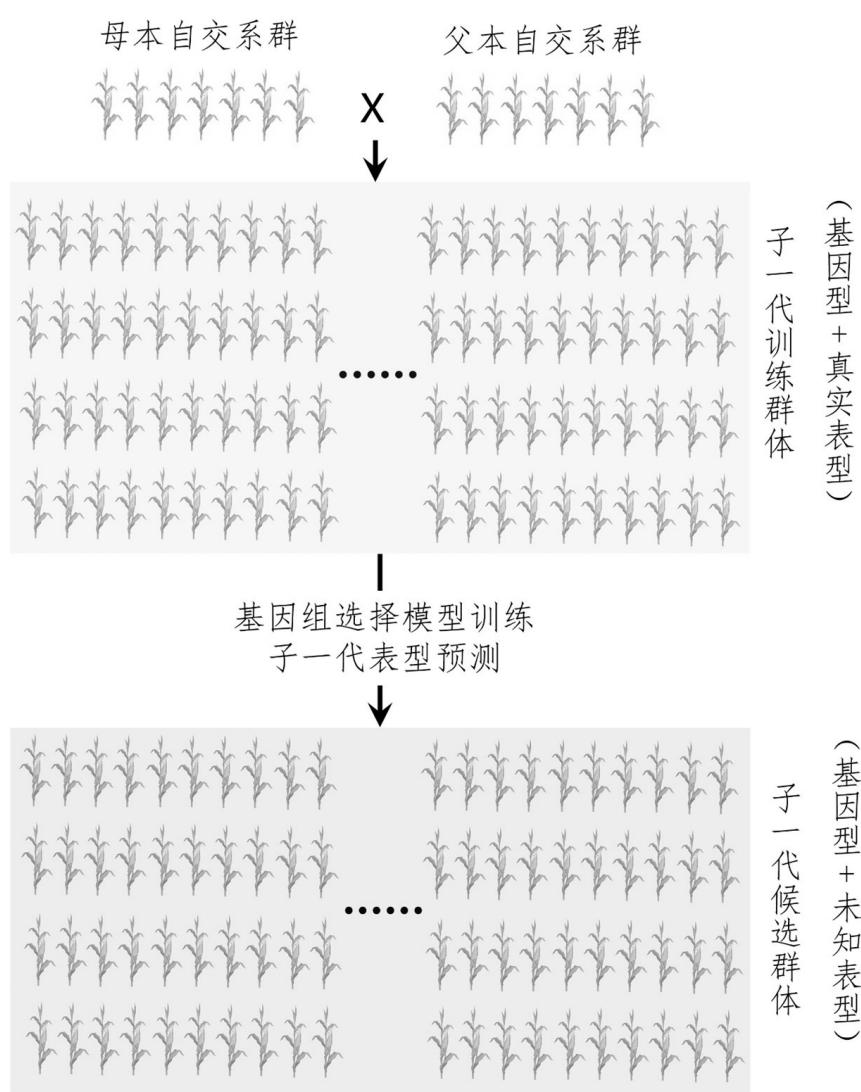


图2 基因组选择育种示意图

Fig.2 Schematic of genome selection breeding

传统基因组选择模型主要应用统计学中的混合线性模型建立基因型与表型间的相关性^[16]。但是,随着高通量基因组测序与表型组监测技术的飞速发展,混合线性模型已无法适应数百万标记和数万计

样本量的大数据分析 with 高性能并行运算。此外,由于玉米产量杂种优势的表现由非加性遗传效应所决定,并且受环境因素影响较重,使用混合线性模型预测产量杂种优势的表现仍存在一定的问題。为突破

混合线性模型的预测精度瓶颈,应用机器学习数据挖掘的策略,整合基因组、表型组、环境参数等多维数据源建立基因组选择模型,是基因组选择育种的下一步发展方向。该领域将重点尝试深度学习、集成学习等人工智能领域中的先进算法建立作物表型预测模型。机器学习技术建立的“基因型到表型”的预测模型育种应用将更为广泛,可以实现玉米基本农艺性状与产量预测、亲本配合力与子代杂种优势预测、玉米营养吸收与高效利用评价、玉米抗病虫与抗逆境评价、玉米最优环境生态区选择等预测功能,最终实现智能化、自动化地辅助育种家设计育种方案。

2.3 基因编辑构建突变体库

玉米基因组中大约有4万个编码蛋白的基因,但仅有不足200个基因得以克隆和功能验证。随着玉米转基因技术的逐渐成熟,采用基于CRISPR/Cas9基因编辑技术高通量敲除玉米基因,建立玉米全基因组范围的突变体库,实现快速筛选与鉴定调控重要农艺性状的关键基因。在作物中应用基因编辑技术对单个基因的功能敲除、改变单一目标农艺性状已有诸多成功的案例^[17]。如在水稻、玉米、大豆、马铃薯等作物中以乙酰乳酸合成酶基因(*ALS*)为靶点编辑创制的*ALS*突变植株,对乙酰乳酸合成酶抑制剂类除草剂的耐受性是野生型的1万倍。玉米中利用基因编辑技术通过对*Wx1*基因的修饰提高直链淀粉含量、对*TMS5*基因的修饰创制热敏雄性不育材料、对*ZmMTL*基因修饰创制单倍体诱导系材料、对*ARGOS8*基因修饰创制耐旱材料,为玉米遗传改良育种和分子设计育种提供重要的供体材料。

采用高通量定向基因编辑技术建立突变体库是作物功能基因组研究的重要方向。2017年6月,Molecular Plant以背靠背的形式发表了李家洋院士团队与百格基因研究团队利用CRISPR/Cas9基因编辑结合寡核苷酸芯片技术,在全基因组范围内对水稻基因进行编辑,构建水稻定向基因编辑突变体库^[18,19]。玉米定向基因编辑突变体库的建立可以参考水稻中的研究思路,在3~5年内逐步完成全基因组范围内的基因敲除。玉米突变体库建成后的工作重点将转向突变体表型的鉴定、功能基因的挖掘,以及关键农艺性状分子机制的解析,为玉米遗传改良与分子设计育种提供基础遗传材料。

2.4 人工合成抗逆基因回路

合成生物学是在系统生物学的基础上,采用工程系统设计的理念合成人工基因回路,为一个物种引入多基因参与的生物通路,使该物种具备新的生

物学功能^[20]。人工合成基因回路的最大优势是可以利用来自微生物等其他物种的外源基因,解决依靠改造作物自身基因组难以实现的性状改良的问题。合成生物学的方法在玉米育种中较为适合解决玉米抗逆性和环境适应性的问题。自然界中各种极端环境微生物包含大量的耐酸、耐盐、耐热、耐寒等基因资源。微生物中的合成生物学研究已积累了大量的抗逆基因元器件与模块。诸多研究表明,微生物与植物的抗逆通路存在生物学共性。随着玉米转基因技术的成熟,将微生物中现有的抗逆基因回路整合到玉米基因组中,快速培育高抗玉米新品种有望在不久的将来实现。

以玉米为底盘生物的合成生物学研究仍有大量前期基础研究工作需要开展,尤其是如何提高微生物的基因与玉米基因组胁迫应答系统的兼容性,以及建立基于玉米环境信号感应的智能反馈系统,控制和调节基因抗逆回路的开与关、高与低。未来10年内,玉米人工合成抗逆回路的研究将分阶段逐步开展,包括玉米抗逆基因元器件的挖掘、评价与筛选,玉米环境信号反馈通路与基因调控网络建模,玉米抗逆基因回路的智能化组装与稳定性表达等。最后,随着数据的积累与基因回路的反复优化,最终实现高适配性、高靶向性玉米人工合成抗逆基因回路的智能设计、组装、表达。

2.5 基因组智能设计育种体系

随着人类社会步入人工智能的新纪元,未来作物育种模式将发生历史性的变革。玉米育种科学将在组学大数据与生物信息学、基因编辑技术与合成生物学、人工智能与机器学习技术等多学科、多领域的共同支撑下,推动玉米育种快速迈进育种4.0阶段。未来十年到二十年里,我国玉米育种领域的核心发展目标之一是创建“玉米基因组智能设计育种”技术体系,集成生命科学与信息科学领域的技术优势,全方位提高我国玉米育种科技的国际竞争力。所谓“基因组智能设计”的核心体现,是应用人工智能模拟的方法,为某一育种群体材料人工设计聚合了所有优势基因,具有“理想基因型”的虚拟基因组;再用机器学习模型预测虚拟亲本基因组与测试亲本基因组组配产生的杂交后代的“理想表型”。所谓“理想表型”则是该育种群体所能创造的杂种后代性状的理论上限,是该育种群体的杂种优势利用的潜力极限。根据智能设计的理想基因型,建立亲本自交系创制方案和杂交育种方案,快速逼近杂交后代的理想表型。

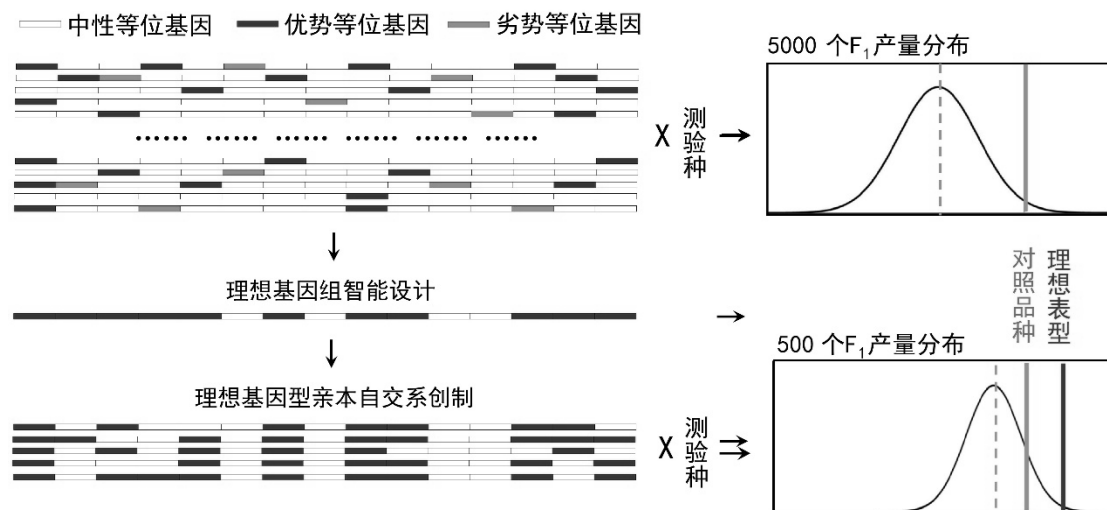


图3 玉米基因组智能设计育种示意图

Fig.3 Schematic diagram of maize genome intelligent design breeding

常规玉米杂交育种的应用场景如图3所示,假设某一育种群体包含5 000个亲本自交系作为母本,与某一骨干父本作为测验种杂交,获得5 000个 F_1 的产量分布与对照品种进行比较,筛选比对照品种产量提高的组合。玉米基因组智能设计育种的流程则分为以下4个环节:首先,鉴定亲本的基因型,以及与测验种杂交后 F_1 的产量表型;其次,通过对5 000个母本自交系的优势等位基因型、劣势等位基因型进行推导,建立针对该杂交育种群体最优的理想基因组,该理想基因组聚合了该群体中所有的优势等位基因,以及尽量排除了所有劣势等位基因,虚拟的理想基因组与测试种杂交获得表型即该群体理想产量的理论上限;第三,根据理想基因组聚合优势等位基因的规律,从母本自交系中选择一定比例的材料,创制新的优良母本自交系材料,如500个理想基因型自交系;第四,500个理想基因型自交系与测验种杂交获得的产量分布更加接近理论上限,同时有更高比例的组合高于对照品种的表现。

3 基因组智能设计育种商业模式—玉米育种4.0

作物育种4.0阶段的核心是大数据驱动的基因组智能设计育种,其跨学科、多技术交叉的特点决定了育种4.0阶段的育种商业运作模式需要多学科领域的技术支撑与产学研一体的研发链条,围绕育种行业的需求为总体目标,整合生命科学、信息科学、数据科学、人工智能等领域的技术专长,推动我国玉米育种产业向着工程化、智能化的方向发展(图4)。

最终,通过育种大数据与人工智能决策,实现辅助育种家精准筛选玉米优良亲本组合,快速创制玉米优良自交系,有效地缩短育种周期,提高育种效率,降低育种成本。

3.1 玉米基因型大数据体系

大数据驱动的智能育种运作模式的前提条件是标准化大数据体系,即建立来自玉米科研领域与育种行业的基因型与表型数据的规范化采集、处理、提炼、分析、存储与管理的数据库系统。玉米科研领域主要贡献种质资源群体的基因组、转录组、代谢组、蛋白组等数据类型,通过多维度生物组学数据的联合分析,挖掘株型、产量、抗病、抗虫、耐盐、耐旱、耐寒等性状相关的重要基因与自然变异,为分子设计育种与基因人工改造提供素材。商业化基因编辑与转基因技术公司为玉米基因改造提供专业化技术服务,大规模建立玉米定向基因突变体库与创制育种供体材料。以玉米为底盘细胞的合成生物学将逐步开展起来,借助极端微生物中的抗性基因创造人工合成抗逆基因回路,突破玉米自身基因抗胁迫的瓶颈。我国也将建立为育种行业提供数据共享服务的玉米基因型与分子标记数据库,为商业育种提供有效的分子检测标记资源。随着二代测序、SNP芯片等不同通量的基因型检测成本的降低,为商业育种的骨干亲本、测试亲本、亲本衍生系、DH系等不同群体提供更加灵活和便利的基因分型解决方案。与杂交组合测试实验的费用相比,越来越多的育种企业将选择更低成本的基于基因型的表型预测辅助每年育种方案的设计和决策。

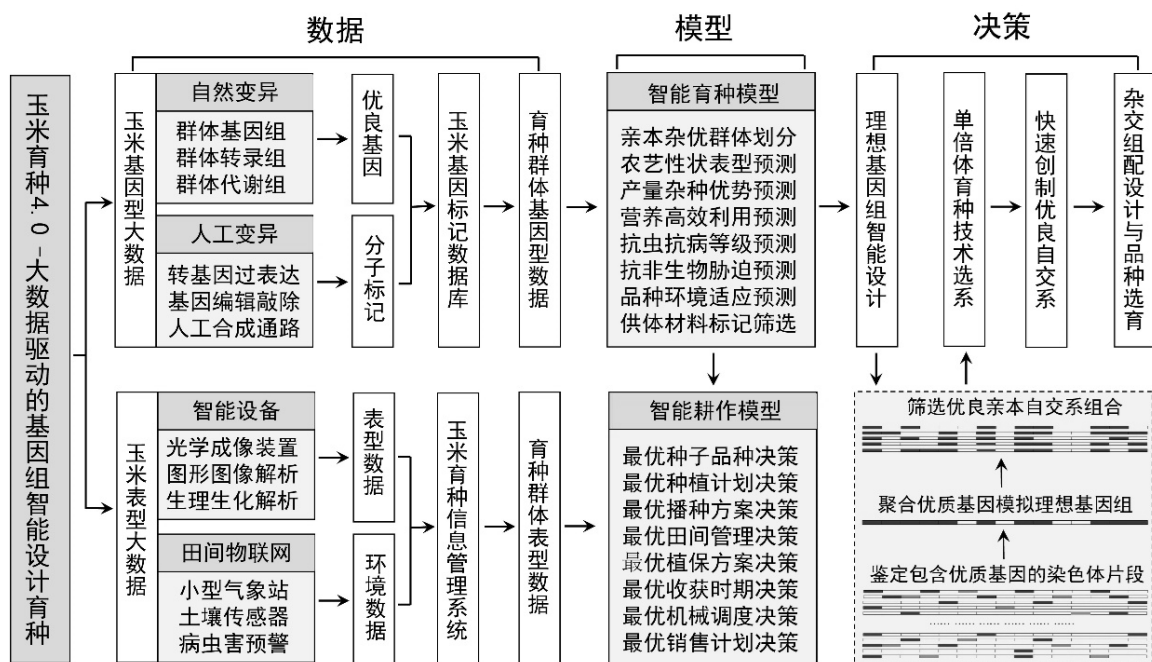


图4 玉米育种4.0商业模式

Fig.4 Business model of corn breeding 4.0

3.2 玉米表型与环境大数据体系

在图形图像技术、人工智能技术、物联网技术的支撑下,玉米育种实验数据与种植生产数据将指数增长,逐步形成玉米表型与环境大数据体系。未来10年里,将有一大批商业化农业服务科技公司崛起,为玉米种业行业提供表型监测的各类技术服务。田间表型数据的采集使用小型无人机、野外机器人、农业机械装备搭载的光学设备、雷达设备在不同玉米生长时期内自动化采集表型图像。图形图像数据经过专业的深度学习模型数字化处理,转换成育种中常用的标准化农艺性状数据。用于采集玉米生长发育过程中的生理生化指标的离子检测传感器等便携式设备将逐步应用起来,用于生理表型数据的获取。生理表型组数据将有助于挖掘抗逆相关基因,以及氮磷钾等营养物质高效吸收利用材料的筛选。作物的田间长势、产量性状与环境因素息息相关。应用农业物联网技术配合田间农情监测系统对测试区域的气象与土壤数据、虫害与病害数据、生长与产量数据,以及收获后测产数据进行自动化采集。各类环境数据通过通信网络传输至云端服务器进行处理、分析、存储与管理,实现对育种实验站点的常年多点式追踪。环境因子数据整合到预测模型中可以显著提高模型预测精度,更加精准的预测产量杂种优势、基因型与环境互作,为玉米品种最优种植生态区的选择提供决策。

3.3 玉米基因组育种模型体系

玉米育种材料基因型数据与田间表型数据的集中式采集与管理为建立育种智能决策模型提供广泛的训练集合。在国际知名种业公司的商业化育种体系中,育种信息管理系统、育种价值评估模型、分子设计育种工具、育种决策工具等一系列信息管理与数据分析模块是衔接整个育种流程中每个环节的重要节点,用数据决策每个环节的育种方案。我国育种4.0阶段的重要任务之一是结合中国的实际情况,为我国玉米种业行业建立工程化育种与信息管理体系,应用大数据与机器学习技术建立一系列育种智能决策模型,覆盖育种流程中的每一环节,为育种企业提供数据分析与决策服务,包括基因型与表型数据的自动化处理、亲本自交系材料的杂优群体划分、玉米基础农艺性状的表型预测、玉米产量与杂种优势的预测、一般配合力与特殊配合力的预测、营养高效利用与吸收效率的预测、育种材料的抗虫与抗病等级评价、抗逆境与非生物胁迫等级评价、玉米品种的环境适应性与最优种植生态区预测,以及供体亲本材料的标记筛选等育种决策服务。未来将有一大批提供育种数据分析与决策服务的农业科技公司涌现,推动玉米育种数据的规模化与标准化,逐步形成针对我国玉米育种材料的训练群体,提高育种决策模型的普遍适用性。

3.4 玉米精准种植模型体系

玉米育种模型与种植模型是相辅相成的。大数据驱动的精准种植模型重点结合天气、土壤、遥感等环境大数据,针对某一玉米杂交种品种建立的气候模型、土壤模型、播种模型、植保模型、营养模型,以及该品种的地上生长模拟模型、根系生长模拟模型,为玉米种植者提供实时、高效、精准的耕作管理决策。气候模型可以通过调取过去10年目标种植区域的卫星遥感图像,以及过去20年气候情况、地形变化、耕作制度等信息,进行综合分析建模。同时,根据拟播种玉米品种的生长特性,计算目标种植区域的气温、降水量、病虫害等环境限制因素,预测拟播种品种的潜在产量或推荐最佳种植品种,确定最佳种植密度和最优单产产量。土壤模型可以通过对耕作土壤地表下100 cm的深度范围进行横切面分析,与全球范围收集的500个土壤样本进行对比,确定土壤土质类型;同时,提取土壤样本在专业化实验室对土壤内的化学元素进行分析建模,精确计算目标区域的土壤供肥能力。播种模型可以根据过去8~10年的历史播种日期、地表温度与5 cm土层土壤温度、历史与实时气象数据、未来10 d降雨与气温预测最佳播种日期,确保播种后8 d内的100%出苗率并达到出苗整齐。在玉米生长过程中,可以利用玉米发育模型推断玉米生长发育进程和根系生长情况是否正常;利用便携式叶绿素仪、离子传感器等设备动态监测玉米植株内营养成分是否正常;同时,利用营养模型预测最佳追肥时间与追肥数量。病虫害预测模型可以根据当年的气温、湿度等情况以及孢子捕获器的数据推测疾病防控的最佳时期,确定农药喷洒的剂量和时间。目前在我国精准种植决策已作为农业科技服务的项目之一,优先在玉米种植中发挥作用。实践表明,通过玉米的精准种植决策,在单产产量不变的情况下有效地降低种植管理成本15到30个百分点,实现成本效益最大化。

3.5 玉米育种方案设计决策体系

玉米育种模型的分析与预测结果将整理成标准化报告,辅助育种家筛选育种材料与设计育种流程。基因组设计育种的智能决策主要在基因组选择的水平上体现两个层面的内容:其一是根据杂种一代的表型预测结果筛选一般配合力和特殊配合力整体优于对照品种的优良亲本自交系,建立杂交育种组合实验方案,该步骤基本可以直接获得可接近品种的玉米杂交种;其二是采用基因组智能设计算法,模拟适用于该育种群体的亲本理想基因组与子代理理想表型,以理想表型为目标,设计亲本自交系的杂交方案并结合单倍体育种技术实现快速创制新的优良

自交系。亲本理想基因组理论上聚合了该育种群体的自交系所涵盖的大部分优势等位基因,同时排除了大部分劣势等位基因。亲本理想基因组与测试亲本杂交产生的子代理理想表型,尤其是产量及其杂种优势相关性状,基本代表了该育种群体所能实现的理论产量的上限。有了理想基因组作为参考,随即可以根据理想基因组中包含优势等位基因的染色体片段的重组规律,从众多亲本自交系中遴选包含优良染色体片段较多的少数亲本组合,建立最优杂交组合方式,结合单倍体育种技术实现优良亲本自交系的快速创制。

3.6 中国育种4.0商业模式初探

育种4.0商业模式的建立需要以数据为基础。与国际种业公司完善的育种实验数据管理系统与基因组选择预测模型相比,我国玉米育种行业内的基因型、表型、环境数据缺乏统一、规范化的管理。因此,当务之急是建立符合我国玉米育种实际情况的基因型、表型与环境数据管理标准体系,以及在此基础上建立起普遍适用于我国玉米育种行业的基因组选择育种的训练群体与预测模型,作为公共平台为我国玉米种业行业提供育种决策服务。

建立公共基因组选择训练群体的基础材料可以由我国玉米育种行业和玉米学科群内的单位提供。首批标准训练群体可以主要针对我国东华北与黄淮海生态区选择应用较为广泛、在各个杂种优势群中有较高代表性的200个左右优良骨干亲本自交系作为基础群体,每个基础群体杂交产生的 F_1 代杂交种再通过单倍体加倍生产200个左右的DH系。DH系群体再与1到2个公共测验种杂交获得 F_1 代杂交种训练群体。那么,标准训练群体将包括4万到8万个左右的 F_1 代杂交种。通过诱导优良基础群体建立的大规模训练群体,不仅提高了群体的遗传多样性,同时也聚合了高产、抗逆等优良等位基因,更加适合基因组选择预测模型的建立以及理想基因组的模拟。DH系亲本群体的基因型通过统一的SNP芯片或SNP捕获测序的方式进行检测,并与测试种的基因型合并模拟出 F_1 代杂交种的基因型。表型性状数据将按照统一、规范的采集方法进行三年五点的调查、录入、处理、存储与管理。该标准训练群体首先将用于实现基本农艺性状与产量的预测,在此群体基础之上再设计更多的胁迫、营养、抗病、抗虫等实验逐步积累数据,最终实现亲本配合力与杂种优势预测、玉米营养吸收与高效利用评价、玉米抗病虫与抗逆境评价、玉米最优环境生态区选择等各个方面的育种决策。此外,建立该标准训练群体的基础自交系

材料是固定的,更有助于研究优良等位基因型的组合模式与优良农艺性状的关系,为理想基因组智能设计与新的优良自交系的创制提供参考。

4 结 语

育种4.0时代的育种技术变革依赖于基因组技术、大数据技术、人工智能技术的紧密结合。本世纪前二十年生物技术与人工智能技术的突飞猛进,为我国玉米育种行业的历史性变革与国际竞争力的提升带来前所未有的机遇与挑战。未来五到十年的时间里,我国玉米育种学科群将紧紧把握历史机遇、协同创新建立适合我国玉米育种产业的育种4.0技术体系。育种4.0的实现将以玉米育种为突破口,逐步拓展到水稻、小麦、大豆等农作物、园艺作物与经济作物的育种应用中,在未来10到20年内实现育种从“艺术”到“科学”到“智能”的革命性转变。

参考文献:

- [1] How to feed the world in 2050[C]. Food and Agriculture Organization of the United Nations, 2009.
- [2] Wallace G, Rodgers-Melnick E, Buckler E. On the road to breeding 4.0: unraveling the good, the bad, and the boring of crop quantitative genomics[J]. Annual Review of Genetics, 2018, 52(1): 421–444.
- [3] Webb S. Deep learning for biology[J]. Nature, 2018, 554: 555–557.
- [4] Ma C, Xin M, Feldmann K A, et al. Machine learning-based differential network analysis: a study of stress-responsive transcriptomes in Arabidopsis[J]. Plant Cell, 2014, 26(2): 520–37.
- [5] Ma C, Zhang H H, Wang X. Machine learning for Big Data analytics in plants[J]. Trends in Plant Sciences, 2014, 19(12): 798–808.
- [6] Jiao Y, Zhao H, Ren L, et al. Genome-wide genetic changes during modern breeding of maize[J]. Nature Genetics, 2012, 44(7): 812–5.
- [7] Li H, Peng Z, Yang X, et al. Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels[J]. Nature Genetics, 2013, 45(1): 43–50.
- [8] Wang X, Wang H, Liu S, et al. Genetic variation in *ZmVPP1* contributes to drought tolerance in maize seedlings[J]. Nature Genetics, 2016, 48(10): 1233–1241.
- [9] Luo J. Metabolite-based genome-wide association studies in plants[J]. Current Opinion in Plant Biology, 2015, 24: 31–38.
- [10] Houle D, Govindaraju D R, Omholt S. Phenomics: the next challenge[J]. Nature Review Genetics, 2010, 11(12): 855–866.
- [11] Ghanem M E, Marrou H, Sinclair T R. Physiological phenotyping of plants for crop improvement[J]. Trends in Plant Sciences, 2015, 20(3): 139–144.
- [12] Ubbens J R, Stavness I. Deep plant phenomics: a deep learning platform for complex plant phenotyping tasks[J]. Frontiers in Plant Sciences, 2017, 8: 1190.
- [13] Wang J, Li H, Zhang X, et al. Molecular design breeding in crops in China[J]. Acta Agronomica Sinica, 2011, 37(2): 191–201.
- [14] Crossa J, Pérez-Rodríguez P, Cuevas J, et al. Genomic selection in plant breeding: methods, models, and perspectives[J]. Trends in Plant Sciences, 2017, 22(11): 961–975.
- [15] Covarrubias-Pazaran G. Genome-assisted prediction of quantitative traits using the R package *sommer*[J]. PLoS ONE, 2016, 11(6): e0156744.
- [16] Endelman J. Ridge regression and other kernels for genomic selection with R package *rrBLUP*[J]. The Plant Genome, 2011, 4(3): 250–255.
- [17] Zhang Y, Massel K, Godwin I D, et al. Applications and potential of genome editing in crop improvement[J]. Genome Biology, 2018, 30: 19(1): 210.
- [18] Meng X, Yu H, Zhang Y, et al. Construction of a genome-wide mutant library in rice using CRISPR/Cas9[J]. Molecular Plant, 2017, 12: 10(9): 1238–1241.
- [19] Lu Y, Ye X, Guo R, et al. Genome-wide targeted mutagenesis in rice using the CRISPR/Cas9 system. Molecular Plant, 2017, 12: 10(9): 1242–1245.
- [20] Brophy J A, Voigt C A. Principles of genetic circuit design[J]. Nature Methods, 2014, 11(5): 508–520.
- [21] 林章凇, 林 敏. 微生物和植物抗逆元器件的合成生物学研究[J]. 生物产业技术, 2013(4): 7–11.
- Lin Z L, Lin M. Synthetic biology of microbial and plant stress-resistant components[J]. Biotechnology, 2013(4): 7–11. (in Chinese)

(责任编辑:李万良)