

DOI: 10.19788/j.issn.2096-6369.190201

植物表型组学大数据及其研究进展

赵春江^{1,2,3}

- (1.北京农业信息技术研究中心, 北京 100097;
- 2.国家农业信息化工程技术研究中心, 北京 100097;
- 3.数字植物北京市重点实验室, 北京 100097)

摘要: 植物表型组学通过集成自动化平台装备和信息化技术手段, 获取多尺度、多生境、多源异构植物表型海量数据, 形成植物表型组学大数据, 从组学高度系统深入地挖掘“基因型-表型-环境型”内在关系、全面揭示特定生物性状的形成机制, 将极大地促进功能基因组学、作物分子育种与高效栽培的进程。本文概括了植物表型组学大数据的发展背景、含义、产生过程和特点, 系统综述了植物表型组学大数据研究进展, 包括植物表型数据获取与解析、植物表型组大数据管理及建库技术、表型性状预测和基于表型组的多重组学分析的进展; 从植物表型数据采集标准、多样化表型配套设施和低成本表型设备研发、开放共享植物表型组大数据平台构建、表型大数据融合与挖掘理论方法、植物表型组学协同共享和互作机制五个方面探讨了当前植物表型组学大数据研究与应用中面临的问题和挑战; 最后从加强植物表型组技术体系设计与标准研究、植物表型-环境感知机理研究和智能化设备研发、植物表型组大数据建设以及人才队伍和协作网络建设四个方面提出具体建议。

关键词: 植物表型组学; 大数据; 数字植物; 数据挖掘; 数据管理; 数据获取; 性状预测; 植物表型组大数据平台
中图分类号: S-1 **文献标识码:** A **文章编号:** 2096-6369 (2019) 02-0005-14

本文引用格式: 赵春江. 植物表型组学大数据及其研究进展[J]. 农业大数据学报, 2019, 1(2): 5-18.

Zhao Chunjiang. Big Data of Plant Phenomics and Its Research Progress[J]. Journal of Agricultural Big Data, 2019, 1(2): 5-18.

Big Data of Plant Phenomics and Its Research Progress

Zhao Chunjiang^{1,2,3}

- (1. Beijing Research Center for Information Technology in Agriculture, Beijing 100097;
2. National Engineering Research Center for Information Technology in Agriculture, Beijing 100097;
3. Beijing Key Laboratory of Digital Plant, Beijing 100097)

Abstract: Plant phenomics is capable of acquiring gigantic multi-dimensional, multi-environment, and multi-source heterogeneous plant phenotyping datasets through integrated automation platforms and information retrieval technologies, based on which the big-data driven plant phenomics research is established. This emerging research domain aims to systematically and thoroughly explore the internal relationship between "gene-phenotype-environment" at the omics level, so that phenomics methods can be utilized to unravel the formation mechanism of specific biological traits in a comprehensive manner. As a result, it is greatly catalyzing the research progress of functional genomics, crop molecular breeding, and efficient cultivation. In this paper, we summarized the

收稿日期: 2019-05-05

基金项目: 北京市农林科学院协同创新中心建设专项 (作物表型组学协同创新中心)

作者简介: 赵春江, 男, 研究员、中国工程院院士, 研究方向: 农业信息技术与智能装备; E-mail: zhaocj@nrcita.org.cn

background, definition, initiation, and features of the big-data driven plant phenomics, followed by a systemic overview of the progress of this field, including the acquisition and analysis of plant phenotyping data, data management and relevant database construction techniques for administering big data generated, the prediction of phenotypic traits, and its connection with the plant omics research. Furthermore, this paper focuses on discussing present problems and challenges encountered by both plant research and related applications, including (1) the standardization of collecting plant phenotypes, (2) research and development (R&D) of diverse phenotyping devices, supporting facilities, and low-cost phenotyping equipment, (3) the establishment of big data platforms that can openly share phenotyping data and phenotypic traits information, (4) theoretical approaches for fusion algorithms and data mining techniques, and (5) collaborative, sharing and interactive mechanisms for the plant phenomics community to adopt. Finally, the paper puts forward suggestions in four aspects that need to be strengthened: (1) systematic design and standards of plant phenomics research, (2) revealing the mechanism of plant phenotype and environment to facilitate intelligent equipment R&D, (3) the establishment of big data for plant phenomics, and (4) the formation of collaborations through academic networks and specialized research groups and laboratories.

Keywords: plant phenomics; big data; digital plant; data mining; data management; data acquisition; trait prediction; big data platform of plant phenomics

1 引言

进入 21 世纪以来, 物联网、移动互联网、大数据、云计算和人工智能等新一代信息技术的发展及其与各行业的深度融合, 正加速推进人类社会向大数据、智能化时代迈进。数据, 已经渗透到当今每一个行业和业务职能领域, 成为重要的生产因素。海量聚集的大数据隐含着巨大的社会、经济和科研价值, 吸引着各学科、各行业乃至各国政府的高度关注^[1]。目前, 以大数据和人工智能技术为核心的算力正成为智能时代的生产力^[2], 并且已经为商业、金融、制造业和医疗等领域带来了深刻变革, 在推动重大科学发现、前沿技术突破和产业模式创新的同时促进传统产业转型升级。

大数据时代, 农业科学正在从理论科学、实验科学和计算科学步入以数据密集型知识发现为研究范式的学科发展阶段^[3]。特别是作物育种技术正在进入以基因组和信息化技术高度融合为主的育种 4.0 阶段^[4], 直接推动作物科学乃至植物科学进入以大数据为核心的组学研究时代^[5]。进入组学时代的农业基础科学, 迫切需要组织开展农学、生命科学、信息科学和工程科学等多学科交叉合作研究, 获取并解析表型组 - 基因组 - 环境大数据, 从组学高度系统

深入地挖掘“基因型 - 表型 - 环境型”内在关联、全面揭示特定生物性状形成机制, 对揭示作物生命科学规律、提高作物功能基因组学和分子育种研究水平等具有重大意义。

随着基因测序和环境传感技术的快速发展, 农业科学领域涌现出越来越多的基因组和环境大数据。同时, 在科研需求的驱动和技术方法快速发展的背景下, 当前已经具备进行高通量、多生境、多维度、精准和经济的植物表型数据获取和解析的必要性和可行性。植物表型组学及其大数据研究引起各国政府的高度关注, 并已成为植物科学领域重要的战略布局方向。

2 植物表型组学大数据

2.1 植物表型组学大数据发展背景

1990 年 10 月 1 日, 美国启动了“人类基因组计划”(Human Genome Project, HGP), 标志着人类正式进入了基因组学(Genomics)时代。在“人类基因组计划”的激励下, 各类微生物、植物和动物的基因组测序工作持续推进, 生命科学领域的数字呈现指数级增长趋势。二十一世纪尤其是 2010 年以来, 随着新一代测序技术、超级计算、人工智能

技术的发展,海量基因组大数据时代来临(从GB、TB级到PB、EB级),例如Illumina公司最新的HISEQ X TEN测序仪3天的测序可获得1.8TB的碱基对数据^[6]。在此期间,植物基因组学研究也取得了巨大进展,水稻、玉米、大豆和小麦等主要作物的基因组已相继完成测序;在“十三五”国家重点研发计划“七大农作物育种”项目中,规划要对17000余份重要种质材料进行全基因组水平的基因型鉴定。通过获得海量基因型数据,作物基因组大数据及数据库建设工作正在快速推进。

测序技术的普及化和待测材料的规模化带来一个新的问题:缺乏合适的高通量表型获取技术获取对应的表型信息。在获得海量作物基因组信息的基础上,如何高分辨、高效地解析基因功能、植物表型及环境响应三者的相互作用机理,以及植物表型与产量、品质和抗性之间的关系,已成为一个全新的挑战^[7]。在此背景下,植物表型组学(Phenomics)应运而生^[8]。

2.2 植物表型组学大数据的含义

植物表型是指能够反映植物细胞、组织、器官、植株和群体结构及功能特征的物理、生理和生化性状,其本质实际是植物基因图谱的时序三维表达及其地域分异特征和代际演进规律^[9,10]。碱基序列构成了基因组数据分析的基本单元,通过对植物样本进行测序即可获得包含基因组全部基因排列和间距信息的一维物理图谱,除少量基因突变外,碱基序列可认为是客观存在的。而植物表型是这些基因与环境相互作用后在时空中的三维表达,它是植物基因在与环境的相互作用中其遗传信息被选择性地表征,并完成一个动态的生活史^[11]。因此,植物表型组所包含的信息量和复杂程度已远远超出人们的预估。针对如此复杂的植物表型特征,植物表型组学旨在集成自动化平台装备和信息化技术手段,系统、高效地获取植物表型信息,并结合基因组学、生物信息学和大数据分析等最新理论技术,从组学高度系统深入地挖掘“基因型-表型-环境型”内在关联,以揭示植物多尺度结构和功能特征对遗传信息和环境变化的响应机制,是突破未来植物科学研究和应

用的一门重要的交叉学科^[9,12,13]。随着植物科学、计算机科学与工程科学等领域科学研究的不断协同,通过环境传感、无损成像、光谱分析、机器人技术、机器视觉、激光雷达和理化分析等手段采集的植物表型数据正逐渐形成一种科学大数据,即植物表型组学大数据,它涵盖植物从细胞到群体的各个尺度、植物性状在多生境下的遗传与变异,以及植物对生物和非生物胁迫的响应等信息。

2.3 植物表型组学大数据的产生过程

植物表型组学大数据包含了从基因与环境相互作用形成的植物表型原始数据、植物表型性状元数据,最终到生物学知识的全集数据,它涵盖基因、生理、生化、生态及生长动态等多维尺度。植物表型组大数据的产生和形成是一项巨大的科学系统工程^[14],它以作物栽培和植物育种的实际需求为导向,依赖表型平台、传感设备、无线通信、数据库和大数据分析等现代信息技术和机械装备,需要农学、植物学、自动化、机械工程、图形图像和计算机科学等多学科人员在大数据形成的各环节紧密协作,进而对植物表型数据进行获取、解析、管理和挖掘,将植物表型组大数据最终转化为生物学、农学新知识(图1)。如德国LemnaTec公司研发的Field Scanalyzer^[15]表型平台,包括龙门吊式行走装置、机械运动自动控制模块、高精度传感阵列、配套的数据采集和分析软件等,具备产生、管理和分析植物表型大数据的能力。LemnaTec的植物表型平台及配套软件和服务已先后被法国农业科学院、中国科学院和杜邦先锋公司等科研机构与企业所采购,应用于植物育种研究工作。

2.4 植物表型组学大数据的特点

植物表型组大数据具有传统大数据的典型3V特点^[16]:(1)数据量大(Volume),以智能化装备和人工智能技术为依托,利用先进表型技术设备获取的表型数据量迅速增加,例如,利用温室高通量表型平台监测玉米动态生长发育,1000盆植株每天产生的各类图像数据可达TB级;利用micro-CT获取籽粒显微图像,500份籽粒的图像信息可达到

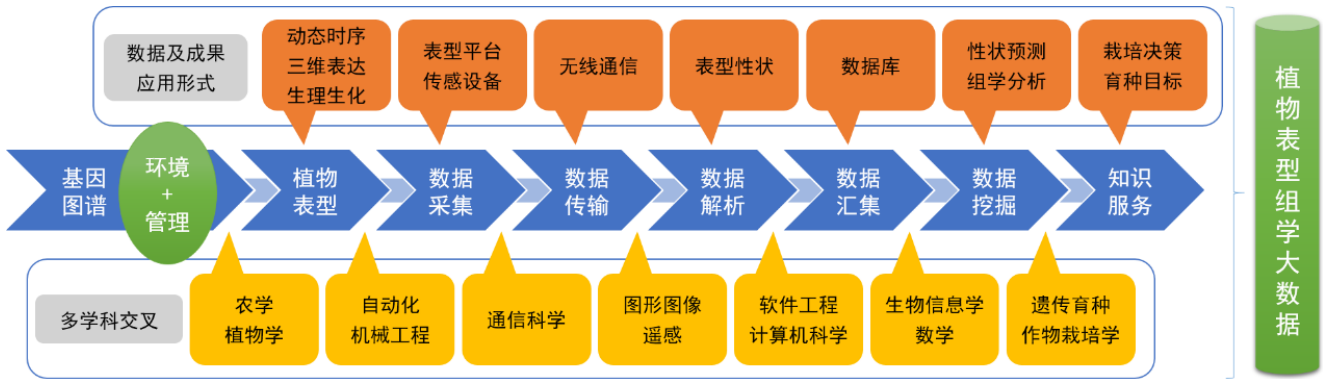


图 1 植物表型组学大数据形成过程

Fig.1 The formation of big data in plant phenomics

注：此处多学科交叉所述专业仅为列举，各环节包含但不限于所述专业。

20TB^[17]；(2) 数据的多态性 (Variety)，植物个体和数据类型的多样性、异质性决定了表型数据的多态性，不仅包括从地下根系和地上植株表型信息，室内到野外条件下从细胞、器官、植株到群体水平的表型数据，也包括近地面到天空的遥感数据^[18]；(3) 数据的时效性 (Velocity)，表型大数据往往以数据流的形式动态、快速地产生，如室内植物表型平台、大田植物表型平台、无人机平台等搭载可见光、近红外、远红外、光谱等传感器获取大量流数据和实时数据。正是因为表型组学大数据具有典型的 3V 特点，需要依靠大数据思维和数据分析策略对植物表型组数据进行清洗、控制、挖掘和转化。同时植物表型组学大数据还具有 3H 特性^[12]：(1) 高维度 (High dimension)，植物表型组学大数据是植物遗传信息与资源环境相互作用后在时空时序三维表达信息的汇集，不仅包括文本数据、试验元数据，还涉及图像和光谱数据、三维点云数据、时序生长数据等，多尺度、多模态、多生境的表型数据决定了其具有高维特点。这些高维度数据为发掘蕴含于植物表型大数据中的深刻规律奠定了基础，同时在大数据整合与分析方面提出了挑战^[19]；(2) 高度复杂性 (High complexity)，植物表型大数据作为植物遗传信息与环境作用下完整生活史的象征，遗传信息的多样性以及地域资源环境的差异性决定了植物表型组学大数据的高度复杂性；(3) 高度不确定性 (High uncertainty)，由于植物遗传信息被选择

性的表征，这种表征受植物生长的地理位置、光温水气热等资源的不同形成差异显著的表型，加之植物表型大数据样本来源广、处理方法多、数据获取标准不统一、存储格式多样，导致表型数据具有低重复性和不确定性的特点^[20,21]。

针对植物表型组学大数据的上述特征，各学科研究人员围绕植物表型组学开展了一系列研究，以期解决植物表型组学大数据在获取、解析、管理、挖掘与应用等方面的难题，进而推动植物科学的快速发展。

3 植物表型组学大数据研究进展

植物表型组学大数据的研究体系一般可以分成三个部分，分别为获取与解析、数据管理以及挖掘与应用。其中获取与解析是指利用传感设备获取表型原始数据，并将其转化为具有明确生物学含义的植物表型性状；数据管理是对全链条大数据的存储、管理与共享；挖掘与应用是实现植物表型组学大数据价值化的重要途径。

3.1 植物表型数据获取与解析

21 世纪以来，以各类新型物理、化学和生物 (生理) 传感器、图形图像技术、人工智能技术及物联网技术为代表的新一代表型获取技术体系，正在为植物研究提供海量表型和环境数据源。1998年，

比利时 CropDesign 公司成功开发世界上首套大型植物高通量表型平台,命名为 TraitMill^[22],该平台打破了几百年来“一把尺子一杆秤”的植物表型性状获取瓶颈,可高通量、自动化获取包括地上部分生物量、株高、总粒数、结实率、粒重、以及收获指数等植物表型信息。此后,环境可控的室内植物表型平台、大田植物表型平台、低成本、便携式表型采集设备及航空机载平台等多层次表型获取技术迅猛发展,整合图像、点云、光谱、红外、X 射线等传感器采集了细胞-组织-器官-植株-群体多尺度的表型数据(见表 1),形成了从室内、温室到田间

的多生境、数万亿字节(terabytes data)的大规模数据^[23]。目前,植物表型数据获取研究者致力于构建表型基础设施、研发便携式低成本的表型获取装置,努力进一步提高表型数据获取的吞吐量、分辨率和自动化程度。

如何把这些初始数据转化为具有生物学意义的信息至关重要。近年来,各类计算机视觉算法、图形图像处理和机器学习分类方法在表型数据解析中得到大规模应用,通过融合科研工作者的专家先验知识,从各种结构化和非结构化信息中自动抽取植物大小、形态、生长动态、病害检测等重要表型特

表 1 多尺度植物表型数据获取与解析案例

Table 1 Examples of data acquisition and analysis of multi-scale plant phenotyping

不同尺度分类	数据类型	表型解析方法	表型参数	植物类别
细胞、组织尺度	激光共聚焦显微图像	随机森林	下胚轴细胞表型	拟南芥 ^[24]
	Micro-CT 图像	随机森林	茎秆维管束表型	玉米 ^[25]
		人工神经网络(ANN)	幼苗病害识别	蝴蝶兰 ^[26]
		随机森林	生物量、叶面积指数、叶长等表型	拟南芥 ^[27]
器官尺度	RGB 图像	卷积神经网络(CNN)	根系、叶片、果穗等识别与分类	小麦 ^[28]
			豆荚中的种子数目	大豆 ^[29]
	深度卷积神经网络(DCNN)	果实成熟度	苹果 ^[30]	
		叶片识别(叶型胁迫的类型)、分类(低、中或高胁迫)及量化(胁迫严重程度)	大豆 ^[31]	
		RGB+深度图像	支持向量机(SVM)	叶片病毒侵染识别
	荧光成像	支持向量机(SVM)	叶片黄龙病检测	柑橘 ^[34]
	高光谱成像	支持向量机(SVM)	病害早期检测	大麦、西红柿、甜菜 ^[35]
		K-均值聚类	叶片黄褐斑病检测	大麦 ^[36]
	光谱、红外、叶绿素 荧光成像	随机森林	小麦黑斑病检测	小麦 ^[37]
			RGB+光谱图像	支持向量机(SVM)
植株尺度	点云数据	Faster R-CNN	植株分割、株高检测	玉米 ^[39]
	RGB 图像	简单线性迭代聚类(SLIC)	大田水稻稻穗识别	水稻 ^[40]
		+卷积神经网络(CNN)	开花表型性状	小麦 ^[41]
	RGB+高程图	卷积神经网络(CNN)	出苗率、生物量	小麦 ^[42]
群体尺度	立体相机成像	深度卷积神经网络(DCNN)	茎秆数目、茎宽表型	高粱、甘蔗、谷物、玉米 ^[43]
	多光谱成像	支持向量机(SVM)	冠层覆盖度、植被指数、开花表型检测	棉花 ^[44]
			黄龙病检测	柑橘 ^[45]
	点云数据	人工神经网络(ANN)	绿叶面积指数(GAI)解析	小麦 ^[46]
浅层卷积神经网络(CNN)		产量性状	莴苣 ^[47]	

征和逻辑关系,在海量大数据中实现表型性状自动精准识别。目前,利用支持向量机(SVM)^[24, 25, 34, 38, 44]、随机森林^[27, 37]、人工神经网络(ANN)^[46]、卷积神经网络(CNN)^[28, 29, 41, 47]、深度卷积神经网络(DCNN)^[31, 43]等算法实现了对不同的植物器官的自动分类和识别、表型性状的高通量解析、以及植物病害性状的自动解析等。另外,通过使用如OpenCV、SciKit-Image、TensorFlow等开源软件库和YOLO^[48]、Capsule Networks^[49]等模型框架设计开发的自动化表型分析流程不断更新,极大地推动了表型大数据的分类、解析与可视化。今后,随着高通量、多维度、多尺度表型数据的进一步积累,适用于大数据分析的软硬件系统升级、分析流程的整合与交互式可视化分析平台的建设将必不可少。目前,植物表型解析研究的重点是增加可观测、可量化、具有明确生物学含义的植物表型性状数量,提高可量化植物表型解析的精度和效率,并针对不同植物的形态结构和生理生态功能,研发相关算法开展定制的表型解析。

3.2 植物表型组大数据管理及建库技术

利用高通量植物表型信息获取平台获取的图像、点云、光谱和环境等非结构化表型数据的大量出现,为植物表型组数据的管理带来了挑战。计算机技术的快速发展,特别是互联网的普及和数据库技术的进步,为有效管理飞速增长的表型组数据提供了可能。数据库的研究始于20世纪60年代中期,历经半个世纪的发展,形成了坚实的理论基础、成熟的商业产品和广泛的应用领域。传统的植物表型数据多为结构化的数值型和字符串型数据,可利用常见的关系数据库完成有效的数据存储、检索和维护。但由于非结构化表型数据格式不一,长度各异,无法用简单的二维表结构来逻辑表达和实现,因此出现了与之匹配的非结构化植物表型组数据管理系统。目前,常用于植物表型组数据的非结构化数据管理系统有基于传统关系数据库系统扩展的非结构化数据管理系统^[50, 51],以及基于NoSQL的非结构化数据管理系统^[52]等。近年来,有关植物表型组学的数据库及管理系统已有报道,如为特定物种的植物本体以及基因和表型注释提供参考的Planteome数据库^[53],

针对物联网传感器和表型平台自动化获取数据的开源信息管理系统CropSight^[51]等,详见表2。

植物表型组数据库构建是对表型数据进行管理、存储和共享的过程,可以利用计算机硬件和软件技术增强数据管理能力;可以使用充足的数据注释和标准化的文件格式提高数据存储质量;可以打破信息孤岛实现表型组大数据的整合与共享。在植物表型组数据库构建中,常用的表型信息标准化原则包括三点^[60]:(1)利用最小信息法(minimum information, MI)来定义表型组数据集的内容,确保表型数据可验证、可分析和可解释,如使用植物表型试验的最低限度信息(Minimum Information About a Plant Phenotyping Experiment, MIAPPE)来标准化描述表型研究,有利于研究人员对表型数据进行再利用和整合^[19, 61];(2)采用本体术语(ontology terms)作为表型数据的唯一和可重复性注释,如株高、叶倾角、叶面积指数等;(3)选择适当的数据格式(data format)来构建表型数据集,如Micro-CT图像格式为BMP,RGB图像格式为PNG或JPG,点云数据获取格式为XYZ。如今,基于“云技术”的数据标准化和存储方案正在成为植物表型组数据存储的发展趋势。云存储系统可以优化植物表型平台系统架构、文件结构和高速缓存等设计,基于web的、最先进的云和存储技术可以有效地收集、分析和可视化植物表型组数据。同时,5G技术快速发展,将有助于实现任意环境下植物表型数据获取-传输-解析-存储-应用整个环节的实时性、在线化和可视化,尤其是将提升海量表型数据传输速率。此外,人工智能的先进技术,为公开可用的可扩展型植物表型组数据管理系统的构建提供支持,实现了数据传输、校准、标注和聚合等过程的有效集成,并加强表型组数据的重利用和安全共享。

3.3 植物表型组大数据挖掘与应用

植物表型组大数据的挖掘与利用目前可归纳为两个方面:面向表型组本身的表型性状预测和结合其他组学的多重组学分析。

3.3.1 表型性状预测

植物表型组学海量聚集的数据规模和愈发繁杂

表 2 主要植物表型组数据库及管理系统

Table 2 Main database and data management system of plant phenomics

名称	简介	发表时间	URL
CropSight ^[51]	针对物联网传感器和表型平台自动化获取数据的开源信息管理系统	2019	https://github.com/Crop-Phenomics-Group/cropsight/releases
PHIS ^[54]	本体驱动的植物混合信息系统——处理植物表型组学中多源、多尺度信息	2019	http://www.phis.inra.fr/
Planteome ^[53]	植物基因组和表型组数据共享平台	2018	http://www.planteome.org
Crop Phenotyping Center ^[55]	华中农业大学作物表型中心	2017	http://plantphenomics.hzau.edu.cn/checkiflogin_en.action
PGP repository ^[56]	植物表型和基因组学数据发布基础平台	2016	http://edal.ipk-gatersleben.de/repos/pgp/
Seed breeding cloud platform	金种子育种云平台	2016	http://ebreed.com.cn/#bz
Phenotyper ^[57]	使用移动终端收集表型数据的软件	2015	http://www.bioinformatics.org/groups/?group_id=1210
SensorDB ^[82]	用于集成、可视化和分析各种生物传感器数据的虚拟实验室	2015	http://sensordb.csiro.au
OPTIMAS-DW ^[50]	玉米的转录组学、代谢组学、离子组学、蛋白质组学和表型组学综合数据资源库	2012	https://apex.ipk-gatersleben.de/apex/f?p=270:1::::
iPlant (又称 CyVerse) ^[58]	植物研究数据管理系统	2011	http://www.cyverse.org
BIOGEN BASE- CASSAVA ^[59]	木薯表型组和基因组信息资源库	2011	http://www.tnagenomics.com/biogenbase/casava.php
BreedDB	收录育种所需数量农艺性状	2009	https://www.wur.nl/en/show/BreedDB.htm

的数据类型有力地促进了表型性状预测技术的发展。大数据基础上的表型性状预测方法发展历经三个阶段：(1) 基于过程机理的模型。虽然土壤数据库、主要农作物品种区域试验数据库和气象数据库为植物表型组大数据的形成奠定了良好基础，但人们无法通过观测和控制实验等方法全面获取植物表型信息全球尺度上的地域分异规律，基于机理过程的计算机模拟技术不仅能够分析全球气候变化对主要农作物产量的影响^[62]，还可以为区域尺度农作物产量和气候数据提供了一种生成技术和融合方法^[63]。(2) 基于统计学习理论的机器学习。遥感数据及其反演算法的快速发展极大丰富了植物面元尺度的表型信息（如植被指数），由于部分地区缺乏地面观测数据且植物表型信息和环境因子间存在复杂的非线性关系，机器学习为植被指数驱动下的植物表型性状的预测提供了重要手段，例如机器学习算法可大幅提升主要农作物产量的预测精度^[64]。(3) 针对大

规模数据的深度学习。高通量的环境信息和植物性状获取设备正在产生越来越多的半结构化和非结构化数据（如图像、点云和光谱），传统的数据分析和表达方法已无法满足解决复杂、抽象问题的需求，国外顶级学术刊物 Nature 曾在 2018 年刊文指出深度学习对于分析生物学研究中的复杂数据具有重要意义^[65]。2017 年斯坦福大学的可持续性和人工智能实验室以遥感影像为主要数据源，利用深度学习实现了对美国县域尺度大豆产量的准确预测^[66]，该技术目前已被美国农业大数据公司 Gro Intelligence 实现商业化应用^[67]，为金融机构、贸易公司、政府和农场主提供决策咨询服务。这些方法为及时、高效、准确地预测不同区域、不同尺度的时序植物表型信息，揭示植物表型性状的地域分异和演化规律，服务植物育种和栽培决策提供了重要手段。

3.3.2 基于表型组的多组学分析

植物生命活动是植物在基因和环境共同作用下

的动态过程,在某种程度上是表型特征与生理功能的集合。植物的生长发育是一个复杂的网络,基因变异、表观遗传的改变、基因表达水平的异常等诸多因素都会影响着生命体特征的改变。随着高通量测序技术的不断发展与完善,单组学研究日趋成熟与完善,而整合多组学数据研究植物生长发育的工作方兴未艾。近年来,将基因组数据与表型数据相结合的组学研究在许多植物中开展起来,迅速解码了大量未知基因的功能,提高了对 G-P 图谱的理解^[68]。2013 年,将小粒型谷物表型性状与基因组信息相关联,剖析籽粒物质积累的遗传结构^[69]。2014 年,水稻 13 个传统农艺性状与 2 个新定义性状相结合,利用 GWAS 鉴定出 141 个相关位点^[70]。2015 年,采用高通量叶片表型获取方法 (HLS) 解析 3 个关键生育期的 29 项叶片表型性状,并进行 GWAS 分析,检测到 73 个调控叶片大小、123 个调控叶片颜色和 177 个调控叶片形状的新位点^[71]。2017 年,结合玉米从苗期至抽穗期的 106 个农艺性状,进行大规模 QTL 作图,共鉴定出 988 个 QTL 位点^[55]。显然,将高通量表型技术与大规模 QTL 或 GWAS 分析相结合,不仅极大地拓展了我们对植物动态发展过程的认识,而且为植物基因组学、基因表征和育种研究提供了一种新的工具。随着表型组数据的积累与完善,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,植物基因组-表型组智能设计将孕育而生。应用人工智能模拟的方法,人工设计聚合优势基因,建立具有“理想基因型”^[72]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”,最终构建基因型-表型-环境多维大数据驱动的精准育种决策系统,实现颠覆性高效育种新模式。

4 面临的问题及挑战

在植物科学需求牵引和国际各研究组织的共同努力下,植物表型组大数据研究进展迅速,然而,植物表型组学目前仍处于发展的初级阶段,在构建基因型-表型-环境的颠覆式育种新模式过程中,

仍面临诸多问题和挑战亟待解决。下面将从植物表型组学大数据的产生、整合、分析和应用各环节探讨其面临的问题及挑战。

4.1 植物表型数据采集缺乏共识统一的标准

基因组学已建立起完善的数据获取与解析标准,在其约束下,研究者可以有序开展大量物种的基因测序以及数据库构建工作。由于高质量植物表型数据采集环境搭建复杂、采集工作量大,世界各国正耗费大量人力物力,重复采集植物表型数据。目前植物表型研究者主要致力于植物表型获取平台的搭建和解析方法的研究,虽然形成了诸多植物表型采集解决方案,但由于缺乏统一的标注、命名、格式、完整性约束等数据采集标准与规范,所得到的植物表型数据存在诸多问题,如格式和精度差异性大、配套信息不完整、数据冗余、数据利用率低等,亟待加强表型数据采集标准体系的建设。

4.2 多样化表型配套设施和低成本表型设备亟待发展

植物表型数据采集基础设施是解决高通量植物表型数据获取最有效的途径,在植物表型数据采集标准的规范下,通过建设植物表型基础设施可以实现标准化、高通量、高精度、全生育期的植物表型信息采集。然而,已有表型基础设施如田间轨道式^[11]或室内传送带式^[73]表型平台,多具有造价昂贵、运行和维护成本高、使用区域扩展性低等问题,无法满足大部分科研单位的植物表型研究需求。因此,亟待研发便携式、低成本、高精度的植物表型采集设备,尤其是面向田间的植物表型基础设施,解决植物栽培和育种等对植物表型获取的实际需求。此外,现有表型数据采集设备的传输主要采用传统的硬盘读写的方式,存在速度慢、效率低、难以实时查看与解析等问题,迫切需要整合在线传输与云计算技术,实现表型数据的高速传输和实时解析。例如,应用远距离无线电 (Long Range Radio, LoRa) 和空白电视频谱 (TV White Space, TVWS) 实现低功耗和远距离的统一,达到信号覆盖范围更广的要求,推进表型配套设施的云操控与云管理。

4.3 开放共享的植物表型组大数据平台建设需要高度重视

目前所构建的植物表型数据库主要是在团队已有的表型获取与解析技术方法基础上,针对特定植物构建的小型数据库,这些数据库规模小、数据完整性不高、数据挖掘潜力有限。考虑到植物表型组大数据高维度、高度复杂性和高度不确定性的特点,构建开放共享的植物表型组大数据平台,仍存在诸多技术问题,例如现有植物表型组数据管理解决方案通常使用非结构化文件系统或关系数据库,而针对非结构化植物表型组的数据管理系统虽然可以解决多模态、多尺度、多维度植物表型组数据的管理问题,但一般不提供对用户查询的即时响应,也不支持快速的数据分析和可视化,因此减缓了表型数据广泛共享和重用的速度,阻碍了团队之间的协作。近年来,植物表型组学数据中生物时间序列传感器数据的生成越来越多,当植物表型获取实验项目变得更大时,这类数据的数据量和复杂性会增加,跟踪位置、状态和起源的任务变得困难和繁重,导致数据丢失、数据恢复不完整和数据重用不良的可能性增加。因此,目前迫切需要建立一个类似于NCBI的综合性开放共享植物表型组数据库,并解决植物表型组大数据平台构建过程中的技术难点,给植物学家和育种学家提供一个表型组信息存储、处理和共享的系统,进一步推动各国研究机构和科研实验室的相互合作。

4.4 表型大数据融合与挖掘理论方法亟待创新

多尺度、多模态、多生境的表型大数据不整合就很难发挥出大数据的巨大价值。由于传感器分辨率限制问题,目前植物表型的获取与解析主要在细胞、组织、器官、单株和群体尺度分别开展,不同尺度之间的表型信息彼此孤立。多尺度表型数据的整合,多尺度、多模态、多生境的表型大数据融合,以及表型-基因型-环境信息的深度融合,涉及到数据格式、数据矛盾等一系列问题,彼此之间能否有效地融合是表型大数据面临的一个重要问题。因此,亟待突破植物表型大数据深度融合技术,从时

间和空间等多角度建立多尺度、多模态、多生境表型大数据的纽带,进而实现植物的生理生态和结构功能等表型的有效整合与解析。在基因组学技术快速发展的基础上,组学分析技术已趋近成熟,在得到植物表型组数据后,采用常规组学分析技术即可实现基于表型组的多重组学分析。但在植物表型组大数据本身的数据挖掘方面,虽然国内外研究者已利用机器学习、人工智能等前沿技术开展了大量的探索,但目前为止仍为有限的、非系统的尝试,仍缺乏面向植物表型组大数据本身的大数据挖掘分析方法,亟待形成基于植物表型组大数据的表型性状识别、分类、定量化和预测技术体系。加之植物表型组大数据具有高维度、高复杂度等特点,多种形式的表型输入数据使得搜索知识的代价较高。当前的数据挖掘工具能够处理数值型的结构化数据,但对文本等非结构化的数据处理能力有限。此外,植物表型组大数据挖掘所面临的问题中,在知识的表达以及解释机制、知识的维护更新方面也尤为突出,在技术支持的局限和其他系统的集成方面仍面临很大挑战。

4.5 缺乏植物表型组学协同共享、互作激励的机制

植物表型组学是一个新兴的研究领域,需要整合农学、生物信息学、计算机图形图像处理、生物传感器、自动化等诸多学科的专业技术力量。目前,围绕植物表型组学已建立了国际植物表型网络(IPPN)、泛欧洲植物表型联盟(EMPHASIS)、英国植物表型网络(PhenomUK)等一批学术组织,我国也建立了中国植物表型网络(CPPN),以推进我国植物表型组学研究的进展。虽然这些表型组学组织定期举办学术研讨会,促进同行学者的交流。然而与植物基因组、农业模型、系统生物学等领域的学术机构相比,目前国际国内植物表型组学协作机制并不完善;受各国、各研究单位科研和保密机制的限制,以及表型大数据获取高昂资金和人力成本的影响,各研究机构间仍存在数据不共享、信息不连通的问题,亟需构建创造性的开放共享、协同创新、激励约束的表型组学协作机制。

5 建议

为了更好地推进植物表型组学大数据的研究与应用,我国植物表型组大数据研究必须坚持创新驱动发展,加快推进表型组大数据技术的自主创新和成果转化;必须适应农业现代化发展要求,以农业产业发展需求为导向,加快构建表型组大数据技术体系、应用体系和服务体系,变革农业传统生产方式,切实推进我国作物育种和栽培的数字化进程;必须立足我国多元化的农业资源禀赋,加快不同生态区多重组学大数据的融合,揭示植物遗传和变异规律,全面提升我国农业的竞争力和自主创新能力。

5.1 加强植物表型组技术体系设计与标准研究

植物表型组学是一个多学科交叉协作的系统工程,由于专业背景的不同,研究人员对植物表型组学的认知存在差异,各学科人员多从本专业角度出发理解并开展相关工作,因此,需要加强植物表型组技术体系设计,使得从业人员对植物表型组学的认知达成共识,明确各专业人员对植物表型组学中的定位,促进基于农业信息学的植物表型组学学科的发展。

植物表型组大数据亟待加强标准研究,提升植物表型组大数据的结构化程度,降低数据噪声、消除数据冗余、降低数据获取与解析成本、减少数据存储与传输成本、提高表型组数据安全性。标准的建立有益于保障大规模数据的采集、筛选和归纳分析,降低海量的数据产生向异性信息的可能性,避免得到与事实完全相反的结论,同时降低数据挖掘和深度学习的难度。由于不同植物在不同尺度下的特征差异显著,因此,建议标准制定需要以植物物种为单位,在细胞-组织-器官-个体-群体等多尺度下细化。随着需求的增多以及研究和应用的不断深入,植物表型数据标准不会是一成不变的,亦应是一个不断迭代修正的过程。

5.2 加强植物表型-环境感知机理研究和智能化技术设备研发

目前组学研究中很多关键植物表型性状无法直

接测量,或测量精度不高,尤其是田间原位表型性状的采集,仅从测量数据的角度结合表型解析方法难以满足植物表型组学研究的需求。因此,建议加强植物表型-环境感知机理研究,通过明确植物表型和环境的互作机制,引入新的植物表型测量技术(如太赫兹、电磁探测和生物小分子检测技术等),提升表型性状获取精度,增加可测表型性状数量。

高通量表型设备是获取高通量植物表型数据最为高效的技术手段,表型设备的智能化程度上决定了植物表型组学大数据发展进程的快慢,因此,建议加强植物表型智能化设备研发,提高植物表型数据获取的效率和精度,在植物表型组学研究过程中早日实现机器代替人力的目标。

5.3 加强植物表型组大数据建设

从表型到表型组,再到表型组大数据,不仅是一个数据积累和堆积的过程,还需要加强植物表型组大数据建设,构建良性的表型组大数据的管理、共享、挖掘和利用机制。植物表型组数据获取成本较高,大数据的采集和建设要坚持应用需求导向,不要求大求全,植物表型组大数据分析处理也要关注可适性问题。建立针对植物表型组大数据的数据挖掘技术体系,加强表型性状预测、植物表型设计等应用与商业化育种等相关产业的结合。推进挖掘语言的标准化,利用标准的数据挖掘语言以及各环节的标准化工作积极促进并形成系统化的植物表型组大数据挖掘。发展可视化数据挖掘技术,结合5G技术实现植物表型组大数据云端实时可视化挖掘结果的精彩呈现。

5.4 加强人才队伍和协作网络建设

植物表型组学是一个协同、多学科交叉合作的新兴领域,需要制定具有突破性和吸引力的政策,吸引高端人才开展合作交叉研究,打破单一学科对表型组认知和应用的壁垒。然而,优秀研究人才在业界竞争下逐步流向其他领域是国内外的普遍现象,特别是交叉学科和计算机、数学专业的人才。当前,我国农业科研和生产对植物表型技术手段存在着大量需求,在此契机下,建议加快具有自主知识产权

的植物高通量表型获取平台和个性化定制的智能表型解析软件研发,降低我国植物表型技术进口率,推动我国植物表型成果转化和产业化进程,直接提升从业人员社会认可度和待遇,逐步形成具有突破性和吸引力的政策,促进植物表型人才队伍建设良性发展。此外,建议加强植物表型组学协同共享、互作激励的机制建设,推进数据和算法的共享,进而推动植物表型组学的快速发展。

参考文献

- [1] 李国杰. 大数据研究的科学价值[J]. 中国计算机学会通讯, 2012, 8(9): 8-15.
Li G J. The scientific value of big data research [J]. Communications of the China Computer Federation, 2012, 8 (9): 8-15.
- [2] 邱晨辉. 院士专家热议“人工智能计算”——人工智能时代,谁将成为“第一生产力”[N]. 中国青年报, 2018.
Qiu C H. Academician experts hotly discuss "artificial intelligence computing" - who will become the "first productivity" in the era of artificial intelligence [N]. China Youth Daily, 2018.
- [3] 中国人工智能学会. 中国人工智能系列白皮书 -- 智能农业[R]. 2016.9.
Chinese Association for Artificial Intelligence. White paper on artificial intelligence in China--Intelligent Agriculture[R]. 2016.9.
- [4] Wallace JG, Rodgers-Melnick E and Buckler ES. 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: 421-444.
- [5] Marx V. The big challenges of big data[J]. Nature, 2013, 498: 255-260.
- [6] illumina. HiSeq X™ series of sequencing systems[A]. Specification Sheet: Sequencing, 2016.
- [7] Furbank RT and Tester M. Phenomics--technologies to relieve the phenotyping bottleneck[J]. Trends in Plant Science, 2011, 16(12): 635-644.
- [8] Zhao C, Zhang Y, Du J, et al. Crop phenomics: current status and perspectives[J]. Frontiers in Plant Science, 2019, 10: 714.
- [9] Tardieu F, Cabrera-Bosquet L, Pridmore T, et al. Plant phenomics, from sensors to knowledge [J]. Current Biology, 2017, 27(15): R770-R783.
- [10] Hickey LT, A NH, Robinson H, et al. Breeding crops to feed 10 billion[J]. Nature Biotechnology, 2019, 37(7): 744-754.
- [11] McCouch S, Baute GJ, Bradeen J, et al. Agriculture: feeding the future[J]. Nature, 2013, 499(7456): 23-24.
- [12] 潘映红. 论植物表型组和植物表型组学的概念与范畴[J]. 作物学报, 2015, 41(2): 175-186.
Pan Y H. Analysis of concepts and categories of plant phenome and phenomics [J]. Acta Agronomica Sinica, 2015, 41 (2): 175-186.
- [13] 周济, Tardieu F, Pridmore T, 等. 植物表型组学:发展、现状与挑战[J]. 南京农业大学学报, 2018, 41(4): 580-588.
Zhou J, Tardieu F, Pridmore T, et al. Plant phenomics: history, present status and challenges[J]. Journal of Nanjing Agricultural University, 2018, 41(4): 580-588.
- [14] Pieruschka R and Schurr U. Plant phenotyping: past, present, and future[J]. Plant Phenomics, 2019: 1-6.
- [15] Virlet N, Sabermanesh K, Sadeghi-Tehran P, et al. Field Sc-analyzer: an automated robotic field phenotyping platform for detailed crop monitoring [J]. Functional Plant Biology, 2017, 44(1): 143.
- [16] 宁康, 陈挺. 生物医学大数据的现状与展望 [J]. 科学通报, 2015, 60(5-6): 534-546.
Ning K, Chen T. Big data for biomedical research: current status and prospective[J]. Chinese Science Bulletin, 2015, 60 (5-6): 534-546.
- [17] Shao M, Zhang Y, Du J, et al. Fast analysis of maize kernel plumpness characteristics through Micro-CT technology[C]. International Conference on Computer and Computing Technologies in Agriculture, 2019: 31-39.
- [18] Hawkesford MJ and Lorence A. Plant phenotyping: increasing throughput and precision at multiple scales[J]. Functional Plant Biology, 2017, 44(1): v-vii.
- [19] Cwiek-Kupczynska H, Altmann T, Arend D, et al. Measures for interoperability of phenotypic data: minimum information requirements and formatting [J]. Plant Methods, 2016, 12: 44.

- [20] Araus JL and Cairns JE. Field high-throughput phenotyping: the new crop breeding frontier [J]. *Trends in Plant Science*, 2014, 19(1): 52-61.
- [21] Chenu K, Deihimfard R and Chapman SC. Large-scale characterization of drought pattern: a continent-wide modelling approach applied to the Australian wheatbelt--spatial and temporal trends[J]. *New Phytologist*, 2013, 198(3): 801-820.
- [22] Reuzeau C, Frankard V, Hatzfeld Y, et al. Traitmill?: a functional genomics platform for the phenotypic analysis of cereals[J]. *Plant Genetic Resources*, 2006, 4(1): 20-24.
- [23] Li L, Zhang Q and Huang D. A review of imaging techniques for plant phenotyping [J]. *Sensors (Basel)*, 2014, 14(11): 20078-20111.
- [24] Hall HC, Fakhrazadeh A, Luengo Hendriks CL, et al. Precision automation of cell type classification and sub-cellular fluorescence quantification from laser scanning confocal images[J]. *Frontiers in Plant Science*, 2016, 7: 119.
- [25] Du J, Zhang Y, Guo X, et al. Micron-scale phenotyping quantification and three-dimensional microstructure reconstruction of vascular bundles within maize stalks based on micro-CT scanning [J]. *Functional Plant Biology*, 2017, 44(1): 10.
- [26] Huang K-Y. Application of artificial neural network for detecting Phalaenopsis seedling diseases using color and texture features [J]. *Computers and Electronics in Agriculture*, 2007, 57(1): 3-11.
- [27] Lee U, Chang S, Putra GA, et al. An automated, high-throughput plant phenotyping system using machine learning-based plant segmentation and image analysis [J]. *PLoS One*, 2018, 13(4): e0196615.
- [28] Pound MP, Atkinson JA, Townsend AJ, et al. Deep machine learning provides state-of-the-art performance in image-based plant phenotyping [J]. *Gigascience*, 2017, 6(10): 1-10.
- [29] Uzal LC, Grinblat GL, Namí as R, et al. Seed-per-pod estimation for plant breeding using deep learning[J]. *Computers and Electronics in Agriculture*, 2018, 150: 196-204.
- [30] Sabzi S, Abbaspour-Gilandeh Y, Garcí a-Mateos G, et al. An automatic non-destructive method for the classification of the ripeness stage of red delicious apples in orchards using aerial video[J]. *Agronomy*, 2019, 9(2): 84.
- [31] Ghosal S, Blystone D, Singh AK, et al. An explainable deep machine vision framework for plant stress phenotyping [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2018, 115(18): 4613-4618.
- [32] Mokhtar U, Ali MAS, Hassanien AE, et al. Identifying two of tomatoes leaf viruses using support vector machine [M]. New Delhi: Springer, 2015: 771-782.
- [33] Raza SE, Prince G, Clarkson JP, et al. Automatic detection of diseased tomato plants using thermal and stereo visible light images[J]. *PLoS One*, 2015, 10(4): e0123262.
- [34] Wetterich CB, Kumar R, Sankaran S, et al. A comparative study on application of computer vision and fluorescence imaging spectroscopy for detection of citrus huanglongbing disease in USA and Brazil [J]. *Frontiers in Optics*, 2013: 841738.
- [35] Rumpf T, Mahlein AK, Steiner U, et al. Early detection and classification of plant diseases with Support Vector Machines based on hyperspectral reflectance[J]. *Computers and Electronics in Agriculture*, 2010, 74(1): 91-99.
- [36] Wahabzada M, Mahlein AK, Bauckhage C, et al. Metro maps of plant disease dynamics--automated mining of differences using hyperspectral images[J]. *PLoS One*, 2015, 10(1): e0116902.
- [37] Odilbekov F, Armoniemi R, Henriksson T, et al. Proximal phenotyping and machine learning methods to identify septoria tritici blotch disease symptoms in wheat[J]. *Frontiers in Plant Science*, 2018, 9: 685.
- [38] Chen D, Neumann K, Friedel S, et al. Dissecting the phenotypic components of crop plant growth and drought responses based on high-throughput image analysis [J]. *Plant Cell*, 2014, 26(12): 4636-4655.
- [39] Jin S, Su Y, Gao S, et al. Deep Learning: individual maize segmentation from terrestrial lidar data using faster R-CNN and regional growth algorithms [J]. *Frontiers in Plant Science*, 2018, 9: 866.
- [40] Xiong X, Duan L, Liu L, et al. Panicle-SEG: a robust image segmentation method for rice panicles in the field based on deep learning and superpixel optimization[J]. *Plant Methods*, 2017, 13: 104.
- [41] Wang X, Xuan H, Evers B, et al. High-throughput phenotyping with deep learning gives insight into the genetic architec-

- ture of flowering time in wheat[J]. bioRxiv, 2019.
- [42] Shubhra Aich, Anique Josuttis, Ilya Ovsyannikov, et al. DeepWheat: estimating phenotypic traits from crop images with deep learning[J]. IEEE Winter Conference on Applications of Computer Vision, 2018: 323-332.
- [43] Harjatin Baweja, Tanvir Parhar, Omeed Mirbod, et al. StalkNet: a deep learning pipeline for high-throughput measurement of plant stalk count and stalk width. Field and Service Robotics. Cham: Springer, 2018.
- [44] Xu R, Li C and Paterson AH. Multispectral imaging and unmanned aerial systems for cotton plant phenotyping[J]. PLoS One, 2019, 14(2): e0205083.
- [45] Garcia-Ruiz F, Sankaran S, Maja JM, et al. Comparison of two aerial imaging platforms for identification of Huanglongbing-infected citrus trees[J]. Computers and Electronics in Agriculture, 2013, 91: 106-115.
- [46] Liu S, Baret F, Abichou M, et al. Estimating wheat green area index from ground-based LiDAR measurement using a 3D canopy structure model[J]. Agricultural and Forest Meteorology, 2017, 247: 12-20.
- [47] Bauer A, Bostrom AG, Ball J, et al. Combining computer vision and deep learning to enable ultra-scale aerial phenotyping and precision agriculture: A case study of lettuce production[J]. Horticulture Research, 2019, 6: 70.
- [48] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: unified, real-time object detection[J]. Computer Science, 2015.
- [49] Sabour S, Frosst N and Hinton GE. Dynamic routing between capsules[J]. Computer Science, 2017.
- [50] Colmsee C, Mascher M, Czauderna T, et al. OPTIMAS-DW: a comprehensive transcriptomics, metabolomics, ionomics, proteomics and phenomics data resource for maize[J]. BMC Plant Biology, 2012, 12: 245.
- [51] Reynolds D, Ball J, Bauer A, et al. CropSight: a scalable and open-source information management system for distributed plant phenotyping and IoT-based crop management[J]. Giga-Science, 2019, 8(3).
- [52] Salehi A, Jimenez-Berni J, Deery DM, et al. SensorDB: a virtual laboratory for the integration, visualization and analysis of varied biological sensor data [J]. Plant Methods, 2015, 11: 53.
- [53] Cooper L, Meier A, Laporte MA, et al. The Planteome database: an integrated resource for reference ontologies, plant genomics and phenomics[J]. Nucleic Acids Research, 2018, 46(D1): D1168-D1180.
- [54] Neveu P, Tireau A, Hilgert N, et al. Dealing with multi-source and multi-scale information in plant phenomics: the ontology-driven Phenotyping Hybrid Information System[J]. New Phytologist, 2019, 221(1): 588-601.
- [55] Zhang X, Huang C, Wu D, et al. High-throughput phenotyping and QTL mapping reveals the genetic architecture of maize plant growth [J]. Plant Physiology, 2017, 173 (3): 1554-1564.
- [56] Arend D, Junker A, Scholz U, et al. PGP repository: a plant phenomics and genomics data publication infrastructure[J]. Database (Oxford), 2016.
- [57] Kohl K and Gremmels J. A software tool for the input and management of phenotypic data using personal digital assistants and other mobile devices [J]. Plant Methods, 2015, 11: 25.
- [58] Goff SA, Vaughn M, McKay S, et al. The iPlant collaborative: cyberinfrastructure for plant biology [J]. Frontiers in Plant Science, 2011, 2: 34.
- [59] Jayakodi M, Selvan SG, Natesan S, et al. A web accessible resource for investigating cassava phenomics and genomics information: BIOGEN BASE [J]. Bioinformatics, 2011, 6 (10): 391-392.
- [60] Krajewski P, Chen D, Cwiek H, et al. Towards recommendations for metadata and data handling in plant phenotyping [J]. Journal of Experimental Botany, 2015, 66 (18): 5417-5427.
- [61] Oellrich A, Walls RL, Cannon EK, et al. An ontology approach to comparative phenomics in plants [J]. Plant Methods, 2015, 11: 10.
- [62] Lobell DB, Schlenker W and Costa-Roberts J. Climate trends and global crop production since 1980 [J]. Science, 2011, 333(6042): 616-620.
- [63] Lobell DB and Burke MB. On the use of statistical models to predict crop yield responses to climate change[J]. Agricultural and Forest Meteorology, 2010, 150(11): 1443-1452.
- [64] Johnson MD, Hsieh WW, Cannon AJ, et al. Crop yield forecasting on the Canadian Prairies by remotely sensed vegeta-

- tion indices and machine learning methods [J]. *Agricultural and Forest Meteorology*, 2016, 218 - 219: 74-84.
- [65] Webb S. Deep learning for biology [J]. *Nature*, 2018, 554 (7693): 555-557.
- [66] You J, Li X, Low M, et al. Deep gaussian process for crop yield prediction based on remote sensing data[M], 2017.
- [67] Gro Intelligence. How to make money with Gro intelligence yield models[EB/OL]. <https://www.gro-intelligence.com/about/blog/how-to-make-money-with-gro-intelligence-yield-models>.
- [68] Huang X, Wei X, Sang T, et al. Genome-wide association studies of 14 agronomic traits in rice landraces [J]. *Nature Genetics*, 2010, 42(11): 961-967.
- [69] Busemeyer L, Mentrup D, Moller K, et al. BreedVision—a multi-sensor platform for non-destructive field-based phenotyping in plant breeding[J]. *Sensors (Basel)*, 2013, 13(3): 2830-2847.
- [70] Yang W, Guo Z, Huang C, et al. Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice[J]. *Nature Communications*, 2014, 5: 5087.
- [71] Yang W, Guo Z, Huang C, et al. Genome-wide association study of rice (*Oryza sativa* L.) leaf traits with a high-throughput leaf scorer[J]. *Journal of Experimental Botany*, 2015, 66 (18): 5605-5615.
- [72] Donald CM. The breeding of crop ideotypes [J]. *Euphytica*, 1968, 17: 385-403.
- [73] Junker A, Muraya MM, Weigelt-Fischer K, et al. Optimizing experimental procedures for quantitative evaluation of crop plant performance in high throughput phenotyping systems [J]. *Frontiers in Plant Science*, 2014, 5: 770.