
机器学习技术加速植物精准设计育种

作者：writer 来源：爱科学

本文原地址：<https://www.iikx.com/news/progress/20159.html>

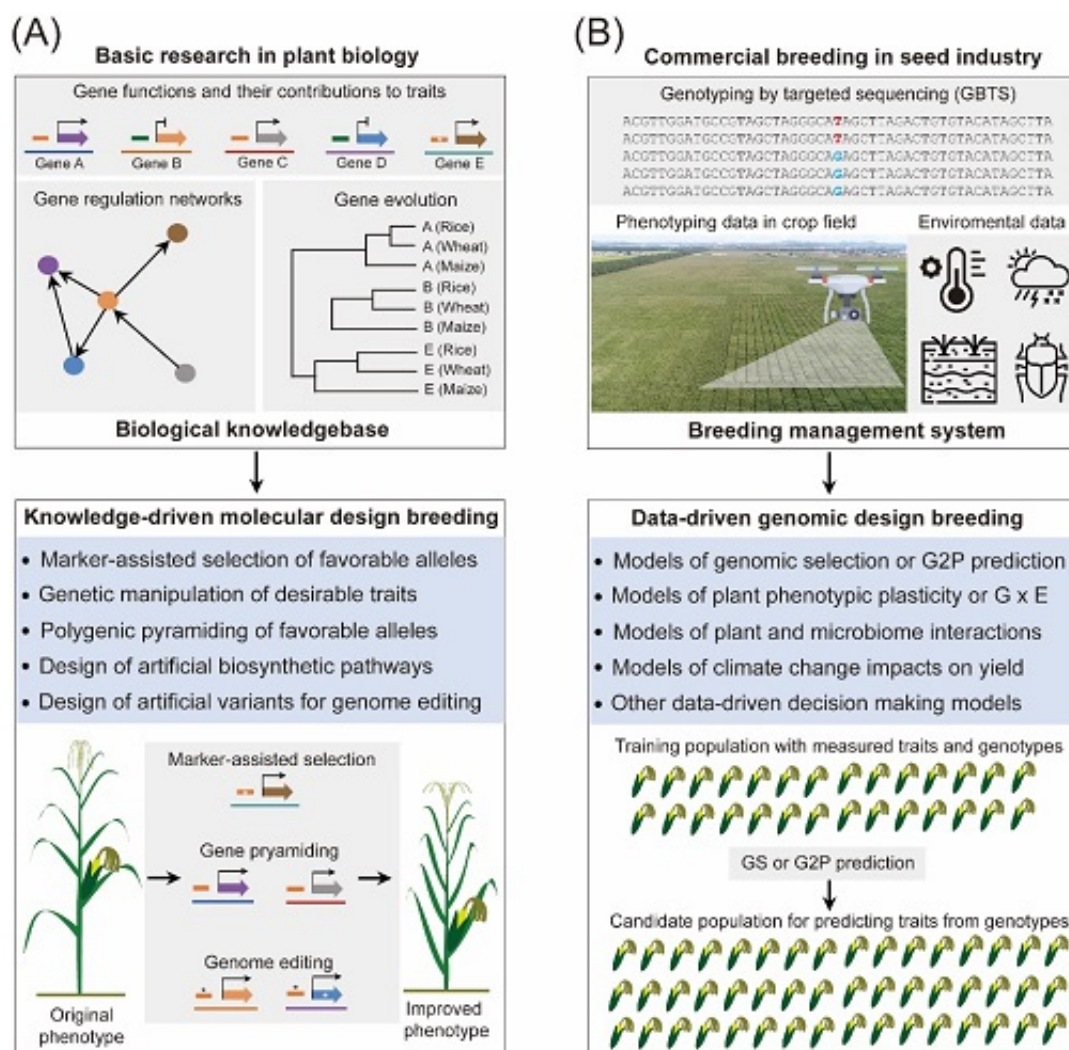
本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

机器学习技术加速植物精准设计育种。

种子被誉为农业的芯片，育种科技创新是推动农业发展的核心动力。未来植物育种的新范式是基因组学、基因编辑、合成生物学等生物技术（BT）与数据科学、机器学习、人工智能等信息技术（IT）的多元化融合。农业农村部十四五规划将智慧种业列在智慧农业领域七大攻关任务之首。任务中明确提出：构建数字化育种平台，探索基因型到表型的智能育种技术体系，加快经验育种向精确育种转变的攻关目标。

9月21日，Cell旗下的植物科学领域顶级综述类期刊Trends in Plant Science在线发表了中国农业大学分子设计育种前沿科学中心、国家玉米改良中心王向峰教授与闫军副教授共同撰写的特邀综述：Machine learning bridges omics sciences and plant breeding。该综述准确定义了精准育种的含义，并将精准设计育种划分为知识驱动分子设计育种与数据驱动的基因组设计育种。论文重点阐述了机器学习技术如何将知识与数据转化成为育种服务的驱动力，以及如何为基础研究与育种实践之间建立桥梁，加速实现植物领域的精准设计育种。

最近几十年，植物生物学基础研究产生了大量新的知识和数据，这些知识与数据最终将为植物育种与性状改良而服务。然而，实现植物精准设计育种的终极目标，还需要解决当前植物基础研究与育种实践脱节的问题。作为人工智能的一个分支，机器学习技术因其在整合复杂多变的生物学知识和组学大数据方面的卓越能力而得到广泛应用。

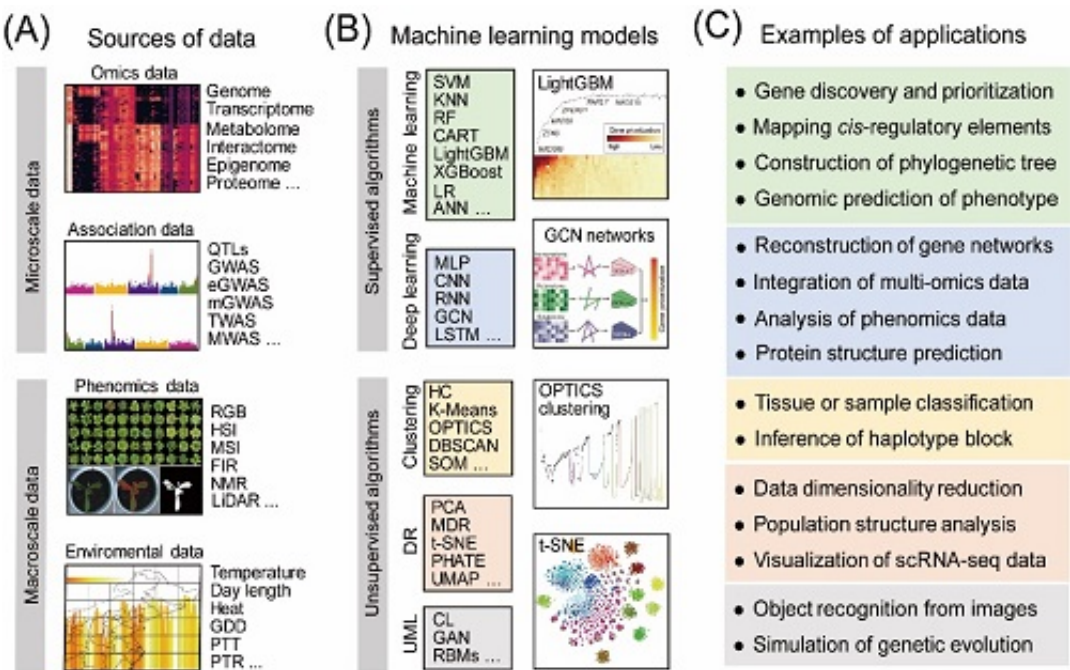


知识与数据驱动精准设计育种 王向峰供图

机器学习主要可以通过两种途径在基础研究和育种实践中建立桥梁。一种途径是从植物生物学的基础研究中认识基因功能和调控机制，从而实现知识驱动的分子设计育种。在明确性状调控基因的功能后，通过分子标记辅助选择、有利等位基因的多基因聚合、基因编辑与合成生物学等技术，对植物品种进行定向改良。另一种途径是直接将机器学习技术应用于商业育种管线，构建各种预测模型和决策算法，从而实现数据驱动的基因组设计育种。

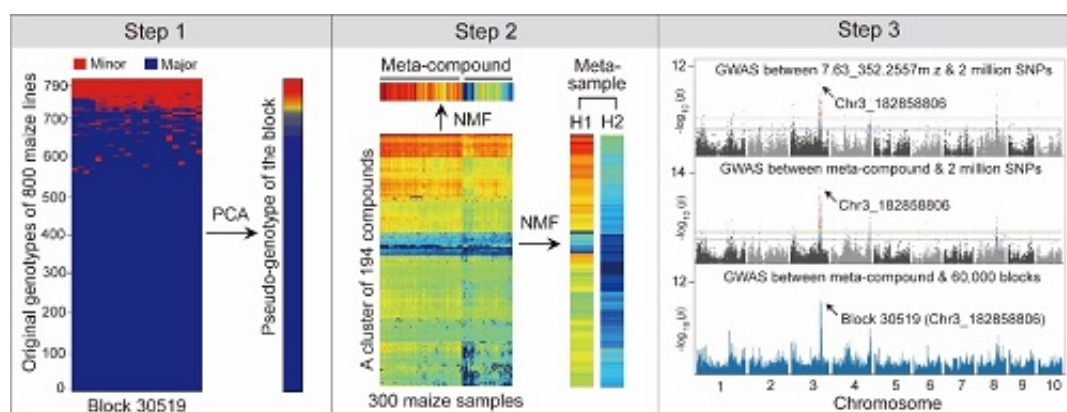
这两种途径在现代商业化育种管线中相互结合并发挥着重要作用。现代育种管线的选择取决于与性状相关的基因或位点的数量：对于主要由遗传背景决定的数量性状，如：产量、生物量、环境适应性等性状，通常采用数据驱动模型来推断表型与全基因组标记之间的相关性；对于由遗传背景决定的多基因性状，如抗病、品质等性状，必须首先明确性状调控基因的分子功能及作用途径，才能将多个基因的优良等位变异精准聚合；对于单基因性状而言，应用基因编辑是人工制造突

变，则是性状改良的最佳快捷方式。事实上，只要在植物生物学和育种方面积累了足够的知识与数据，机器学习技术便可以发挥功效以推动植物精确设计育种目标的实现。



各类机器学习算法在植物多组学研究中的应用 王向峰供图

论文首先介绍了现代机器学习技术的主要类型（包括监督式学习、半监督式学习、非监督学习、深度学习等）与最新进展；其次，综述了如何将现代机器学习算法应用于高维多组学数据降维、基因调控网络推断、多组学数据关联分析与基因挖掘，以及候选基因的优先级决策等植物学基础研究中；再次，介绍了基于半监督学习框架的深度学习算法在植物表型组学中的应用进展；最后，介绍了机器学习技术在全基因组选择辅助育种、基因型到表型预测，以及基因型与环境互作建模中的应用进展。在论文的结论与展望部分，讨论了目前机器学习和人工智能技术在植物研究中面临的挑战和潜在解决方案。



案例：应用NMF降维算法提高基因挖掘的效率 王向峰供图

此外，本综述还提供了一个应用非监督学习案例，即：如何利用NMF非负矩阵分解算法提高玉米多组学数据关联分析的效率与基因挖掘的精度。（来源：中国科学报张晴丹）

相关论文信息：<https://doi.org/10.1016/j.tplants.2022.08.018>

作者：王向峰等 来源：《植物科学趋势》

更多 科学进展 请访问 <https://www.iikx.com/news/progress/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](https://www.iikx.com)转发