科学通报 2026 年 第 71 卷 第 20 期: 4710 ~ 4722
生物与信息技术赋能现代农业绿色发展专辑 评 述
《中国科学》杂志社 SCIENCE CHINA PRESS

多组学融合与人工智能驱动的大豆智能设计育种: 进展与展望

苏鑫[1,2], 朱志锋[3], 谷晓峰[4], 邱丽娟[1,3*]

1. 中国农业科学院作物科学研究所, 北京 100081
2. 神农种业实验室/河南省作物分子育种研究院, 郑州 450002
3. 先正达集团中国/中国种子集团有限公司, 三亚 572000
4. 中国农业科学院生物技术研究所, 北京 100081
* 联系人, E-mail: qiulijuan@caas.cn

2026-04-30 收稿, 2026-05-28 修回, 2026-06-17 接受, 2026-06-22 网络版发表
中央级公益性科研院所基本科研业务费专项(Y2026CG02)资助

摘要 农业生物育种正加速从经验育种转向生物技术(biotechnology, BT)、信息技术(information technology, IT)和人工智能(artificial intelligence, AI)融合的智能设计新阶段, 是我国保障粮食安全与推进农业绿色转型的重要抓手. 大豆作为关键的粮油饲兼用作物, 其单产与广适性提升长期受制于产量、品质、抗逆等性状的多基因基础与强烈的基因型-环境互作(G×E), 以及水土肥药等资源高效利用约束, 使得传统育种难以在稳产前提下实现多性状协同突破. 本文重点围绕“多组学性状解析→种质工程化创制→G×E智能预测→良种+良法+智慧装备落地”全链条主线, 系统综述大豆BT+IT+AI融合的前沿进展, 在BT方面, 梳理了泛基因组与多维组学对调控网络的解析价值, 以及基因编辑、免组培转化、合成生物学通路重构等技术突破, 和人工智能、高通量表型组、环境型鉴定、智慧农业等IT技术赋能路径, 阐述BT+IT融合在大豆复杂农艺性状解析、种质智能创制、G×E高效预测与“良种+良法+智慧装备”协同应用中的核心作用. 梳理了基于AI的G×E预测模型、全基因组智能选配(whole-genome intelligent crossing and prediction, WPICP)及数字孪生系统的技术演进, 总结了绿色防控、养分高效利用与智慧农艺装备的融合创新范式. 我国在大豆耐逆、固氮、智能工厂化育种方向已形成特色积累; 未来需要突破的路径包括生命科学大模型驱动的生成式智能设计, 合成生物学对复杂代谢网络系统重构, 以及基因型-环境-管理措施(G×E×M)协同优化, 为大豆精准育种与绿色产能提升提供理论与技术支撑.

关键词 大豆, 生物育种, BT+IT融合, 基因型-环境互作, 智能设计育种, 智慧农业

当前国际农业生物育种前沿深度融合生物技术(biotechnology, BT)与信息技术(information technology, IT), 已成为保障粮食安全、应对全球气候变化、推动农业绿色转型的核心支撑[1]. 在我国耕地资源紧张、气候变化影响加剧、大豆对外依存度居高不下的多重压力下, 提升本土大豆单产、品质、抗逆性与广适性, 发展绿色高效大豆产业, 已成为国家粮食安全领域的重大需求[2]. 系统构建大豆核心种质资源、深度解析重要农艺性状遗传调控机制, 并结合人工智能(artificial Intelligence, AI)模型开展遗传位点预测、功能基因挖掘与品种性状定向改良, 是连接植物基础遗传学研究与现代分子育种应用的前提[3]. 近年来, 我国在作物功能基因组学、分子设计育种、农业遥感、智慧农业等前沿领域实现跨越式发展, 特别是人工智能技术的引

引用格式: 苏鑫, 朱志锋, 谷晓峰, 等. 多组学融合与人工智能驱动的大豆智能设计育种: 进展与展望. 科学通报, 2026, 71: 4710–4722
Su X, Zhu Z-F, Gu X-F, et al. Multi-omics integration and AI-driven intelligent design for soybean breeding: progress and prospects (in Chinese). Chin Sci Bull, 2026, 71: 4710–4722, doi: 10.1360/CSB-2026-0608

© 2026《中国科学》杂志社

入，极大提升了海量基因组数据挖掘与高通量表型识别的效率，部分研究方向已跻身国际先进行列[4,5]. 尤其是代表性强、遗传多样性丰富的大豆核心种质的构建[6~8]，不仅促进了优异种质的精准鉴定与基因发掘，还显著提升了大豆资源的利用效率.

然而，大豆产量、品质与抗逆等重要农艺性状多为受基因型与环境(G×E)互作控制的复杂数量性状，通过传统经验育种难以实现预期定向改良. 现代生物育种突破的核心在于推动育种范式由“被动筛选”向“智能设计”转变. 一方面，泛基因组、多维组学与高通量表型技术的协同演进，与AI驱动的算法(如深度神经网络、图表示学习等)结合后，突破了传统统计学模型的算力与维度瓶颈，实现了复杂性状调控网络的系统解析，为基因组修饰提供了确切靶标；另一方面，底层递送平台的革新与CRISPR/Cas系统的迭代，使多基因协同编辑与代谢通路重构成为现实. 随着人工创制基因型空间的指数级扩张，依托人工智能构建的高精度全基因组选择预测模型，使跨环境稳定性预测已成为智能设计育种的核心枢纽；而优异种质遗传潜力的最终释放，须依托“基因型-环境-管理措施(G×E×M)”的匹配与智慧农艺决策闭环.

因此，本文紧扣“多组学性状解析→种质工程化创制→G×E智能预测→良种＋良法＋智慧装备落地”的全链条主线，系统综述大豆生物技术与信息技术(BT+IT)深度交叉的最新研究进展. 文章全面梳理了BT+IT+AI(生物技术+信息技术+人工智能)赋能农业绿色高质量发展的技术路径及其科学意义，并指出未来需突破的生命科学大模型构建、合成生物学系统重构及G×E×M协同优化方向，以期为大豆精准智能设计育种的发展提供系统的理论与技术参考(图1).

1 大豆复杂农艺性状的精准解析：多组学与大数据的交汇

大豆高产、优质、抗逆、抗病等重要农艺性状大多数是受多基因控制、易受环境互作(G×E)调控的复杂数量性状，传统遗传学方法存在解析效率低、微效多基因难挖掘等瓶颈. BT与IT的融合正为这类性状的解析从“经验关联”推向“多组学驱动的因果建模”，为大豆分子育种突破提供了技术支撑.

1.1 多组学时代的泛基因组构建与演化解析

近20年来，植物基因组学研究实现了从单一参考基因组测序，到物种水平遗传多样性系统解析的跨越式发展. 2010年大豆Williams 82参考基因组的成功破译，为大豆遗传学研究与分子育种奠定了重要基础，但单一参考基因组难以全面覆盖大豆群体中的结构变异、存在/缺失变异等遗传变异，无法完整表征大豆物种的全部遗传信息[9~11].

2014年构建并解析的植物领域首个“野生大豆泛基因组”，阐析了大豆种内/种间结构变异和基因组演变特征[12]，被*Nature*评为“近20年基因组领域的里程碑事件”，开启了真核生物泛基因组学研究的新篇章，推动植物基因组学从单一参考基因组研究，迈入群体水平的泛基因组学研究新阶段[13,14]. 近年来，随着图形化泛基因组、端粒到端粒(telomere-to-telomere, T2T)无间隙基因组组装技术的日趋成熟，作物基因组学研究正式进入多元参考基因组时代，实现了全物种遗传多样性的精准、全面挖掘[15].

在大豆研究领域，图形泛基因组技术成功鉴定出4.7万个高质量结构变异，覆盖野生大豆、地方品种与栽培大豆三大类群，挖掘出90%以上传统研究难以捕获的稀有功能基因，为大豆抗逆性提升、品质定向改良提供了丰富的新型遗传资源[16]；同时，大豆全生命周期三维时空转录组图谱的构建，搭建起宏观表型-单细胞-空间转录的三级研究体系，精准定位*GmHBs*、*GmPMTs*等调控大豆共生固氮与生长发育的关键基因靶点[17,18]，为大豆育种从传统“经验盲选”向分子精准设计转型提供了理论支撑.

通过大规模种质重测序与群体遗传学分析，系统构建了大豆核心种质群体，明确了大豆驯化过程中的遗传瓶颈、多样性演变规律及平行选择分子机制[19]；结合表观基因组学数据，进一步解析了非编码区结构变异、DNA甲基化等表观遗传修饰，在大豆重要性状调控及基因型-环境(G×E)互作中的关键调控效应，完善了大豆遗传变异-性状表型的调控网络[20,21]. 大豆多组学集成数据库的构建(如SoyOmics)为上述多层数据对齐提供了统一入口[22]，而大豆蛋白组参考资源则为“表型→蛋白证据→基因模型”提供可追溯锚点[23].

1.2 核心农艺性状基因发掘与调控网络

大豆高产、优质、抗逆等重要农艺性状多属于受多基因调控的复杂数量性状. 传统遗传学研究受限于“单组学、静态化”视角，难以解析基因型与复杂表型间的非线性关联. 随着BT与IT的深度融合，大豆研究

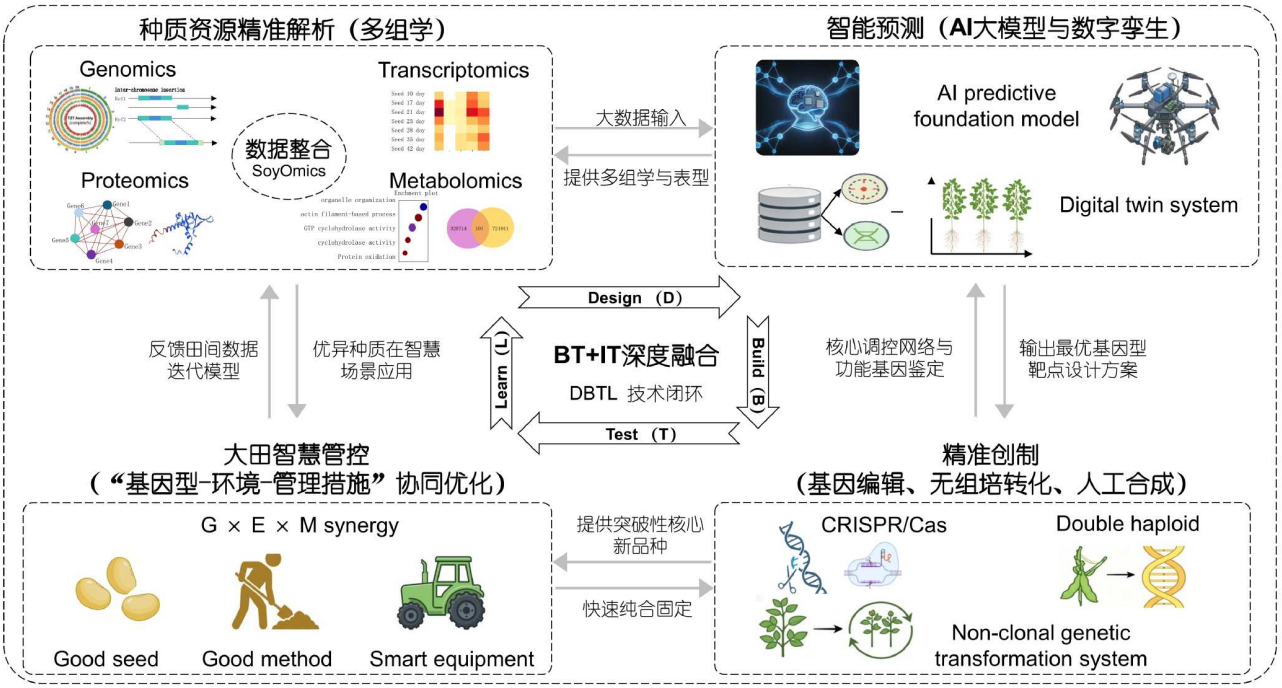


图 1 大豆BT+IT+AI智能设计育种全景图. 图中大脑元素在作者的指导下借助人工智能工具豆包和Gemini生成, 其余元素使用BioRender.com绘制

Figure 1 Panoramic view of BT+IT+AI intelligent design breeding of soybeans. Brain elements in this figure were created via artificial intelligence tools Doubao and Gemini under the authors' guidance, while other elements were created with BioRender.com

已跨入“多组学融合与AI挖掘”的新阶段, 实现了从单一基因鉴定, 升级为多维度、时空动态、网络级的系统调控研究.

在株型遗传调控方面, Dt_1、Dt_2等关键基因的克隆与功能验证, 系统阐明了大豆结荚习性、株高及顶端分生组织发育的分子机制, 为培育耐密植、抗倒伏、适宜机械化作业的大豆新品种提供了重要理论依据[24,25]. 在广适性改良方面, *E*系列、*J*、*Tof11/12*、*FT13-3*等光周期核心位点的解析, 揭示了株型发育与光温适应性的遗传互作规律, 有效破解了高产株型与广生态适应性难以协同改良的育种瓶颈[26~29].

产量与品质协同改良是大豆育种的核心目标. 依托SoyOmics等多维组学数据库与全基因组关联分析技术, 多效基因的挖掘取得显著突破. 例如, *GmSWEET10a*被证实可协同调控籽粒大小、油脂与蛋白质积累, 实现产量与品质的协同提升[30]; 整合全基因组关联分析(genome-wide association study, GWAS)和全转录组关联分析(transcriptome-wide association study, TWAS)鉴定的关键基因*GmSop20*, 通过直接激活糖转运蛋白, 调控种子碳源分配, 有效打破了高油与高蛋白间的遗传负相关, 为高油、高蛋白专用品种的定向选育提供了关键基因靶标[31]. 此外, 针对大豆脂质生物合成与油脂积累等复杂代谢途径, 人工智能算法能够基于转录组和代谢组数据构建高精度的表型预测模型, 预测非天然代谢途径的通量分配, 从而在庞大的基因互作网络中高效定位调控油分与蛋白比例的关键枢纽基因.

在养分高效利用与生物固氮领域, 一系列调控磷、氮、钾高效吸收与转运的关键基因(如*GmPTs*、*GmNRTs*、*GmAKTs*)的系统鉴定, 为培育减肥减施、养分高效的绿色大豆品种奠定了重要遗传基础[32~34]. 此外, 结合单细胞转录组(scRNA-seq)与空间转录组技术, 研究者精准解析了大豆根瘤发育与共生互作的分子调控网络, 明确了*GmNIN*、*GmRIC1*、*GmHBs*等关键基因的功能, 为提升大豆自身固氮能力、减少化肥投入提供了关键途径[35~37]. 这种基于多组学时空图谱的解析方式, 为利用AI大模型进行代谢通路的人工模拟与生成式设计提供了底层逻辑.

在抗病与逆境领域, *GmSALT3*作为耐盐碱主效基因, 通过维持胞内离子稳态显著提升大豆耐盐碱性, 为盐碱地等边际土地的大豆种植提供重要基因资源[38];

针对大豆胞囊线虫、拟茎点霉种子腐烂病等病害，“中品03-5373”近无间隙(near-gapless)基因组的构建，成功揭示了结构变异在胞囊线虫抗性中的关键作用[39]，在此基础上，*GmSNAP11*等广谱抗病位点与免疫受体基因被定位，时序表型组与多环境验证技术进一步实现了抗病基因动态效应的系统解析，为抗病分子育种提供了稳定可靠的遗传标记[40].

此外，蛋白组、代谢组、离子组学技术的纵深发展进一步拓展了复杂性状的解析维度. 在蛋白组学中，现已从2-DE时代(双向凝胶电泳时代)走向4D-DIA(四维全景定量时代)全景定量[41]，Komatsu等人[42]对亚细胞(细胞器)蛋白组与非生物胁迫响应框架(ROS清除–分子伴侣–Ca^{2+}/ABA信号交联)的综述，系统揭示了非生物胁迫响应机制. 在代谢组学中，以*Phytophthora sojae*大豆为模型，用代谢组把抗性锁定到广谱抑菌植保素(coumestrol)等苯丙烷支路产物，并由代谢物GWAS牵引候选区间[43]. 在离子组学中，其关键价值是元素积累本身可遗传、可进行GWAS分析，并直接指向转运蛋白，例如对钼酸根转运蛋白*GmMOT1.1*/*GmMOT1.2*自然变异的克隆，把“种子Mo→Mo依赖IAA合成→叶面积/产量”串成可育种闭环，其多样性种质种子元素谱GWAS则给出“离子型性状标记化”的规模示范[44,45](表1).

1.3 “空-地-室”智能感知表型组学研究新体系

破解大豆复杂数量性状遗传解析瓶颈，核心在于突破传统人工表型鉴定通量低、主观误差大且难以捕捉动态变化的局限，进而获取海量、精准、全周期的多维度表型数据[58]. 近年来，高通量传感技术与人工智能算法深度融合，成功构建起专用于大豆的“空-地-室”一体化表型监测体系. 该体系实现了表型特征的三维空间全覆盖，成为串联基因组变异与复杂表型的纽带[59~61].

一是“空”基大尺度感知. 依托搭载多光谱、高光谱及轻型激光雷达的无人机平台，实现对大范围田间群体表型的高效无损监测，可高通量获取全生育期冠层结构(如郁闭度、叶面积指数)、生物量演变及非生物胁迫响应特征. 例如，结合无人机时序影像获取的冠层覆盖度动态数据与全基因组关联分析，研究者成功鉴定出35个相关遗传位点(其中10个与已知开花期、株高QTL重合)，表明了动态表型捕获技术可以有效发掘传统静态研究遗漏的新位点，还原田间性状的时空变化规律[62,63].

二是“地”基中下层精细探测. 结合田间物联网传感器网络、地面高通量表型车及农业机器人，并深度融合机器视觉算法，该平台能够穿透冠层，实现对大豆分枝角度、结荚习性、节间长度及冠层微气候(温度、湿度、光合有效辐射等)的连续追踪，有效弥补了航空遥感难以观测冠层内部结构的缺陷. 融合自主导航、多源传感与无损检测技术，并搭载人工智能算法的田间自动化表型平台，可在大田环境中实现作物表型高通量、动态、精准采集[64].

三是“室”内地下部与微观精准量化. 针对大豆地下部及生理生化鉴定难题，依托人工气候室、X射线微

表 1 大豆复杂性状解析的多组学资源

Table 1 Multi-omics resources for complex soybean trait analysis

组学类型	代表性应用	成果
基因组学	阐析种内/种间结构变异，鉴定4.7万个高质量结构变异，挖掘90%以上的稀有功能基因	W82参考基因组发布[9]；植物首个野生大豆泛基因组[12]；图形泛基因组[16]
转录组学	捕捉大豆根瘤发育过程中的细胞过渡态，解析根瘤菌共生互作与固氮效率分子调控网络	转录组测序定量基因表达、时空转录组[35~37]
蛋白组学	4D-DIA全景定量14个器官12855种蛋白及m^6A修饰；量化亚细胞蛋白组与非生物胁迫响应框架	大豆全景蛋白组[41]；亚细胞蛋白组[42]
单细胞组学	系统解析转录分化和顺式调控演化模式；构建单细胞基因表达和顺式作用元件图谱，分析固氮相关转录因子调控网络	重复基因的表达图谱和染色质可及性图谱[46]、顺式调控元件图谱[47]、种子基因表达图谱[48]
表型组学	深度学习方法提取表型特征并预测产量；无人机的多光谱植被指数表征大豆基因型对淹水胁迫的响应；大豆表型检测、抗逆评价和管理	集合模型RF-PLSR[49]，淹水胁迫响应最佳线性模型[50]，性状优先的抗逆评价框架和闭环流程[51]
表观组学	新生的表观突变在驯化瓶颈后增加表观遗传多样性；解析南方大豆表观遗传调控的潜在分子机制	DNA甲基化图谱[52,53]，抗病基因表观调控机制[54]
代谢组学	大豆品种营养品质差异源的代谢物积累与基因表达调控差异；野生大豆种子成熟中期的细胞特性和分化过程	培育高品质大豆品种提供依据与分子靶点[55]，响应胁迫的分子机制[56,57]

型计算机断层扫描(micro-CT)、微流控芯片及高通量数字考种设备，实现了可控生境下根系三维构型、单株籽粒特征及细胞生理的自动化精准提取[65,66]. 同时，植物细胞传感器将环境胁迫信号转化为可定量检测的物理信号; 基于荧光探针与柔性微纳器件的技术, 实现了体内钙离子、活性氧等信号分子的可视化成像; H_2O_2微针可穿戴传感器则实现了病害与非生物胁迫的早期快速预警，全方位补齐了生理抗性研究中的表型数据短板[67]，这种“空-地-室”三维立体协同的监测网络，使表型组学不再是孤立的研究分支. 此外，在室内或可控生境下，高通量数字考种设备深度融合机器视觉算法，可实现对单株籽粒特征(如粒型、表观色泽、百粒重等)的全天候自动化、高精度量化提取. 这种基于AI的图像解析技术，极大地消除了人工鉴定的误差，是实现大豆高通量抗逆表型筛选的关键.

2 种质资源定向创制: 前沿生物技术与AI的深度融合

在系统解析农艺性状调控网络的基础上，实现大豆种质资源的定制化创制，已成为现代作物育种的核心目标. 合成生物学、基因编辑、遗传转化与双单倍体技术(doubled haploid, DH)并非孤立的工具，而是深入交融、不可分割的有机整体. 在AI模型的驱动下, 这些技术共同组成了现代大豆智能设计育种中的“设计-构建-测试-学习”(design-build-test-learn，DBTL)的标准化技术闭环.

2.1 AI赋能的调控元件与代谢网络重构

传统育种受限于物种内的自然遗传变异，而合成生物学则通过底层遗传元件的重构，为大豆复杂农艺性状的系统级改良提供了全新路径[68]. 在这一过程中，AI与机器学习的引入重塑了合成生物学的研究范式，成为连接“数据”与“创制”的桥梁[69]. 在靶标挖掘与通路设计阶段，深度学习算法能够处理海量多组学数据，并预测非天然代谢途径的通量分配. 基于大语言模型的生成式AI和AlphaFold等蛋白质结构预测工具，现已被成功应用于大豆光合固碳酶的理性设计及人工调控元件的生成[70,71]. AI驱动的蛋白质从头设计在生物工程诸多领域展现出强大能力，借助人工智能定向设计功能性蛋白, 并将其应用于作物, 可实现农艺性状的模块化改造与功能提升[72].

例如，针对大豆光合效率与共生固氮系统的高耗能瓶颈, 研究人员依托AI辅助设计, 正尝试重构大豆的光呼吸旁路(photorespiration bypass)、构建人工碳浓缩机制, 以期突破光合作用的生理极限[73]. 在共生固氮研究中，借助智能模型解析大豆根瘤发育的铁信号-结瘤耦合通路(如*GmNAS1*/*GmNAP1*、*GmBTSa*等)，通过计算机模拟外源代谢途径与大豆内源网络的精准匹配，定向改造结瘤信号通路，从而提升逆境下的共生固氮效率[74]. AI不仅为合成生物学提供了“施工蓝图”, 更确保了系统级代谢网络重构的精确性与可行性.

2.2 多重基因组编辑体系的迭代与叠加

为了将AI与合成生物学设计的复杂“蓝图”在体内实现，高度依赖于基因组编辑技术的迭代升级. CRISPR/Cas系统在大豆中的应用已从早期的单基因敲除跃升至多基因靶向编辑、底物单碱基编辑(CBE/ABE)及引导编辑等修饰阶段[75]. 特别是面对合成生物学设计中需要多基因协同的复杂性状叠加(trait stacking)需求，多重编辑系统展现出巨大潜力. 例如，开发我国自主知识产权工具(Cas12i3-5M)的高效多重编辑系统(pSoC12i-V3), 在大豆中成功创制出无豆腥味、高油酸的突变体[76]; 利用大豆内源启动子*GmRPS5-9*驱动优化的*LbCas12a*变体，在T_0代成功实现了多个靶点基因的高效同步编辑[77]. 此外，高效的*CRISPR*/*LbCpf1*多基因编辑系统实现了多靶标的同时编辑，完成了10 kb的染色体小片段删除，拓展了大豆多基因编辑的操作维度. 在具体性状创制上，靶向敲除*GmCKX3*可通过细胞分裂素途径实现种子细胞扩张与脂质积累，同步提升产量与含油量[78]; 多重编辑*GmNIN*/*GmRIC1*则成功实现了减氮30%、增产10%~20%的固氮效率提升[79]. 在产业化层面，我国高油酸、长童期基因编辑大豆已获首批生物安全证书, 美国*Calyxt*高油酸大豆亦实现商业化, 标志着操作技术已全面迈入产业化阶段.

2.3 突破受体基因型限制的无组培转化技术

大豆作为典型的难转化作物，长期受基因型依赖性强、组培周期长及体细胞变异等问题的制约. 为突破基因型的限制，研究人员通过理化融合方法协同促进农杆菌侵染，并结合以磷酸甘露糖异构酶(phosphomannose isomerase, PMI)为筛选基因的方法[80], 成功实现了优良育成品种(如中黄6106)的直接转化, 推动了大豆转化从模式植物向优良育种材料的突破. 尽管基因编辑与合成生物学提供了一定的创制方法，但这些遗

传改良元件与外源通路的导入依赖于高效的基因递送系统. 近年来, 通过遗传转化技术的创新, 大豆基因递送的技术瓶颈被逐步突破. 一方面, 利用发育调节因子(developmental regulators, DRs)可避开传统的外植体再生障碍. 例如, 过表达形态发生调节因子*Bbm*和*Wus2*突破了特定受体基因型限制[81]; 共表达*WUS2*和*IPT*基因, 更在不添加外源激素和筛选剂的条件下, 实现了14.6%~22.3%的较高转化效率, 并将转化周期大幅缩短至9~11周[82]. 另一方面, 在无组培遗传转化体系中, 基于RUBY可视化的大豆原位无组培遗传转化系统(RUBY-assisted one-shot tissue-culture-free in-planta soybean-transformation, ROTIS), 可利用萌发种子作为外植体进行侵染, 规避了体细胞离体再生环节, 在打破基因型屏障的同时, 将转基因植株的获取周期缩短至40~80天[83]. 此外, 纳米载体介导法及花粉管通道法等新型递送技术的日益成熟[84], 使得直接对优良大豆主栽品种进行系统级遗传修饰成为现实.

2.4 育种周期的缩短与双单倍体技术

在完成靶向修饰与高效递送后, 获得的突变材料多处于杂合状态. 在传统育种流程中, 将优异等位变异固定为纯合体通常需要历经数代的连续自交[85]. 因此, 双单倍体技术作为种质资源智能化创制的核心手段, 在加速基因纯合方面发挥着关键作用, 是缩短作物品种选育周期的重要技术[86]. 尽管*ATL*/*PLA1*、*PLD*等单倍体诱导基因在单子叶作物中功能保守, 但在双子叶作物中面临诱导效率低下的技术瓶颈[87]. 在大豆中鉴定出单倍体诱导同源基因*GmDMP1*和*GmDMP2*, 通过多重编辑创制的双突变体在其自交后代中成功筛选出单倍体植株. 虽然目前平均单倍体诱导率维持在0.56%~0.61%左右, 但研究发现有22.9%的大豆单倍体可通过自发染色体加倍产生可育种子[85,86]. 随着该技术的持续优化, 大豆纯系的创制周期有望被极限压缩至1~2年. 这一进展使得由AI理性设计、高新技术创制的新型代谢网络能够被高效固定为纯合种质, 以上技术的融合与应用完成了大豆智能设计育种技术体系的关键闭环.

3 基因型–环境互作(G×E)高效预测: 人工智能与大数据深度赋能

大豆表型(P)受基因型(G)、环境(E)及其互作(G×E)共同调控, 经典模型为: $P=\mu+G+E+(G\times E)+\varepsilon$($\mu$为群体总体均值, ε为随机误差). 针对大豆强光温敏感性与G×E互作效应, 环境型鉴定(Envirotyping)将离散环境因素转化为结构化、可建模特征向量, 为解析环境效应提供全新范式[88~90].

通过整合高时空分辨率气象、土壤、遥感大数据, 构建了大豆全生育期环境协变量网络, 实现微环境异质性量化与生态敏感度解析[91,92]. 为了将基因型数据转化为预测能力, 开发了“中豆芯”系列、SoyDNGP、中科豆芯和问豆等高通量分型平台等SNP育种芯片, 能够捕获大豆遗传多样性, 为大规模群体的全基因组扫描和G×E分析提供了强有力的工具平台[93]. 全基因组智能选配(whole-genome intelligent crossing and prediction, WPICP)模型依托深度学习实现表型预测, 可淘汰80%~90%无效杂交组合, 为大规模G×E分析提供技术支撑. 全基因组选择(genomic selection, GS)是现代育种的核心技术, 其在实现“降低田间测试负载”与“保持高预测精度”的平衡、加速育种进程方面展现出巨大潜力与显著优势[94]. 目前, 大豆GS已从“方法学演示”正式进入“育种程序集成”阶段, 对于蛋白与油分等较高遗传力的性状, GS已成为成熟的可落地工具; 而对于产量等低遗传力复杂性状, GS则发挥着高价值辅助决策的作用, 还需结合G×E等模型来进一步提升预测精度. 传统GBLUP、rrBLUP模型难以解析复杂非线性效应, 深度学习算法有效弥补这一缺陷. 卷积神经网络(convolutional neural network, CNN)挖掘SNP与环境数据的非线性关联, 长短期记忆网络(long short-term memory, LSTM)捕捉气象时序累积效应, 实现跨生态区、跨世代表型预测[95]. 融合转录组、代谢组数据, 耦合DSSAT、APSIM生长模型构建大豆数字孪生系统, 提升极端环境下预测可靠性; 结合气候预测模型, 评估基因型气候稳定性, 为广适性品种选育提供依据[79]. 基于多组学大数据的智能育种决策系统, 实现亲本组配、后代选择的智能优化, 推动大豆育种迈入数据驱动新阶段[96](表2).

4 资源高效利用新范式与智慧农业系统构建

生物技术与信息技术的深度交叉融合, 贯穿大豆种质创制、品种培育、生产管控全流程, 既是突破传统大豆育种技术瓶颈的动力, 也是构建现代化智慧农业生产体系、破解农业资源粗放利用、化肥农药过量投入等行业痛点的关键路径. 其最终目标是搭建全域环境智能感知、多源数据云端决策、全流程装备作业

表 2 大豆BT+IT+AI融合关键技术体系及其育种应用

Table 2 Key technology systems for BT+IT+AI integration in soybean and their breeding applications

技术类别	代表方法	解决的育种问题	成果/代表基因
多组学性状解析	图形化泛基因组、多维组学网络分析[15,29]	解析受多基因控制、G×E互作的复杂数量性状调控网络	*GmSop2d*[31]、*GmSWEET10a*[30]
智能表型鉴定	“空-地-室”一体化表型监测、植物细胞传感器[48]	传统鉴定通量低、误差大、难以捕获动态变化的局限	35个冠层发育动态相关位点[65]; H_2O_2微针传感器[67]
种质精准创制	多重编辑、无组培转化、AI辅助合成生物学[54,59,66]	受体基因型限制、离体再生时间长及多基因协同改良困难	无豆腥味高油酸突变体[75]; *GmNIN*/*GmRIC1*[78]; *GmDMP*[85]
预测与智慧管控	全基因组选择、深度学习、环境型鉴定、数字孪生等[73,77~79]	跨环境/生态区表型预测精度低, 良种难以释放最大潜力	全基因组智能选配, “中豆芯”系列[90]

的智慧农业闭环系统，实现大豆生产从育种到栽培、植保、管理的全程精准化管控，全面提升水土肥药等农业资源利用效率，推动大豆产业向绿色化、智能化、高效化转型，为保障国家粮食安全与农业可持续发展提供技术支撑[97,98].

生物技术与信息技术深度融合，是打通大豆育种、栽培、植保全链条，构建智能感知—数据决策—精准作业智慧农业体系的支撑，最终实现大豆生产资源高效利用与绿色低碳发展，为保大豆产能安全提供关键技术路径[99,100].

4.1 大豆绿色防控与养分高效利用

在大豆绿色植保方面，深度学习与无人机遥感技术可实现大豆疫霉根腐病、拟茎点种腐病、胞囊线虫等主要病害的早期智能识别与田间制图，结合区域流行病学模型, 实现病害预警与农药减量施用, 构建大豆全程绿色防控体系[101]. 在养分高效利用方面，依托多组学技术挖掘大豆氮、磷、钾高效吸收及根系构型优异等位变异, 定向培育养分高效大豆种质; 解析大豆根系—根际微生物互作机制, 研发专用生物菌肥, 活化土壤养分、提升利用效率，从遗传与栽培双路径实现化肥减施增效[102,103].

4.2 “良种+良法+智慧装备”大豆生产融合创新

大豆株型、抗炸荚、耐密植等性状的分子遗传改良，是支撑机械化、集约化栽培的生物学基础[104]. 结合人工智能构建出大豆基因型×管理措施(G×M)的耦合模型，可实现品种与播期、密度、水肥运筹方案的匹配, 充分释放优良品种的产量潜力[105,106].

在BT+IT+AI的融合框架下，物联网田间传感网络实时监测土壤、微气候与大豆长势，云端决策系统生成管理处方，智能农机装备实现变量施肥、喷药、高效播种等智慧作业，显著提升水肥利用效率、降低碳排放. 该体系构建了良种赋能、精准管理、智能装备协同的大豆生产新范式，实现资源高效利用与生态效益、产量提升的协同统一，推动大豆产业向绿色、智能、高效高质量转型(表3).

5 现存问题与挑战

当前，全球大豆生物育种正从单一生命科学驱动, 迈向“BT+IT+AI”三位一体的智能设计育种新阶段，传统经验育种模式正被数据驱动的决策彻底颠覆. 我国在大豆耐逆、共生固氮及智能工厂化育种领域已形成特色优势; 而美国、巴西等国则在抗除草剂、抗虫及高油酸品种产业化方面保持领先. 得益于AI与多组学的深度融合, 全球大豆育种技术代差正持续缩小. 面向国家粮食安全与农业绿色高质量发展的需求，未来大豆科研需聚焦以下三大核心方向实现突破.

5.1 实现从“经验育种”到“大模型驱动的智能设计”的育种范式跨越

打破多源数据壁垒，依托大豆多维组学数据库(如SoyOmics)，整合泛基因组、表型组与环境型数据. 利用AI模型挖掘海量组学数据中的潜在关联，逐步实现基于目标性状的定向设计与基因组重组. 同时, 推进大豆自动授粉系统与加速育种设施的工程化应用, 以提升繁育效率并缩短育种周期. 然而, 当前大豆智能育种仍面临数据标准化缺失与整合困难等瓶颈. 跨地域、跨平台的农业高通量表型数据(如无人机多光谱与田间传感数据)缺乏统一的解析标准与元数据规范; 涵盖极端环境胁迫的高质量、多模态标签数据集相对匮乏，导致深度学习模型在跨生态区预测时面临泛化能力不足与可解释性较差等挑战. 未来需重点建设国家级大豆标准测试网络及开源共享的标准化

表 3 AI在大豆育种中的主要应用场景

Table 2 Main application scenarios of AI in soybean breeding

应用场景	算法/技术手段	应用成效
复杂调控网络挖掘	机器学习、多组学大数据挖掘与网络分析	从海量数据中破译基因型与复杂表型间的非线性关联, 定位协同调控网络中的关键基因节点[31]
表型特征提取	计算机视觉、CNN、深度学习算法	将无人机时序影像转化为结构化表型特征, 识别细微动态变化, 挖掘出与已知QTL重合的新位点[62,65]
合成生物回路设计	生成式AI、AlphaFold大语言模型	处理多组学数据并预测非天然途径通量, 辅助设计光合固碳酶, 模拟大豆内源与外源通路匹配[71,73]
预测与智能选配	CNN、长短期记忆网、全基因组智能选配	捕捉SNP与气象时序的非线性及累积效应, 实现跨生态区/世代预测[91,93]
智慧农业与数字孪生	耦合DSSAT、APSIM生长模型与病害病学模型	构建大豆数字孪生系统, 实现品种与播期、密度、水肥的匹配[94]

多维组学数据库, 为大豆智能育种模型提供可靠的数据支撑.

5.2 实现从“单一基因靶点修饰”到“复杂性状调控网络”的系统级重构

结合蛋白质结构解析技术与全基因组多组学(multi-omics)图谱, 从系统生物学层面解析并重构大豆物质与能量代谢途径. 通过探索人工碳浓缩机制或优化光呼吸途径, 改善光合作用效率; 通过定向改造根际共生系统, 增强植株在逆境下的养分吸收与固氮能力. 此外, 整合遗传学、表观遗传学及多组学数据, 耦合植物叶片贴附式柔性传感器与细胞级传感技术, 实现对“基因型-环境”互作机制的微观与动态解析, 进而有效拓宽现有种质的遗传基础.

5.3 实现从“单一品种选育”到“基因型-环境-管理措施”智慧农业全要素协同优化

将品种设计前置, 深度耦合机械化、轻简化及绿色低碳生产场景, 利用AI决策模型, 量化特定大豆基因型对极端气候的适应性, 动态优化播期、种植密度与水肥运筹方案. 通过构建“良种+良法+智慧装备”的自适应匹配体系, 最大化提升农业生产力并降低系统碳足迹, 推动大豆产业向绿色、高效、智能化方向全面转型. 总体而言, 面对全球极端气候频发与提升大豆产能的需求, 大豆种业科技创新需深度融合生物技术、信息技术与人工智能. 通过加速突破智能设计育种的关键核心技术瓶颈, 将为大豆产业的产能安全与可持续发展提供坚实的科学依据与技术支撑.

需要强调的是, 随着科技进步, 原有的性状解析、种质创制、智能预测到智慧管理全链条的线性关系与因果次序正在被深度重构; 各环节不再是单向传递的接力棒, 而是形成双向耦合、多方输入输出的反馈环路. 这种改变使技术成果反过来成为科技进一步突破的条件—数据更丰→模型更准→设计更精→产出更优—从而驱动育种科技进入一个持续加速的螺旋式上升过程. 每一次循环都在提升“预测可信度”与“设计自由度”的上限, 也正是这一机制, 构成了智能育种范式的动力学来源, 促进BT+IT+AI育种的跨越式发展.

致谢 感谢江苏省现代作物生产省部共建协同创新中心对大豆复杂农艺性状多组学解析与智能育种理论探讨方面给予的学术交流平台与指导支持; 感谢吉林农业大学魏健、王婉, 神农种业实验室王俊在本文撰写及修改过程中, 对文献梳理校对以及文章内容的补充完善所提供的宝贵帮助.

参考文献

1 Wei X, Wan X Y. Planning path and scientific and technological support for the development of green and efficient agriculture (in Chinese). Chin Sci Bull, 2026, 71: 4666–4673 [魏珣, 万向元. 绿色高效农业发展的规划路径与科技支撑. 科学通报, 2026, 71: 4666–4673]

2 Ray D K, Gerber J S, MacDonald G K, et al. Climate variation explains a third of global crop yield variability. Nat Commun, 2015, 6: 5989

3 Zhu W, Li W, Zhang H, et al. Big data and artificial intelligence-aided crop breeding: progress and prospects. J Integr Plant Biol, 2025, 67: 722–739

4 Zhao C, Li J, Feng X. Development strategy of smart agriculture for 2035 in China (in Chinese). Chin J Eng Sci, 2021, 23: 1–9 [赵春江, 李瑾, 冯

献. 面向2035年智慧农业发展战略研究. 中国工程科学, 2021, 23: 1–9]
5 Gu X F, Zhang L C, Li H H, et al. Agro-biological intelligent design breeding (in Chinese). J Agric Sci Tech, 2025, 27: 1–13 [谷晓峰, 张立超, 李慧慧, 等. 农业生物智能设计育种. 中国农业科技导报, 2025, 27: 1–13]
6 Wang L, Guan Y, Guan R, et al. Establishment of Chinese soybean *Glycine max* core collections with agronomic traits and SSR markers. Euphytica, 2006, 151: 215–223
7 Song X E, Li Y H, Chang R Z, et al. Population structure and genetic diversity of the micro-core collection of Chinese cultivated soybean (*Glycine max* (L.) Merr.) (in Chinese). Sci Agric Sin, 2010, 43: 2209–2219 [宋喜娥, 李英慧, 常汝镇, 等. 中国栽培大豆(*Glycine max* (L.) Merr.)微核心种质的群体结构与遗传多样性. 中国农业科学, 2010, 43: 2209–2219]
8 Qiu L J, Li Y H, Guan R X, et al. Establishment, representative testing and research progress of soybean core collection and mini core collection (in Chinese). A S, 2009, 35: 571–579 [邱丽娟, 李英慧, 关荣霞, 等. 大豆核心种质和微核心种质的构建、验证与研究进展. 作物学报, 2009, 35: 571–579]
9 Schmutz J, Cannon S B, Schlueter J, et al. Genome sequence of the palaeopolyploid soybean. Nature, 2010, 463: 178–183
10 Qiu L J, Xing L L, Guo Y, et al. A platform for soybean molecular breeding: the utilization of core collections for food security. Plant Mol Biol, 2013, 83: 41–50
11 Sun C, Zhang Z, Liu M, et al. Comparison of grain traits and genetic diversity between Chinese and Uruguayan soybeans (*Glycine max* L.). Front Plant Sci, 2024, 15: 1435881
12 Li Y, Zhou G, Ma J, et al. *De novo* assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. Nat Biotechnol, 2014, 32: 1045–1052
13 Liu Y, Du H, Li P, et al. Pan-genome of wild and cultivated soybeans. Cell, 2020, 182: 162–176.e13
14 Torkamaneh D, Lemay M, Belzile F. The pan-genome of the cultivated soybean (PanSoy) reveals an extraordinarily conserved gene content. Plant Biotechnol J, 2021, 19: 1852–1862
15 van der Laan L, Parmley K, Saadati M, et al. Genomic and phenomic prediction for soybean seed yield, protein, and oil. Plant Genome, 2025, 18: e70002
16 Bayer P E, Golicz A A, Scheben A, et al. Plant pan-genomes are the new reference. Nat Plants, 2020, 6: 914–920
17 Pereira W J, Conde D, Perron N, et al. Investigating biological nitrogen fixation via single-cell transcriptomics. J Exp Bot, 2024, 76: 931–949
18 Jia K H, Zhang X, Li L L, et al. Telomere-to-telomere genome assemblies of cultivated and wild soybean provide insights into evolution and domestication under structural variation. Plant Commun, 2024, 5: 100919
19 Zhang C, Shao Z, Kong Y, et al. High-quality genome of a modern soybean cultivar and resequencing of 547 accessions provide insights into the role of structural variation. Nat Genet, 2024, 56: 2247–2258
20 Fan J, Shen Y, Chen C, et al. A large-scale integrated transcriptomic atlas for soybean organ development. Mol Plant, 2025, 18: 669–689
21 Liu Z, Kong X, Long Y, et al. Integrated single-nucleus and spatial transcriptomics captures transitional states in soybean nodule maturation. Nat Plants, 2023, 9: 515–524
22 Liu Y, Zhang Y, Liu X, et al. SoyOmics: a deeply integrated database on soybean multi-omics. Mol Plant, 2023, 16: 794–797
23 Ohyanagi H, Sakata K, Komatsu S. Soybean Proteome Database 2012: update on the comprehensive data repository for soybean proteomics. Front Plant Sci, 2012, 3: 110
24 Tian Z, Wang X, Lee R, et al. Artificial selection for determinate growth habit in soybean. Proc Natl Acad Sci USA, 2010, 107: 8563–8568
25 Ping J, Liu Y, Sun L, et al. *Dt2* is a gain-of-function MADS-domain factor gene that specifies semideterminacy in soybean. Plant Cell, 2014, 26: 2831–2842
26 Xia Z, Watanabe S, Yamada T, et al. Positional cloning and characterization reveal the molecular basis for soybean maturity locus *E1* that regulates photoperiodic flowering. Proc Natl Acad Sci USA, 2012, 109
27 Lu S, Zhao X, Hu Y, et al. Natural variation at the soybean J locus improves adaptation to the tropics and enhances yield. Nat Genet, 2017, 49: 773–779
28 Lu S, Dong L, Fang C, et al. Stepwise selection on homeologous PRR genes controlling flowering and maturity during soybean domestication. Nat Genet, 2020, 52: 428–436
29 Li Y, Zhang L, Wang J, et al. Flowering time regulator *qFT13-3* involved in soybean adaptation to high latitudes. Plant Biotechnol J, 2024, 22: 1164–1176
30 Wang S, Liu S, Wang J, et al. Simultaneous changes in seed size, oil content and protein content driven by selection of *SWEET* homologues during soybean domestication. Natl Sci Rev, 2020, 7: 1776–1786
31 Zheng H, Feng X, Wang L, et al. GmSop20 functions as a key coordinator of the oil-to-protein ratio in soybean seeds. Adv Sci, 2025, 12: e05181
32 Qin L, Zhao J, Tian J, et al. The high-affinity phosphate transporter GmPT5 regulates phosphate transport to nodules and nodulation in soybean. Plant Physiol, 2012, 159: 1634–1643

33 Wang J H, Chen M, Zhao X, et al. Regulatory network of soybean potassium efficient absorption gene *GmAKTs* and its related transcription factors (in Chinese). Chin Sci Bull, 2024, 69: 1588–1602 [王建华, 陈明, 赵雪, 等. 大豆钾高效吸收基因*GmAKTs*及其相关转录因子的调控网络研究. 科学通报, 2024, 69: 1588–1602]

34 Wang X, Zhao J, Fang Q, et al. GmAKT1 is involved in K^+ uptake and Na^+/K^+ homeostasis in *Arabidopsis* and soybean plants. Plant Sci, 2021, 304: 110736

35 Fu M, Sun J, Li X, et al. Asymmetric redundancy of soybean nodule inception (*NIN*) genes in root nodule symbiosis. Plant Physiol, 2022, 188: 477–489

36 Lim C W, Lee Y W, Hwang C H. Soybean nodule-enhanced CLE peptides in roots act as signals in GmNARK-mediated nodulation suppression. Plant Cell Physiol, 2011, 52: 1613–1627

37 Du M, Gao Z, Li X, et al. Excess nitrate induces nodule greening and reduces transcript and protein expression levels of soybean leghaemoglobins. Ann Bot, 2020, 126: 61–72

38 Guan R, Qu Y, Guo Y, et al. Salinity tolerance in soybean is modulated by natural variation in *GmSALT 3*. Plant J, 2014, 80: 937–950

39 Wang X, Wang J, Huang L, et al. A near-gapless genome assembly of soybean line Zhongpin 03-5373 enables structural variation based dissection of SCN3 resistance. J Integr Agr, 2026, doi: 10.1016/j.jia.2026.04.032

40 Shaibu A S, Zhang S, Ma J, et al. The GmSNAP11 contributes to resistance to soybean cyst nematode race 4 in *Glycine max*. Front Plant Sci, 2022, 13: 939763

41 Yang Q, Hou Z Y, Li L, et al. Landscape and m^6A post-transcriptional regulation of soybean proteome. Cell Genomics, 2025, 5: 100926

42 Wang X, Komatsu S. Subcellular proteomics to elucidate soybean response to abiotic stress. Plants, 2023, 12: 2865

43 Tian M, Sun Y, Zhang G, et al. Metabolomics navigates natural variation in pathogen-induced secondary metabolism across soybean cultivar populations. Proc Natl Acad Sci USA, 2025, 122: e2505532122

44 Zhang J, Liu S, Liu C B, et al. Natural variants of molybdate transporters contribute to yield traits of soybean by affecting auxin synthesis. Curr Biol, 2023, 33: 5355–5367.e5

45 Ziegler G, Nelson R, Granada S, et al. Genomewide association study of ionomic traits on diverse soybean populations from germplasm collections. Plant Direct, 2018, 2: e00033

46 Li X, Zhang X, Schmitz R J. From duplication to divergence: single-cell insights into transcriptional and *cis*-regulatory landscapes in soybean. Plant Cell, 2025, 37: koaf279

47 Zhang X, Luo Z, Marand A P, et al. A spatially resolved multi-omic single-cell atlas of soybean development. Cell, 2025, 188: 550–567.e19

48 Pelletier J M, Chen M, Lin J Y, et al. Dissecting the cellular architecture and genetic circuitry of the soybean seed. Proc Natl Acad Sci USA, 2025, 122: e2416987121

49 Sun L Q, Zhang Q Y, He Y H, et al. RT-DETR-based multiphenotypic measurement and yield prediction of soybean plants (in Chinese). Trans Chin Soc Agric Eng, 2026, 42: 192–200 [孙岚琪, 张琦沅, 贺昱航, 等. 基于RT-DETR的大豆植株多表型测量及产量预测. 农业工程学报, 2026, 42: 192–200]

50 Lima C S, Junior D F U, Carvalho I R, et al. Use of phenomics in the selection of UAV-based vegetation indices and prediction of agronomic traits in soybean subjected to flooding. AgriEngineering, 2024, 6: 3261–3278

51 Guan H, Li J, Ma X, et al. Research progress on rapid detection technology of soybean phenotypic indicators under saline-alkali stress. Planta, 2026, 264: 6

52 Jiang X, Zhang M, Yuan X, et al. Population epigenomics reveals epigenetic drivers of replicated evolution and missing heritability in soybean. Mol Plant, 2026, 19: 1100–1116

53 Fang Y, Han Y, Fang Y, et al. Epigenetic regulation modulates seasonal temperature-dependent growth of soybean in southern China. Plant Biotechnol J, 2025, 23: 4580–4601

54 Jin L, Zhang Y, Guo J, et al. Epigenetically poised chromatin states regulate PRR and NLR genes in soybean. aBIOTECH, 2025, 6: 411–423

55 Zhao X, Xu Y, Cai J, et al. Metabolomic and transcriptomic analyses of nutritional quality differences and underlying molecular mechanisms in soybeans with different coat colors. J Agric Food Chem, 2025, 73: 32922–32934

56 Li S, Zhang Y, Dong H, et al. Integrated transcriptomics, proteomics, and metabolomics offer novel insights into salt resistance in soybean. Industrial Crops Products, 2025, 237: 122212

57 Liu P, Li M, Ma P, et al. Spatiotemporal transcriptomic and metabolomic landscapes of wild soybean seed development reveal regulatory mechanisms of nutrient accumulation. Plant Commun, 2026, 7: 101580

58 Tardieu F, Cabrera-Bosquet L, Pridmore T, et al. Plant phenomics, from sensors to knowledge. Curr Biol, 2017, 27: R770–R783

59 Araus J L, Cairns J E. Field high-throughput phenotyping: the new crop breeding frontier. Trends Plant Sci, 2014, 19: 52–61

60 Chandnani R, Qin T, Ye H, et al. Application of an improved 2-dimensional high-throughput soybean root phenotyping platform to identify novel genetic variants regulating root architecture traits. Plant Phenomics, 2023, 5: 0097

61 Yang W, Feng H, Zhang X, et al. Crop phenomics and high-throughput phenotyping: past decades, current challenges, and future perspectives. Mol Plant, 2020, 13: 187–214

62 Taghavi Namin S, Esmaeilzadeh M, Najafi M, et al. Deep phenotyping: deep learning for temporal phenotype/genotype classification. Plant Methods, 2018, 14: 66

63 Wen W L, Guo X Y, Zhang Y, et al. Standard framework construction of technology and equipment for big data in crop phenomics (in Chinese). Engneer, 2023, 25: 227–238 [温维亮, 郭新宇, 张颖, 等. 作物表型组大数据技术及装备发展研究. 中国工程科学, 2023, 25: 227–238]

64 Pauli D, Chapman S C, Bart R, et al. The quest for understanding phenotypic variation via integrated approaches in the field environment. Plant Physiol, 2016, 172: 622–634

65 Li D, Bai D, Tian Y, et al. Time series canopy phenotyping enables the identification of genetic variants controlling dynamic phenotypes in soybean. J Integr Plant Biol, 2023, 65: 117–132

66 Ubbens J R, Stavness I. Deep plant phenomics: a deep learning platform for complex plant phenotyping tasks. Front Plant Sci, 2017, 8: 1190

67 Falk K G, Jubery T Z, Mirnezami S V, et al. Computer vision and machine learning enabled soybean root phenotyping pipeline. Plant Methods, 2020, 16: 5

68 Meng X, Li Y, Wang H. Synthetic biology in plant and agricultural research. J Integr Plant Biol, 2022, 64: 11–20

69 Wu H, Luo M, Liu Y, et al. Integrated biotechnological and artificial intelligence innovations for plant improvement. Front Plant Sci, 2025, 16: 1736707

70 Alok A, Raman V, D'agostino L, et al. Developmental regulators enable rapid and efficient soybean transformation and CRISPR-mediated genome editing. Plant Physiol, 2026, 200: kiaf640

71 Wurtzel E T. Changing form and function through carotenoids and synthetic biology. Plant Physiol, 2019, 179: 830–843

72 Li G, An L, Yang W, et al. Integrated biotechnological and AI innovations for crop improvement. Nature, 2025, 643: 925–937

73 South P F, Cavanagh A P, Liu H W, et al. Synthetic glycolate metabolism pathways stimulate crop growth and productivity in the field. Science, 2019, 363: eaat9077

74 Musazade E, Liu Y, Chen X, et al. Advances and challenges in haploid induction for warm-season legumes. J Agric Food Chem, 2025, 73: 6315–6332

75 Zhu H, Li C, Gao C. Applications of CRISPR-Cas in agriculture and plant biotechnology. Nat Rev Mol Cell Biol, 2020, 21: 661–677

76 Kong X, Fan K, Xin C, et al. GmRPS5 promoter-driven CRISPR/LbCas12a efficiently generates soybean sextuple mutants. Plant Biotechnol J, 2026, 24: 3494–3496

77 Lin W, Wu H, Kuang H, et al. Developing a robust multiplex CRISPR/Cas12i3-5M system for trait stacking in soybean. J Integr Plant Biol, 2026, jipb.70233

78 Li X, Qian X, Zhao F, et al. Knockout of GmCKX3 enhances soybean seed yield via cytokinin-mediated cell expansion and lipid accumulation. Plants, 2025, 14: 2207

79 Zhang P, Du H, Wang J, et al. Multiplex CRISPR/Cas9-mediated metabolic engineering increases soya bean isoflavone content and resistance to soya bean mosaic virus. Plant Biotechnol J, 2020, 18: 1384–1395

80 Lowe K, Wu E, Wang N, et al. Morphogenic regulators *Baby boom* and *Wuschel* improve monocot transformation. Plant Cell, 2016, 28: 1998–2015

81 Goshisht M K. Machine learning and deep learning in synthetic biology: key architectures, applications, and challenges. ACS Omega, 2024, 9: 9921–9945

82 Zhao X, Meng Z, Wang Y, et al. Pollen magnetofection for genetic modification with magnetic nanoparticles as gene carriers. Nat Plants, 2017, 3: 956–964

83 Li Y, Lu K, Feng S, et al. ROTIS: a RUBY-assisted one-shot tissue-culture-free system for *in-planta* soybean transformation from germinated seedlings. Plant Physiol, 2025, 199: kiaf609

84 Wu X, Wang Y, Ni Q, et al. GmYSL7 controls iron uptake, allocation, and cellular response of nodules in soybean. J Integr Plant Biol, 2022, 65: 167–187

85 Zhong Y, Yang M, Cheng D, et al. Mutation of *GmDMP* genes triggers haploid induction in soybean. bioRxiv, 2024, doi: 10.1101/2024.03.24.585499

86 Zhong Y, Chen B, Li M, et al. A DMP-triggered *in vivo* maternal haploid induction system in the dicotyledonous *Arabidopsis*. Nat Plants, 2020, 6: 466–472

87 Jang J H, Seo H S, Widiez T, et al. Loss-of-function of gynoecium-expressed phospholipase pPLAIIγ triggers maternal haploid induction in Arabidopsis. New Phytol, 2023, 238: 1813–1824.

88 Xu Y. Envirotyping for deciphering environmental impacts on crop plants. Theor Appl Genet, 2016, 129: 653–673

89 Jarquín D, Crossa J, Lacaze X, et al. A reaction norm model for genomic selection using high-dimensional genomic and environmental data.

Theor Appl Genet, 2014, 127: 595–607
90 Montesinos-López O A, Montesinos-López A, Pérez-Rodríguez P, et al. A review of deep learning applications for genomic selection. BMC Genomics, 2021, 22: 19
91 Crossa J, Pérez-Rodríguez P, Cuevas J, et al. Genomic selection in plant breeding: methods, models, and perspectives. Trends Plant Sci, 2017, 22: 961–975
92 Jubair S, Domaratzki M. Crop genomic selection with deep learning and environmental data: a survey. Front Artif Intell, 2023, 5: 1040295
93 Sun R, Sun B, Tian Y, et al. Dissection of the practical soybean breeding pipeline by developing ZDX1, a high-throughput functional array. Theor Appl Genet, 2022, 135: 1413–1427
94 Technow F, Messina C D, Totir L R, et al. Integrating crop growth models with whole genome prediction through approximate Bayesian computation. PLoS One}, 2015, 10: e0130855
95 Khalilzadeh Z, Kashanian M, Khaki S, et al. A hybrid deep learning-based approach for optimal genotype by environment selection. Front Artif Intell, 2024, 7: 1312115
96 Walter A, Finger R, Huber R, et al. Smart farming is key to developing sustainable agriculture. Proc Natl Acad Sci USA, 2017, 114: 6148–6150
97 Kamilaris A, Prenafeta-Boldú F X. Deep learning in agriculture: a survey. Comput Electron Agr, 2018, 147: 70–90
98 Basso B, Antle J. Digital agriculture to design sustainable agricultural systems. Nat Sustain, 2020, 3: 254–256
99 Zhao C, Zhang Y, Du J, et al. Crop phenomics: current status and perspectives. Front Plant Sci, 2019, 10: 714
100 Smith P, Martino D, Cai Z, et al. Greenhouse gas mitigation in agriculture. Phil Trans R Soc B, 2008, 363: 789–813
101 Singh A, Ganapathysubramanian B, Singh A K, et al. Machine learning for high-throughput stress phenotyping in plants. Trends Plant Sci, 2016, 21: 110–124
102 Li X X, Yang Y Q, Zhong Y J, et al. Genetic improvement of nutrient-efficient root systems of legume crops for adaptation to acid soils (in Chinese). J South China Agric Univ, 2019, 40: 186–194 [李欣欣, 杨永庆, 钟永嘉, 等. 豆科作物适应酸性土壤的养分高效根系遗传改良. 华南农业大学学报, 2019, 40: 186–194]
103 Bulgarelli D, Schlaeppi K, Spaepen S, et al. Structure and functions of the bacterial microbiota of plants. Annu Rev Plant Biol, 2013, 64: 807–838
104 Dong Y, Yang X, Liu J, et al. Pod shattering resistance associated with domestication is mediated by a NAC gene in soybean. Nat Commun, 2014, 5: 3352
105 Tian Z X, Liu B H, Yang Y P, et al. Achievements and prospects of soybean molecular design breeding in China (in Chinese). Bull Chin Acad Sci, 2018, 33: 915–921 [田志喜, 刘宝辉, 杨艳萍, 等. 我国大豆分子设计育种成果与展望. 中国科学院院刊, 2018, 33: 915–921]
106 Cooper M, Messina C D, Podlich D, et al. Predicting the future of plant breeding: complementing empirical evaluation with genetic prediction. Crop Pasture Sci, 2014, 65: 311–336

Summary for “多组学融合与人工智能驱动的大豆智能设计育种: 进展与展望”

Multi-omics integration and AI-driven intelligent design for soybean breeding: progress and prospects

Xin Su[1,2], Zhi-Feng Zhu[3], Xiao-Feng Gu[4] & Li-Juan Qiu[1,3*]

[1] *Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing 100081, China*
[2] *Shennong Seed Industry Laboratory/Henan Crop Molecular Breeding Research Institute, Zhengzhou 450002, China*
[3] *Syngenta Group China/China Seed Group Co., Ltd., Beijing 100081, China*
[4] *Institute of Biotechnology, Chinese Academy of Agricultural Sciences, Beijing 100081, China*
* Corresponding author, E-mail: qiulijuan@caas.cn

Agricultural biological breeding is rapidly accelerating its paradigm shift from traditional, empirical breeding toward a new era of Artificial Intelligence (AI) design, characterized by the deep integration of Biotechnology (BT) and Information Technology (IT). This transition serves as a crucial strategic lever for ensuring national food security and propelling the green transformation of the agricultural sector in China. Soybean, functioning as a vital multi-purpose crop for food, edible oil, and livestock feed, has long confronted significant bottlenecks in the enhancement of its unit yield and broad environmental adaptability. These enduring challenges fundamentally stem from the highly complex polygenic basis governing essential agronomic traits—such as yield potential, nutritional quality, and stress tolerance. Furthermore, these traits are heavily influenced by profound genotype-by-environment (G×E) interactions and are strictly constrained by the imperative need for the efficient utilization of resources, including water, soil, fertilizers, and pesticides. Consequently, traditional breeding methodologies struggle to achieve synergistic breakthroughs across multiple desirable traits while simultaneously maintaining yield stability.

To address these limitations, this comprehensive review systematically synthesizes the frontier progress of BT and IT integration within soybean research. The paper is structured around a whole-chain trajectory: “multi-omics trait dissection → engineered germplasm creation → intelligent G×E prediction → the integrative application of superior varieties, optimal agronomic practices, and smart equipment”. In the realm of Biotechnology (BT), we outline the unprecedented value of pan-genomics and multi-dimensional omics in deciphering core genetic regulatory networks, alongside pivotal technological breakthroughs such as gene editing, tissue-culture-free transformation methodologies, and the structural reconstruction of metabolic pathways via synthetic biology. Concurrently, we examine IT-empowered pathways—specifically encompassing AI, high-throughput phenomics, advanced envirotyping, and smart agriculture. Through this dual-lens approach, we elucidate the core transformative role of BT+IT integration in dissecting complex agronomic traits, intelligently formulating novel germplasm, facilitating highly efficient G×E predictions, and enabling the synergistic deployment of elite cultivars with optimized management and intelligent hardware.

Furthermore, this article meticulously traces the ongoing technological evolution of AI-driven G×E prediction models, whole-genome intelligent crossing and prediction (WPICP) pipelines, and agricultural digital twin systems. It summarizes the integrative innovation paradigms surrounding green prevention and control, highly efficient nutrient utilization, and intelligent agronomic equipment. Ultimately, the paper highlights that China has already established distinctive, cumulative advantages in the specialized domains of soybean stress tolerance, biological nitrogen fixation, and intelligent factory-based breeding. Looking forward, the critical pathways requiring imminent breakthroughs include generative intelligent design propelled by large foundational models in life sciences, the systematic reconstruction of complex metabolic networks utilizing synthetic biology, and the multidimensional synergistic optimization of genotype-by-environment-by-management (G×E×M) interactions. Collectively, these advancements will provide robust theoretical foundations and technological support for the precision breeding and sustainable green productivity enhancement of soybeans.

soybean, biological breeding, convergence of BT and IT, genotype-by-environment interaction, intelligent design breeding, smart agriculture

doi: 10.1360/CSB-2026-0608