



人工选择对作物基因组的影响: 从演化记录到改良策略

乔天乐, 黄洪宇, 宋雨茹, 苏晓, 王妍, 林涛*

中国农业大学园艺学院, 北京 100193

* 联系人, E-mail: lintao35@cau.edu.cn

收稿日期: 2025-08-08; 接受日期: 2026-01-05; 网络版发表日期: 2026-04-08

国家自然科学基金优秀青年科学基金(批准号: 32422078)和天津市自然科学基金(批准号: 25JCYBJC00550)资助

摘要 人类对农作物的定向选择是文明发展的基石, 为现代社会提供坚实的物质基础. 在过去约一万年中, 通过无意识与有意识的选择过程, 野生植物保留了更符合人类需求的有益性状. 这些历史选择的印记, 被永久地记录在作物基因组之中, 构成当今作物遗传改良的核心驱动力. 随着测序技术、基因组组装和群体遗传学的飞速发展, 现阶段已能够深入解析这些关键的遗传变异. 然而, 实现基于基因组学的精准设计育种, 仍面临诸多挑战. 本文旨在系统综述作物及群体基因组学的研究进展, 重点分析人工选择对全球前68种主要作物基因组产生的深远影响, 并对检测选择信号的主流方法进行归纳总结. 最后, 立足于可持续农业与全球粮食安全的迫切需求, 本文提出了面向未来的作物改良策略, 为精准培育具有气候适应性与资源高效性的新一代作物品种构建理论框架.

关键词 人工选择, 作物, 驯化, 群体遗传学, 基因功能

自21世纪以来, 全球对农业产出的需求攀升至前所未有的高度. 在此背景下, 现代粮食生产已形成一套集约化体系. 联合国粮食及农业组织(Food and Agriculture Organization of the United Nations, FAO)的研究表明, 农业生产活动已覆盖全球超过三分之一的陆地面积, 其大规模扩张引发了包括淡水资源短缺、土地沙漠化与土壤盐渍化在内的一系列生态环境问题. 在作物驯化与改良的过程中, 其演化动力主要源于人工选择与自然选择的共同作用, 其中人工选择占据主导地位, 而自然选择则扮演辅助角色^[1]. 具体而言, 自然选择是一个无意识的筛选机制, 它在资源竞争与环境波动的压力下, 优先将资源分配给繁殖力更强的个体,

以最大化后代数量, 其最终目标是提升个体在胁迫环境中的适应度. 与之相反, 人工选择则表现出强烈的目的性, 通过有意识或无意识的筛选, 持续改良并强化作物的经济性状, 诸如非落粒性、贮藏器官肥大化及株型紧凑化等. 正如Doebley所指出的, 这种选择的核心理念在于最大化作物的可收获部分, 从而极大地推动农业产出效率的提升^[2].

作物基因组学的迅猛发展, 是解析人工选择机制的核心驱动力. 其技术演进主要经历三个关键阶段: 首先, 水稻、玉米等关键作物参考基因组的成功测序, 为该领域提供作物基因组草图^[3~5]; 随后, 2006年下一代测序(next-generation sequencing, NGS)技术的出现,

引用格式: 乔天乐, 黄洪宇, 宋雨茹, 等. 人工选择对作物基因组的影响: 从演化记录到改良策略. 中国科学: 生命科学, 2026, 56: 1015–1026

Qiao T L, Huang H Y, Song Y R, et al. The impact of artificial selection on crop genomes: from evolutionary records to improvement strategies (in Chinese). Sci Sin Vitae, 2026, 56: 1015–1026, doi: 10.1360/SSV-2025-0245

因成本大幅降低, 驱动研究焦点从“单一基因组解析”迈向“群体遗传变异挖掘”, 如基于NGS的水稻基因组变异图谱, 便在此阶段厘清粳稻的起源与籼稻的驯化路径^[6,7]; 近年来, PacBio SMRT等单分子长读长测序技术的应用, 进一步攻克复杂基因组的组装难题, 成功构建玉米、小麦等作物的T2T完整基因组图谱, 为精准解析选择信号在复杂基因组区域的分布提供高质量的参考依据^[8~10]。

尽管长达万余年的人工选择塑造现代农作物的遗传结构, 并成功支撑全球人口的生存与发展, 然而, 当前作物遗传增益仍难以满足可持续农业与粮食安全的需求^[11,12]。面对这一挑战, 作物生产系统亟需一场旨在同步提升作物环境适应性与突破产量潜力的“双轨突破”^[13]。要实现这一宏伟目标, 必须首先深刻理解农作物的演化历程。这不仅有助于把握作物育种的内在规律, 更能有效规避未来育种对生态系统可能造成的负面影响。

当前已有大量研究致力于探索人工选择对作物演化过程的影响^[14~16], 但仍需从单一基因组解析向群体水平多尺度的跨越, 系统阐释不同作物中人工选择的作用机制及检测方法。鉴于作物基因组学和群体遗传学在挖掘重要农艺性状遗传变异中具有核心作用, 本综述将为作物科学与育种实践提供多维度参考。首先总结重要农作物的基因组测序与组装技术进展, 阐明如何精准组装大型、复杂的杂合作物基因组; 继而分析育种如何通过海量基因组数据重塑作物基因组, 探讨群体遗传分析的新兴工具, 以及挖掘调控重要农艺性状的关键基因; 最后对未来育种策略提出展望, 以应对2050年全球97亿人口的粮食安全与质量挑战。

1 农作物参考基因组

全球人类社会的存续与发展主要依赖于约100种作物的遗传资源。其中, 前68种核心作物(包括11种谷物、3种块根作物、2种糖料作物、11种豆类、7种油料作物、15种蔬菜、2种木本粮食作物和17种水果作物, 图1和网络版附表1)构成现代农业的支柱。从植物分类学区分, 多数谷物属于单子叶植物, 而水果和蔬菜则主要属于双子叶植物。此外, 除小麦和玉米等特例外, 多数作物的基因组相对紧凑。这可能反映自然选择与人工选择的协同演化机制: 较小的基因组规模有

助于提高重组率, 从而加速有利等位基因的固定, 并加速驯化出符合人类需求的农艺性状。为全面解析人工选择在作物基因组中的印记, 首要任务是构建染色体级别的高质量参考基因组。

1.1 作物基因组组装技术的演进与突破

基因组组装质量的核心在于测序技术与组装策略的协同发展。早期作物(如水稻^[3,4]、葡萄^[17]、高粱和玉米^[5,18])的基因组测序依赖于全基因组鸟枪法与BAC/fosmid文库的结合。但Sanger测序的高成本与基因组本身的复杂性(高杂合度、高重复序列), 限制该策略在作物上的广泛应用。

NGS技术(如Roche/454 FLX, SOLiDTM和Illumina/Solexa)在2006年后的崛起^[6,19], 标志着新阶段的开始^[20]。初期由于成本限制, 仅完成草莓^[21]和番茄^[22]等简单基因组的组装。尽管Illumina GA因成本优势成为主流, 但其短读长起初仅能适用于小型基因组。2009年, SOAPdenovo对黄瓜基因组的组装是首个重要突破^[23,24]。随后, 算法与组装策略的发展解决作物杂合性难题, 构建马铃薯高质量基因组^[25]。而多平台数据整合策略提升了水稻组装质量^[26]。尽管如此, 多倍体基因组的组装仍是难点, 例如油菜组装中亚基因组的同源性导致归属歧义^[27]。面包小麦的组装则需依赖费时且繁杂的染色体臂纯化策略^[28,29]。当前, PacBio SMRT和ONT单分子测序技术的超长读长(>10 kb)彻底改变局面, 极大地提升组装的连续性与复杂区域的解析能力^[8,30,31]。尽管早期存在高正确率与高成本的问题, 但HiFi reads等算法与校正技术的成熟已推动其广泛应用, 并显著改善结构变异的检测。然而, 高同源/杂合区域导致的组装正确、成本与计算复杂度, 依然是该技术面临的核心挑战^[32]。

1.2 作物基因组组装的挑战与技术突破

测序技术的革新仅是驱动基因组学发展的一个维度。面对高杂合、高重复的作物基因组, 新型文库构建策略与组装算法的创新同样至关重要。早期研究主要依赖BAC、Fosmid和双端(mate-pair)文库来连接重叠群; 然而, 随着技术进步, 10× Genomics、高通量染色体构象捕获(high-throughput chromosome conformation capture, Hi-C)和Bionano光学图谱等技术相继问世, 显著提升Scaffolding的水平。当前, 结合长读长与链接读

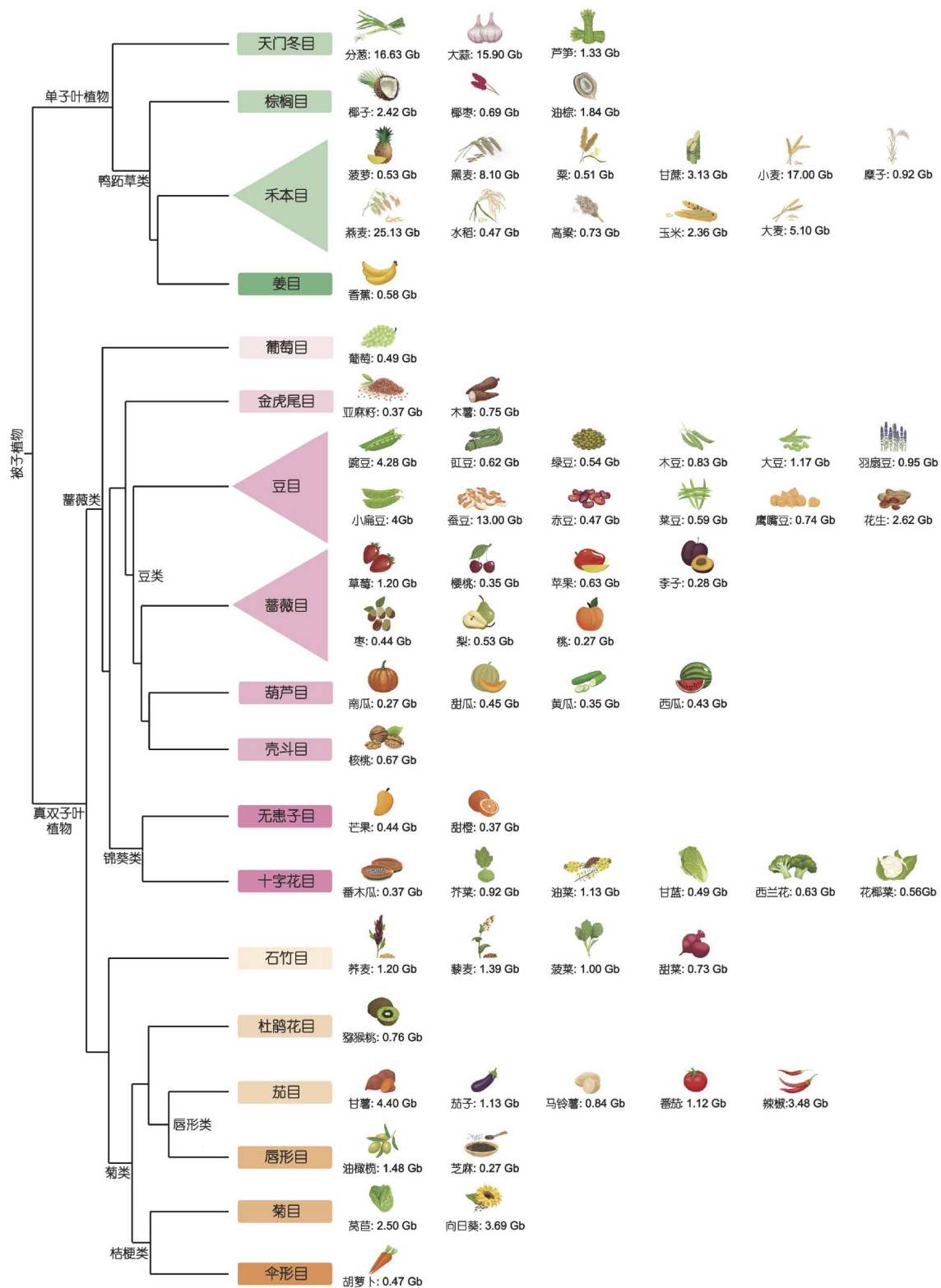


图1 68种主要作物的基因组测序

Figure 1 Genome sequencing of the top 68 crops

长的联合策略已成为解析复杂重复序列的有效手段, 并成功推动玉米、六倍体小麦等大型基因组的端粒到端粒(telomere-to-telomere, T2T)组装, 深入揭示着丝粒区卫星DNA的动态扩增机制^[9,10].

近十年来, 多技术联合的组装策略已显著提升作物基因组在连续性、准确率与完整度方面的表现. 在此基础上, 作物基因组学研究已全面进入群体规模时代. 通过对野生种与栽培种质资源进行大规模重测序, 研究者得以在全基因组范围内系统解析包括单核苷酸多态性(single nucleotide polymorphism, SNP)、插入缺失(insertion-deletion, Indel)及结构变异(structural variation, SV)在内的遗传多样性, 并在此基础上构建基因型-表型关联网络. 综上所述, 测序技术、文库构建与组装算法的协同创新, 正在全球范围内推动新一轮作物基因组从头测序的浪潮.

2 作物群体的重测序

在作物演化过程中, 人工选择塑造了典型的“驯化综合征”, 其特征性状组合如果实增大、风味改良及株型健壮等, 共同满足人类对产量与品质的需求^[33]. 然而, 仅依赖单一基因组测序无法全面解析这些性状的遗传机制. 群体基因组重测序为此提供关键突破, 它通过系统分析不同地理来源的种质资源, 构建覆盖全谱系的遗传变异图谱, 从而揭示目标性状的遗传基础. 随着测序成本持续下降, 针对多种作物的大规模、系统性测序已成为可能.

对群体基因组数据中的遗传变异进行比较分析, 是揭示作物演化轨迹的重要窗口. 然而, 依赖短读长序列比对至单一参考基因组的方法, 在鉴定高度多样化个体中的结构变异时仍存在明显局限. 值得注意的是, SVs可能影响约三分之二的重要农艺性状, 在作物改良中扮演关键角色^[34]. 例如, 小麦4B染色体短臂上500 kb的缺失, 以及辣椒中*Up*基因的8 kb缺失, 均导致显著的表型变化^[35,36]. 在作物基因组中, SVs所影响的碱基对范围远超SNPs, 其主要受三大因素调控: (i) 驯化历史: 野生与栽培种质在驯化过程中产生显著遗传分化, 野生种质通常保有更高的遗传多样性; (ii) 交配系统: 自交作物(如水稻)可通过自然选择清除高危害性SVs^[34]; 克隆繁殖作物(如葡萄)在驯化过程中可能将有害SVs维持在杂合状态^[37], 而异交物种(如玉

米)通常表现出更显著的基因组结构变异; (iii) 繁殖方式: 马铃薯、甘蔗和香蕉等无性繁殖作物的个体间基因组变异更为丰富, 这可能与其经历的遗传瓶颈效应较弱有关^[38].

近年来, 长读长测序与高质量基因组组装技术为群体测序捕获大片段SV提供新的路径. 由此催生出泛基因组研究策略, 旨在整合核心基因集与非必需基因集, 实现作物遗传变异的全景解析. 该策略通常选取代表性种质及野生近缘种进行从头测序和组装, 以此阐明基因获得与丢失事件的演化轨迹.

展望未来, 整合群体基因组变异数据与转录组、表观基因组等多组学信息, 将极大地推动作物育种中关键遗传资源的挖掘. 为此, 未来亟需建立全球统一标准的作物基因组变异数据库体系, 可依托北京基因组研究所(Beijing Institute of Genomics, BIG)、美国国家生物技术信息中心(National Center for Biotechnology Information, NCBI)等平台, 并由国际研究联盟牵头启动“全作物种质百科全书”计划, 对全球核心种质资源进行系统解析, 以共同推动作物科学的长期发展.

3 人工选择的轨迹

在人类定向选择重要农艺性状的过程中, 作物基因组经历了深刻重塑, 具体表现为有利等位基因的固定、全基因组或局部区域遗传多样性的衰减, 以及连锁累赘效应的打破. 然而, 现阶段对这些选择轨迹的认知仍不完整. 深入解析这一过程, 不仅能揭示其背后的遗传与分子机制, 更将为作物的重新驯化与定向设计育种提供坚实的理论框架.

3.1 从关键基因到全基因组的动态演进

作物的早期驯化过程主要源于人类对关键农艺性状的定向筛选, 其驯化模式通常聚焦于少数主效基因. 这些基因通过单一位点突变或表达水平的变化, 直接调控作物的核心性状. 例如, 水稻和玉米中控制籽粒灌浆的*Sweet4*基因, 以及大豆中调控种子休眠的*G*基因, 均在驯化过程中被强烈选择, 凸显关键基因在驯化早期的核心作用^[39,40].

随着泛基因组等技术的突破, 研究视野得以从关键基因扩展至全基因组尺度. 基于野生种、地方品种与现代栽培种的群体重测序分析, 研究者能够在全基

基因组范围内识别出大量与重要数量性状位点(quantitative trait loci, QTL)共定位的选择信号。这些信号点共同构成一个多基因协同调控的复杂驯化网络, 标志着人工选择策略已从“单点突破”转向“网络调控”^[41]。

综上所述, 通过对多物种研究的整合分析发现, 人工选择的轨迹是一个动态演化过程: 它通常以少数关键基因为初始锚点, 继而通过全基因组范围内的多基因协同优化, 最终实现作物性状的全面改良。这种“关键基因选择-全基因组协同选择-性状优化”的递进模型, 系统地揭示作物驯化与改良的遗传学本质。

3.2 人工选择的双重轨迹——遗传增益与代价

人类通过定向选择, 在作物中固定一系列有利等位基因, 使其逐渐形成依赖人工栽培的生存模式。例如, 水稻的*qSH1*基因及小麦的*Btr1/2*基因簇协同控制种子不落粒性状, 是作物驯化中定向选择作用典型例证^[42,43]。然而, 性状改良常伴随功能上的权衡。例如, 葫芦科作物中苦味合成基因簇的丢失虽然提升食用品质, 却导致防御性代谢物缺失与病原易感性增强^[44,45], 这一现象揭示“一因多效”的普遍遗传机制: 当选择压力集中于特定农艺性状时, 可能打破与其他性状相关的基因网络平衡。

与此同时, 人类定向选择对作物遗传多样性产生深刻的影响。(i) 正向选择促使目标性状相关的等位基因在群体中快速固定; (ii) 遗传瓶颈导致全基因组范围内的多样性降低。在早期育种中, 由于仅从野生种中筛选少数具目标性状的个体, 致使栽培种的全基因组遗传多样性显著低于其野生祖先^[34]。此外, 而人类迁徙与环境变化引发的群体瓶颈效应, 进一步加剧遗传多样性的流失。这种全基因组水平的遗传侵蚀, 不仅反映人工选择的强度, 也预示着未来育种可能因遗传基础狭窄而面临挑战。

有利等位基因的固定常伴随邻近基因的“遗传搭车”现象。这一过程受局部重组率制约, 往往导致目标区域遗传多样性降低及有害突变的被动积累, 进而产生诸多负面效应, 例如增加作物对病虫害的易感性^[46]。连锁累赘效应是其主要表现, 在番茄中, 源自秘鲁番茄(*Solanum peruvianum*)的*Tm-2a*抗病毒基因, 由于染色体倒置导致重组抑制, 在9号染色体(51.7~54.7 Mb)区域形成顽固的连锁累赘, 致使多个果实发育异常基因无法被有效剔除^[47,48], 同时, 果实重量基因*fw11.3*在驯

化过程中受到选择时, 其两个侧翼基因也因搭车效应被固定, 这些基因与7种代谢物显著关联, 可能因此引入潜在的不利代谢性状^[49]。这些基因在野生番茄向栽培种的选择性清除中发生搭车效应, 可能成为潜在的有害基因。

尽管对人工选择轨迹的解析极大深化学界对农艺性状遗传基础的理解, 但遗传多样性的衰减与搭车效应仍是当前作物遗传改良的主要瓶颈。有研究表明, 番茄基因组中约有一万个基因位点在驯化、改良及抗病基因渗入过程中被修饰, 这些区域有望通过优化的育种策略被逐步剔除^[47]。展望未来, 随着作物群体基因组测序的完善, 未来有望精准定位控制有利性状的因果基因, 进而通过变异组引导筛选稀有重组事件, 或利用基因组编辑技术直接打破连锁累赘, 从而为解除遗传搭车效应开辟新的技术路径。

4 人工选择的检测方法

考古证据已证实作物演化与新石器时代定居农业存在密切关联^[50]。然而, 人工选择在作物基因组中留下的演化轨迹仍存在大量认知空白。作物单基因组与群体重测序技术的革新, 为解析人工选择留下的选择信号提供重要契机。与此同时, 连锁图谱分析与群体遗传学等方法已被广泛应用于检测作物中的选择信号。

4.1 遗传关联分析与高通量测序技术在选择特征检测中的应用

遗传关联分析旨在通过连锁不平衡(linkage disequilibrium, LD)原理, 识别与目标表型显著相关的遗传变异。传统研究多通过构建野生种与栽培种的杂交分离群体, 进行性状的精细定位, 从而检测人工选择留下的选择信号。例如, 玉米的植株结构基因*tb1*和*tga1*、水稻的落粒性基因*sh4*和*qSH1*、小麦的*Q*基因以及番茄果实重量基因*fw2.2*等, 均是通过此类方法在栽培种与其野生祖先的杂交群体中成功定位^[51~57]。然而, 该策略耗时长、效率低, 且需要依赖于大量的遗传标记才能精确锁定控制性状的基因座。

近年来, 高通量测序技术的兴起显著推动选择信号检测方法的革新。基于NGS的集群分离分析(bulked segregant analysis, BSA-seq)与全基因组重测序技

术^[58-61], 已成为快速筛选候选基因组区域的主流手段。其中, QTL-seq和MutMap构成BSA-seq体系中两种应用最广泛的技术路线^[62,63]。QTL-seq能够从栽培种与野生种的分离群体中鉴定出多个QTL, 并已成功应用于定位番茄从野生醋栗番茄到樱桃番茄, 再到大果栽培番茄的“两步演化”过程中受选择的关键QTL^[47]; 而MutMap则更适用于由单个主效基因控制的质量性状研究, 例如, 在黄瓜中成功鉴定出调控果实心皮发育的ACC氧化酶基因, 以及与矮化性状相关的*F-box*蛋白基因^[64,65]。尽管NGS技术极大地加速QTL定位进程, 但其检测范围在庞大的作物基因组中仍相对有限, 仅能覆盖被选择的部分区域。

4.2 全基因组关联研究与群体遗传分析在作物选择特征检测中的应用

4.2.1 GWAS在农艺性状基因定位中的突破

多数农艺性状由复杂的多基因网络调控, 传统依赖单一分离群体的QTL定位方法在解析全基因组关联基因时存在一定局限。随着测序成本的下降, 利用NGS技术对数百份作物种质及突变体进行大规模基因组测序已成为可能。海量的基因组数据结合高质量的参考基因组与遗传变异图谱, 共同推动全基因组关联研究(genome-wide association study, GWAS)与群体遗传分析成为解析人工选择的选择信号和作物驯化机制的核心策略。

GWAS通过构建高分辨率的基因组变异图谱, 已成为解析复杂农艺性状遗传基础的核心工具, 能够在全基因组尺度上定位相关基因并鉴定因果遗传变异。该方法已在多种作物群体中取得显著进展。例如, 通过268个水稻品种的联合分析, 在有效校正群体结构后, 精准定位到8个重要农艺性状的关键调控突变^[66]; 基于540份玉米自交系群体的研究, 揭示*YIG1*基因通过生长素信号通路调控穗长的分子机制^[67]; 此外, 对600余份番茄种质分析, 成功鉴定出一系列与果皮颜色、果实重量及代谢物积累相关的候选基因^[47,49,68]。需要强调的是, GWAS的统计效力受群体规模、群体结构、变异图谱质量以及表型精度等多重因素影响, 这些因素共同决定结果的可靠性。

4.2.2 群体遗传统计学在选择区域检测中的应用

群体遗传学参数, 如遗传多样性(π , Watterson's

θ), 等位基因频率谱(Tajima's D)以及群体分化指数(F_{st} , XP-CLR), 可在不依赖特定表型的前提下有效检测基因组中经历自然或人工选择的区域。首先, 通过遗传多样性分析可以揭示选择与瓶颈效应的影响。驯化过程导致的遗传瓶颈使得栽培作物的全基因组多样性普遍低于其野生近缘种^[69], 而在选择区域, 局部多样性会进一步降低。基于全基因组变异图谱, 研究人员在水稻中定位55个驯化相关位点(如*sh4*, *PROG1*), 并据此揭示栽培稻的起源路径: 粳稻起源于中国南方珠江流域的野生稻群体, 而籼稻则是粳稻与当地野生稻杂交的后代^[7]。类似地, 对番茄的演化研究则解析了从野生种经樱桃番茄到大果栽培番茄的两次选择历程, 并鉴定了诸如调控果皮颜色的*MYB*基因等关键位点^[47]。其次, Tajima's D 统计量通过比较遗传多样性的分布来推断选择压力, 其负值通常指示了近期定向选择或群体扩张事件。在黄瓜的驯化研究中, Tajima's D 分析的结果验证了基于核苷酸多样性所发现的选择扫荡区域^[70]。此外, 基于群体分化的方法, 如XP-CLR, 可通过比较等位基因频率的差异来定位选择区域。在玉米中, 该方法检测到484个与驯化(从野生种到地方品种)相关的区域和695个与改良(从地方品种到现代自交系)相关的区域^[71]; 在大豆中, 则鉴定出121个驯化扫描区域和109个改良扫描区域^[72]。同样, F_{st} 通过量化群体间的等位基因频率差异来揭示局部适应机制。例如, 籼稻与粳稻在粒形调控基因*GS3*上表现出中等的遗传分化^[7]; 而在番茄中, 高分化位点(如控制可溶性固形物的*ssc5.1-5.3*基因和控制果实硬度的*fir5.1*基因)则清晰地标示加工番茄与鲜食番茄的分化特征^[47]。

4.3 表观遗传标记在选择检测中的补充作用

表观遗传修饰(如DNA甲基化、组蛋白修饰)是基因组与环境互作的关键媒介, 其变异模式能够反映长期人工选择所施加的间接效应。在番茄驯化研究中, 果实大小相关基因启动子区域的甲基化水平变化与人工选择强度显著相关, 并且这类表观标记能够稳定地跨代遗传^[73]。通过整合ChIP-seq与群体表观基因组数据, 可系统构建“表观选择图谱”, 有效检测基因组非编码区域中传统方法难以发现的选择信号, 并为全面理解作物驯化的分子机制提供全新的维度和见解。

4.4 多组学整合分析在作物选择基因鉴定中的应用与展望

尽管单一组学技术已成功鉴定出部分农艺性状相关基因,但在大数据背景下,其较高的假阳性率仍需通过多维度数据加以验证.为克服这一局限,整合基因组学、转录组学、代谢组学、表观基因组学及表型组学的多维组学分析,已成为系统解析选择基因生物学机制的主流策略^[74].这种多组学协同分析能够有效弥补单一数据集的信息缺失或可靠性不足^[75,76].受模式生物和人类ENCODE项目^[75,77]的启发,作物多组学数据正呈指数级增长,为解析复杂性状的调控网络与精准育种提供新的研究思路.例如,对玉米基因组、转录组与表型变异的联合分析,揭示现代育种中复杂的环境适应机制^[78];而对数百份番茄种质进行变异组、转录组及代谢组的整合挖掘,则系统阐释全球育种背景下果实代谢物的重塑历程^[49,79].后者从多组学视角重构番茄的代谢育种史,为代谢组辅助育种奠定理论基础.同样,在水稻研究中,通过整合不同组织及发育时期的转录组、代谢组与表观组数据,构建全球首个水稻多器官单细胞多组学图谱,并配套建立基因扰动模拟、转录因子功能预测和共表达网络等分析工具,为作物智能育种提供全新的方法论支撑^[80].

4.5 机器学习与人工智能在选择信号检测中的应用

随着对作物基因组中人工选择印记研究的深入,解析漫长驯化与改良过程中积累的复杂选择压力,日益依赖于机器学习(machine learning, ML)与人工智能(artificial intelligence, AI)所具备的高维数据解析能力.全基因组重测序所产生的庞大数据集(包含数百万计的SNP, Indel和SV)呈现出高度复杂的非线性关系,传统的群体遗传学方法在捕捉微弱、复杂或存在交互作用的选择信号时已显不足.而ML/AI算法,尤其是具备卓越模式识别与特征学习能力的深度学习(deep learning, DL)模型,正成为挖掘这些“隐藏”选择指纹的关键工具.

在此背景下,机器学习为解析复杂基因组数据中的选择信号提供全新的研究范式^[81].基于监督学习的模型能够整合GWAS结果与环境因子数据,有效区分自然选择与人工选择的选择信号,并快速筛选出受选择的关键基因^[82].同时,深度学习中的卷积神经网络

(convolutional neural network, CNN)等算法,展现出直接从原始基因组序列中识别选择区域特征模式的巨大潜力.

在利用上述数据解析人工选择信号时, GWAS、群体遗传参数分析与机器学习各具优势. GWAS基于连锁不平衡关联基因型与表型,适用于目标性状的直接遗传解析,其发现的关键位点能为分子标记开发和基因克隆提供明确方向;然而,该方法严重依赖大规模的基因分型与精准表型鉴定,且漫长的连锁不平衡区间易引入假阳性信号.群体遗传参数分析无需表型信息,通过检测对中性演化模型的偏离来快速筛选分化位点,但其短板在于难以区分人工选择与自然选择的作用.机器学习则擅长处理高维、复杂的多组学数据,抗噪声能力强,但其模型可解释性不足常被视为“黑箱”,最终仍需功能实验验证其预测.因此,构建一个高效、可靠的整合式基因鉴定策略显得至关重要.将连锁定位、群体遗传学分析与多组学数据相融合,能够揭示从基因到表型的完整调控网络,并通过多层次证据的交叉验证,实现基因的精细定位.未来,作物遗传研究必将实现从单一基因定位向多源大数据深度融合的范式跨越,从而为设计育种遴选更可靠的分子靶标,最终引领作物育种全面步入智能化时代.

5 作物基因组学与群体遗传学驱动的重新驯化策略

5.1 基于泛基因组的精准基因渗入技术

在泛基因组框架下,利用“核心基因”与“可变基因”的二分策略,定向导入野生种质中的有利等位基因,为作物遗传改良开辟新路径.在大豆研究中,基于泛基因组图谱成功鉴定出NDD2基因组所特有的13个结构变异,并进一步通过结构变异全基因组关联分析(SV-GWAS),系统发掘出一系列与大豆产量及种子品质显著关联的SVs.该研究不仅揭示大豆产量与品质形成的遗传基础,更构建了“变异-基因-表型”的系统解析框架,为复杂农艺性状的遗传解密提供新视角^[83].在棉花研究中,通过整合二倍体与四倍体棉花的基因组数据,构建首个跨倍性棉花图泛基因组,并证实驯化过程对不同倍性棉花种质资源的遗传模块存在差异化塑造^[84].水稻泛基因组研究则实现从“单一参考”

到“群体参考”的范式突破。通过整合全球多样的水稻种质资源(如籼稻、粳稻), 研究者构建高密度的图形泛基因组。结合meta-GWAS分析, 成功定位到如*GW10.2*等与产量相关的基因, 证实其复等位变异通过剂量效应调控穗粒数, 充分体现泛基因组在捕获稀有变异和解析复杂性状位点方面的强大能力^[85]。

尽管基于泛基因组的精准渗入与基因组编辑驱动的头驯化策略, 为突破传统育种瓶颈、拓展作物遗传资源提供新的路径, 但在实际应用中仍面临生态风险、技术瓶颈与遗传安全性等多重挑战。在生态层面, 编辑后的基因可能通过花粉基因流扩散至野生近缘种, 引发不可控的生态影响; 在技术层面, 该策略高度依赖于高质量的泛基因组, 而基因组高度复杂(如多倍体、高重复序列作物)或野生种质资源稀缺的作物仍存在组装与注释瓶颈, 如在多倍体作物中, 亚基因组间的高度同源性导致序列组装时易出现错配和重叠群归属模糊; 高重复序列区域即使采用长读长测序仍难以实现连续且准确的组装; 野生种质资源的匮乏则进一步限制遗传变异的全面捕获, 影响泛基因组的代表性和应用广度。更为关键的是, 在导入或编辑目标基因时, 若未能精准识别并规避连锁累赘区域, 可能导致有害等位基因随有益变异一同渗入, 进而影响作物的适应性或品质性状。因此, 未来需结合高精度泛基因组图谱与多组学功能注释, 建立针对结构变异和连锁区块的精准评估体系, 并通过基因编辑与重组优化技术切断不利连锁, 从而在提升育种效率的同时, 最大限度地控制遗传风险。

5.2 单碱基编辑技术在解除连锁累赘中的突破

碱基编辑器(如CBE, ABE)能够在不诱导DNA双链断裂的情况下实现精准的单碱基替换, 为打破作物育种中的连锁累赘提供关键工具。为此, 研究人员在小麦中靶向编辑*FT-D1*基因第三个外显子的特定碱基, 成功实现对小穗数和抽穗期的独立调控, 从而解耦这两个在传统育种中紧密连锁的性状。这验证了单碱基编辑技术在解除连锁累赘方面的高效性, 充分展示其在作物精准设计育种中的巨大应用潜力^[83]。

5.3 异源多倍体的从头驯化策略

从头驯化策略旨在通过定向改造野生近缘种的关键驯化基因, 使其快速获得不落粒、花期同步等典型

栽培性状。该策略对异源多倍体作物意义尤为显著: 这类野生种虽富含优异的抗逆基因, 却因缺乏关键的驯化性状而难以被直接应用于育种实践, 而从头驯化正可突破此限制。例如, 研究人员通过对异源四倍体野生稻的落粒性、开花时间及株型等相关基因进行编辑, 使其快速获得符合栽培需求的表型, 同时完整保留野生种的抗逆优势。这种“野生基因库与驯化性状”的有机融合, 为高效利用多倍体野生资源开展作物改良提供创新范式^[86]。

6 前景和挑战

本综述系统梳理了作物线性基因组与群体测序的技术突破、人工选择对作物基因组的遗传塑造, 以及针对核心作物的从头驯化等前沿策略。尽管遗传改良始终是提升作物产量与培育新品种的核心手段, 但当前的育种效率已难以满足可持续农业与全球粮食安全的迫切需求^[11]。有鉴于此, 为实现“培育更优作物以保障未来粮食与能源安全”的宏伟目标, 未来作物育种亟需在以下两个方向寻求突破: (i) 深度挖掘野生近缘种中蕴藏的丰富遗传变异, 系统性构建覆盖野生-栽培种质的泛基因组资源库; (ii) 充分利用基因组编辑(如CRISPR-Cas9)、精准重组调控及合成生物学等颠覆性技术, 实现作物性状的定向设计与优化。这些策略将共同推动作物育种从传统的“经验选择”范式, 向现代化的“精准设计”范式转型, 从而为应对日益严峻的全球粮食危机提供坚实的科技支撑。

6.1 种质资源保存与利用的现状与未来策略

过去一个世纪, 全球在种质资源保护方面取得显著进展, 尤其是在野生种质资源的发掘与利用上, 这对于扩展食物来源多样化与提升作物遗传多样性至关重要。国际组织在资源采集与保护中发挥核心主导作用, 各国科研机构亦持续投入巨额资金用于种质收集与保存。然而, 当前种质资源的利用效率依然偏低, 其在产量、营养品质等复杂性状改良中的贡献尤为有限。因此, 必须将种质资源的保护与高效利用深度融合, 方能突破现有遗传资源的限制。随着基因组测序技术的广泛应用, 种质库中蕴藏的遗传多样性亟待系统挖掘分析。为响应可持续农业与粮食安全需求, 本文建议对全球约3000种主要食用作物开展系统的基因

组测序、组装与功能解析,并将相关数据整合至统一共享平台,以实现科研资源的全球互通^[87]。此举将极大提升种质资源的利用价值,为保障全球粮食安全与农业可持续发展注入核心动力。

6.2 作物多性状协同改良的挑战与理性设计策略

在保障作物产量的基础上,品质与抗逆性已成为现代育种的核心目标^[88,89]。然而,产量提升常伴随品质与抗性的协同衰退,导致在多优性状聚合时效率低下。理性设计育种是应对该挑战的前沿方案,特别适用于多个复杂性状的金字塔式聚合。例如,通过聚合影响产量与米质的主效基因培育高产优质超级稻^[90],但此类复杂性状的聚合通常耗时十余年,并伴随着高昂的成本。鉴于传统育种的固有局限,应用从头驯化等新兴技术已显得尤为迫切。实现理性设计育种的关键,在于系统解析重要农艺性状的基因调控网络与信号通路,从而为新作物的创制提供精确的分子设计蓝图。

6.3 跨学科协作推动育种范式革新

作物育种正经历一场从传统“经验驱动”向“数据密集型”交叉学科的根本性变革。这一转型的核心是构建“基因型-环境-表型”数字化孪生系统,通过整合多尺度作物生理模型、基因组选择算法与机器学习等关键技术,在虚拟空间中高通量模拟数千种基因型在不同环境下的表型效应,从而大幅优化育种决策、缩短育种周期。具体而言,在从头驯化野生资源时,该系统能精准预测编辑关键驯化基因对植株整体生长和抗逆性的系统性影响,优先筛选最优编辑组合;在品质改良方面,通过整合代谢组与基因组数据,人工智能可反向设计优化特定营养成分的基因编辑靶点,实现风味与营养价值的理性调控。最终,这种“预测-设计-验证”一体化平台将彻底革新依赖试错的传统育种模式,推动作物育种步入可计算、可预测的智能化设计新纪元。为实现这一愿景,亟须建立融合基因组学、农学、计算科学等多学科的协同创新平台,为应对全球粮食安全挑战提供系统性解决方案。

参考文献

- 1 Darwin C. On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life. London: John Murray, 1859
- 2 Doebley J F, Gaut B S, Smith B D. The molecular genetics of crop domestication. *Cell*, 2006, 127: 1309–1321
- 3 Goff S A, Ricke D, Lan T H, et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science*, 2002, 296: 92–100
- 4 Yu J, Hu S, Wang J, et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *Indica*). *Science*, 2002, 296: 79–92
- 5 Schnable P S, Ware D, Fulton R S, et al. The B73 maize genome: complexity, diversity, and dynamics. *Science*, 2009, 326: 1112–1115
- 6 Bentley D R. Whole-genome re-sequencing. *Curr Opin Genet Dev*, 2006, 16: 545–552
- 7 Huang X, Kurata N, Wei X, et al. A map of rice genome variation reveals the origin of cultivated rice. *Nature*, 2012, 490: 497–501
- 8 van Dijk E L, Naquin D, Gorrison K, et al. Genomics in the long-read sequencing era. *Trends Genet*, 2023, 39: 649–671
- 9 Chen J, Wang Z, Tan K, et al. A complete telomere-to-telomere assembly of the maize genome. *Nat Genet*, 2023, 55: 1221–1231
- 10 Liu S, Li K, Dai X, et al. A telomere-to-telomere genome assembly coupled with multi-omic data provides insights into the evolution of hexaploid bread wheat. *Nat Genet*, 2025, 57: 1008–1020
- 11 Tilman D, Balzer C, Hill J, et al. Global food demand and the sustainable intensification of agriculture. *Proc Natl Acad Sci USA*, 2011, 108: 20260–20264
- 12 Gao C. Genome engineering for crop improvement and future agriculture. *Cell*, 2021, 184: 1621–1635
- 13 Godfray H C J, Beddington J R, Crute I R, et al. Food security: the challenge of feeding 9 billion people. *Science*, 2010, 327: 812–818
- 14 Abbo S, Lev-Yadun S, Gopher A. Plant domestication and crop evolution in the near east: on events and processes. *Crit Rev Plant Sci*, 2012, 31: 241–257
- 15 Curtin S, Qi Y, Peres L E P, et al. Pathways to *de novo* domestication of crop wild relatives. *Plant Physiol*, 2021, 188: 1746–1756
- 16 Wu D, Lao S, Fan L. De-domestication: an extension of crop evolution. *Trends Plant Sci*, 2021, 26: 560–574
- 17 Jaillon O, Aury J-M, Noel B, et al. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*, 2007, 449: 463–467
- 18 Paterson A H, Bowers J E, Bruggmann R, et al. The *Sorghum bicolor* genome and the diversification of grasses. *Nature*, 2009, 457: 551–556

- 19 Margulies M, Egholm M, Altman W E, et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 2005, 437: 376–380
- 20 Mardis E R. Next-generation DNA sequencing methods. *Annu Rev Genom Hum Genet*, 2008, 9: 387–402
- 21 Shulaev V, Sargent D J, Crowhurst R N, et al. The genome of woodland strawberry (*Fragaria vesca*). *Nat Genet*, 2011, 43: 109–116
- 22 Tomato Genome Consortium. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature*, 2012, 485: 635–641
- 23 Li R, Zhu H, Ruan J, et al. *De novo* assembly of human genomes with massively parallel short read sequencing. *Genome Res*, 2010, 20: 265–272
- 24 Huang S, Li R, Zhang Z, et al. The genome of the cucumber, *Cucumis sativus* L. *Nat Genet*, 2009, 41: 1275–1281
- 25 Xu X, Pan S, Cheng S, et al. Genome sequence and analysis of the tuber crop potato. *Nature*, 2011, 475: 189–195
- 26 Kawahara Y, de la Bastide M, Hamilton J P, et al. Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice*, 2013, 6: 4
- 27 Chalhou B, Denoeud F, Liu S, et al. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science*, 2014, 345: 950–953
- 28 Šafář J, Bartoš J, Janda J, et al. Dissecting large and complex genomes: flow sorting and BAC cloning of individual chromosomes from bread wheat. *Plant J*, 2004, 39: 960–968
- 29 Mayer K F X, Rogers J, Doležel J, et al. A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science*, 2014, 345: 1251788
- 30 Chen S, Huang X. DNA sequencing: the key to unveiling genome. *Sci China Life Sci*, 2020, 63: 1593–1596
- 31 Wang H, Chen M, Wei X, et al. Computational tools for plant genomics and breeding. *Sci China Life Sci*, 2024, 67: 1579–1590
- 32 Garg V, Bohra A, Mascher M, et al. Unlocking plant genetics with telomere-to-telomere genome assemblies. *Nat Genet*, 2024, 56: 1788–1799
- 33 Meyer R S, Purugganan M D. Evolution of crop species: genetics of domestication and diversification. *Nat Rev Genet*, 2013, 14: 840–852
- 34 Gaut B S, Seymour D K, Liu Q, et al. Demography and its effects on genomic variation in crop domestication. *Nat Plants*, 2018, 4: 512–520
- 35 Song L, Liu J, Cao B, et al. Reducing brassinosteroid signalling enhances grain yield in semi-dwarf wheat. *Nature*, 2023, 617: 118–124
- 36 Zhang K, Yu H, Zhang L, et al. Transposon proliferation drives genome architecture and regulatory evolution in wild and domesticated peppers. *Nat Plants*, 2025, 11: 359–375
- 37 Zhou Y, Muyle A, Gaut B S. Evolutionary genomics and the domestication of grapes. In: Cantu D, Walker M A, eds. *Pollen: The Grape Genome*. Cham: Springer, 2019. 39–55
- 38 White O W, Biswas M K, Abebe W M, et al. Maintenance and expansion of genetic and trait variation following domestication in a clonal crop. *Mol Ecol*, 2023, 32: 4165–4180
- 39 Soso D, Luo D, Li Q B, et al. Seed filling in domesticated maize and rice depends on SWEET-mediated hexose transport. *Nat Genet*, 2015, 47: 1489–1493
- 40 Wang M, Li W, Fang C, et al. Parallel selection on a dormancy gene during domestication of crops from multiple families. *Nat Genet*, 2018, 50: 1435–1441
- 41 Huang X, Huang S, Han B, et al. The integrated genomics of crop domestication and breeding. *Cell*, 2022, 185: 2828–2839
- 42 Ishikawa R, Castillo C C, Htun T M, et al. A stepwise route to domesticate rice by controlling seed shattering and panicle shape. *Proc Natl Acad Sci USA*, 2022, 119: e2121692119
- 43 Zhao Y L, Li M M. Research progress on molecular mechanism of seed shattering in triticeae crops (in Chinese). *J Triticeae Crops*, 2022, 42: 272–278 [赵永亮, 李美美. 麦类作物落粒控制的分子机制研究进展. *麦类作物学报*, 2022, 42: 272–278]
- 44 Zhong Y, Xun W, Wang X, et al. Root-secreted bitter triterpene modulates the rhizosphere microbiota to improve plant fitness. *Nat Plants*, 2022, 8: 887–896
- 45 Luo F, Huang Y, Sun Y, et al. Transcription factor CmWRKY13 regulates cucurbitacin B biosynthesis leading to bitterness in oriental melon fruit (*Cucumis melo* var. *Makuwa Makino*). *J Agric Food Chem*, 2024, 72: 24697–24710
- 46 Charlesworth B, Harvey P H, Barton N H. Genetic hitchhiking. *Phil Trans R Soc Lond B*, 2000, 355: 1553–1562
- 47 Lin T, Zhu G, Zhang J, et al. Genomic analyses provide insights into the history of tomato breeding. *Nat Genet*, 2014, 46: 1220–1226
- 48 Tanksley S D, Bernachi D, Beck-Bunn T, et al. Yield and quality evaluations on a pair of processing tomato lines nearly isogenic for the *Tm2a* gene for resistance to the tobacco mosaic virus. *Euphytica*, 1998, 99: 77–83
- 49 Zhu G, Wang S, Huang Z, et al. Rewiring of the fruit metabolome in tomato breeding. *Cell*, 2018, 172: 249–261.e12
- 50 Bilgic H, Hakki E E, Pandey A, et al. Ancient DNA from 8400 Year-Old Çatalhöyük wheat: implications for the origin of neolithic agriculture. *PLoS One*, 2016, 11: e0151974

- 51 Doebley J. The genetics of maize evolution. *Annu Rev Genet*, 2004, 38: 37–59
- 52 Doebley J, Stec A, Hubbard L. The evolution of apical dominance in maize. *Nature*, 1997, 386: 485–488
- 53 Kosugi S, Ohashi Y. DNA binding and dimerization specificity and potential targets for the TCP protein family. *Plant J*, 2002, 30: 337–348
- 54 Li C, Zhou A, Sang T. Rice domestication by reducing shattering. *Science*, 2006, 311: 1936–1939
- 55 Simons K J, Fellers J P, Trick H N, et al. Molecular characterization of the major wheat domestication gene *Q*. *Genetics*, 2006, 172: 547–555
- 56 Wang H, Nussbaum-Wagler T, Li B, et al. The origin of the naked grains of maize. *Nature*, 2005, 436: 714–719
- 57 Frary A, Nesbitt T C, Frary A, et al. *fw2.2*: a quantitative trait locus key to the evolution of tomato fruit size. *Science*, 2000, 289: 85–88
- 58 Darvasi A, Soller M. Selective DNA pooling for determination of linkage between a molecular marker and a quantitative trait locus. *Genetics*, 1994, 138: 1365–1373
- 59 Giovannoni J J, Wing R A, Ganai M W, et al. Isolation of molecular markers from specific chromosomal intervals using DNA pools from existing mapping populations. *Nucleic Acids Res*, 1991, 19: 6553–6568
- 60 Mansur L M, Orf J, Lark K G. Determining the linkage of quantitative trait loci to RFLP markers using extreme phenotypes of recombinant inbreds of soybean (*Glycine max* L. Merr.). *Theoret Appl Genet*, 1993, 86: 914–918
- 61 Michelmore R W, Paran I, Kesseli R V. Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *Proc Natl Acad Sci USA*, 1991, 88: 9828–9832
- 62 Takagi H, Abe A, Yoshida K, et al. QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *Plant J*, 2013, 74: 174–183
- 63 Abe A, Kosugi S, Yoshida K, et al. Genome sequencing reveals agronomically important loci in rice using MutMap. *Nat Biotechnol*, 2012, 30: 174–178
- 64 Chen H, Sun J, Li S, et al. An ACC oxidase gene essential for cucumber carpel development. *Mol Plant*, 2016, 9: 1315–1327
- 65 Lin T, Wang S, Zhong Y, et al. A truncated F-box protein confers the dwarfism in cucumber. *J Genet Genomics*, 2016, 43: 223–226
- 66 Yu J, Zhu C, Xuan W, et al. Genome-wide association studies identify OsWRKY53 as a key regulator of salt tolerance in rice. *Nat Commun*, 2023, 14: 3550
- 67 Luo Y, Zhang M, Liu Y, et al. Genetic variation in *YIGE1* contributes to ear length and grain yield in maize. *New Phytol*, 2022, 234: 513–526
- 68 Tieman D, Zhu G, Resende Jr. M F R, et al. A chemical genetic roadmap to improved tomato flavor. *Science*, 2017, 355: 391–394
- 69 Doebley J. Isozymic evidence and the evolution of crop plants. In: Soltis D E, Soltis P S, Dudley T R, eds. *Pollen: Isozymes in Plant Biology*. Dordrecht: Springer, 1989. 165–191
- 70 Qi J, Liu X, Shen D, et al. A genomic variation map provides insights into the genetic basis of cucumber domestication and diversity. *Nat Genet*, 2013, 45: 1510–1515
- 71 Hufford M B, Xu X, van Heerwaarden J, et al. Comparative population genomics of maize domestication and improvement. *Nat Genet*, 2012, 44: 808–811
- 72 Zhou Z, Jiang Y, Wang Z, et al. Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat Biotechnol*, 2015, 33: 408–414
- 73 Guo H, Cao P, Wang C, et al. Population analysis reveals the roles of DNA methylation in tomato domestication and metabolic diversity. *Sci China Life Sci*, 2023, 66: 1888–1902
- 74 Holzinger E R, Ritchie M D. Integrating heterogeneous high-throughput data for meta-dimensional pharmacogenomics and disease-related studies. *Pharmacogenomics*, 2012, 13: 213–222
- 75 Ritchie M D, Holzinger E R, Li R, et al. Methods of integrating data to uncover genotype-phenotype interactions. *Nat Rev Genet*, 2015, 16: 85–97
- 76 Li H, Yang X, Shang Y, et al. Vegetable biology and breeding in the genomics era. *Sci China Life Sci*, 2023, 66: 226–250
- 77 The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*, 2012, 489: 57–74
- 78 Liu H, Wang X, Warburton M L, et al. Genomic, transcriptomic, and phenomic variation reveals the complex adaptation of modern maize breeding. *Mol Plant*, 2015, 8: 871–884
- 79 Zhu G, Gou J, Klee H, et al. Next-Gen approaches to flavor-related metabolism. *Annu Rev Plant Biol*, 2019, 70: 187–212
- 80 Wang X, Huang H, Jiang S, et al. A single-cell multi-omics atlas of rice. *Nature*, 2025, 644: 722–730
- 81 Liu J, Li J, Wang H, et al. Application of deep learning in genomics. *Sci China Life Sci*, 2020, 63: 1860–1878
- 82 Liu Z, Wang N, Su Y, et al. Grapevine pangenome facilitates trait genetics and genomic breeding. *Nat Genet*, 2024, 56: 2804–2814

- 83 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
- 84 Li J, Liu Z, You C, et al. Convergence and divergence of diploid and tetraploid cotton genomes. *Nat Genet*, 2024, 56: 2562–2573
- 85 Yang L, He W, Zhu Y, et al. GWAS meta-analysis using a graph-based pan-genome enhanced gene mining efficiency for agronomic traits in rice. *Nat Commun*, 2025, 16: 3171
- 86 Yu H, Lin T, Meng X, et al. A route to *de novo* domestication of wild allotetraploid rice. *Cell*, 2021, 184: 1156–1170.e14
- 87 Borlaug N E. Contributions of conventional plant breeding to food production. *Science*, 1983, 219: 689–693
- 88 Francis D, Finer J J, Grotewold E. Challenges and opportunities for improving food quality and nutrition through plant biotechnology. *Curr Opin Biotechnol*, 2017, 44: 124–129
- 89 Martin C, Li J. Medicine is not health care, food is health care: plant metabolic engineering, diet and human health. *New Phytol*, 2017, 216: 699–719
- 90 Zeng D, Tian Z, Rao Y, et al. Rational design of high-yield and superior-quality rice. *Nat Plants*, 2017, 3: 17031

The impact of artificial selection on crop genomes: from evolutionary records to improvement strategies

QIAO TianLe, HUANG HongYu, SONG YuRu, SU Xiao, WANG Yan & LIN Tao*

College of Horticulture, China Agricultural University, Beijing 100193

* Corresponding author, E-mail: lintao35@cau.edu.cn

Directional artificial selection in crops has been a cornerstone of civilization, providing the material foundation for modern society. Over the past 10000 years, unconscious and conscious selection have transformed wild plants, endowing them with favorable traits that better meet human demands. The genomic signatures of these historical selection events are permanently encoded within crop genomes, serving as a primary driver for modern genetic improvement. Rapid advances in sequencing technologies, genome assembly, and population genetic now enable the deep characterization of these critical genetic variations. However, achieving genomics-based precision breeding remains a significant challenge. This review systematically summarizes progress in crop and population genomics, focusing on the profound impact of artificial selection on the genomes of 68 major global crops and reviewing mainstream approaches for detecting selection signals. Finally, addressing the urgent needs of sustainable agriculture and global food security, we propose future-oriented strategies for crop improvement, establishing a theoretical framework for the precision breeding of next-generation varieties characterized by climate resilience and resource-use efficiency.

artificial selection, crops, domestication, population genetics, gene function

doi: [10.1360/SSV-2025-0245](https://doi.org/10.1360/SSV-2025-0245)



林涛, 中国农业大学园艺学院教授, 博士生导师, 主要研究方向为基因组学与品质育种。2016年于中国农业科学院蔬菜花卉研究所获得博士学位。2019年被引进中国农业大学。2024年获得国家自然科学基金委青年科学基金项目(B类)。一直致力于利用基因组学推动番茄等蔬菜作物的遗传育种基础研究, 揭示番茄驯化和育种的基因组学历史, 解析果重与品质协同调控的遗传机制。共发表高水平SCI论文20余篇, 并开发基因结构注释等新算法2项, 培育番茄新品种2个。研究成果对功能基因组学研究和基因鉴定等具有重要的指导作用, 可推动番茄种质资源和相关分子标记的遗传育种应用。