

王亚琦, 孙子淇, 郑 峥, 等. 作物分子标记辅助选择育种的现状与展望[J]. 江苏农业科学, 2018, 46(5): 6-12.
doi:10.15889/j.issn.1002-1302.2018.05.002

作物分子标记辅助选择育种的现状与展望

王亚琦¹, 孙子淇², 郑 峥², 黄冰艳², 董文召², 汤丰收², 张新友^{1,2}

(1. 河南科技大学农学院, 河南洛阳 471000; 2. 河南省农业科学院经济作物研究所/农业部黄淮海油料作物重点实验室/河南省油料作物遗传改良重点实验室, 河南郑州 450002)

摘要:分子标记辅助选择育种可以通过精准选择目标性状提高育种效率, 加快育种进程。分子标记辅助选择已在主要农作物育种中广泛应用, 全基因组选择正逐步成为研究热点, 有望推动分子标记辅助选择育种技术的更快发展。综述了分子标记辅助育种的基础及在农作物中的应用进展, 全基因组选择的原理、方法、优势以及全基因组选择在植物育种方面的应用, 并对作物分子选择育种作出了展望。

关键词:作物育种; 分子标记辅助选择; 全基因组选择; 基础; 应用; 展望

中图分类号: S33 **文献标志码:** A **文章编号:** 1002-1302(2018)05-0006-07

传统的植物育种主要依赖于育种家对植株表现型的选择。环境条件、基因间互作、基因型与环境互作等多种因素均会影响植株表型选择准确性从而降低育种效率。一个优良品种的培育往往需花费 7~8 年甚至十几年时间。如何提高选择效率, 是育种工作的关键。近年来, 随着分子生物学和基因组学的快速发展, 分子选择育种应运而生, 利用已掌握的植物表型与基因型相关信息, 研究人员可以直接利用基因型数据对表型进行选择。分子选择育种包括前景选择和背景选择, 前景选择建立在基因定位和 QTL 作图的基础之上, 前景选择的可靠性取决于标记与目标基因之间的连锁程度, 标记与目标基因连锁越紧密, 则标记辅助育种的准确率越高, 而背景选择主要指遗传背景的恢复, 育种材料间遗传距离、亲缘关系分析等。

分子选择育种经历了 2 个重要的阶段, 第 1 个阶段为分子标记辅助选择 (molecular marker assisted selection, MAS), 第 2 个阶段为全基因组选择 (GS)^[1]。分子标记辅助育种, 采用与目标基因紧密连锁的分子标记, 筛选具有特定基因型的个体, 并结合常规育种方法选育优良品种, 此方法建立在 QTL 作图和基因定位的研究基础和数据之上。全基因组选择可以被定义为在全基因组水平上进行分子选择育种的方法。全基因组策略包括对收集的种质资源的全基因组测序, 分析重要、优异 (或低劣) 性状单倍型、重要基因区段、功能基因的分子标记开发等。全基因组选择要配合建立多个环境下目标性状高效的表型鉴定系统, 同时考虑影响基因、基因型、完整植株表现的相关环境因素信息的整合。

相比于传统育种, 分子选择育种具备以下优点: (1) 可缩

短育种年限, 提高育种效率, 节约大量的人力、物力; (2) 分子标记可在作物生长的各个阶段进行检测, 大幅度减少田间性状调查时间; (3) 分子标记选择不受外界环境的影响, 可在任何环境条件 (温室、异地等) 下进行分子标记选择, 提高了某些特殊性状, 如仅在特定环境下表达的性状抗病基因、耐胁迫基因被选择的有效性; (4) 相对于表型选择, 分子标记选择对于遗传力低的性状 (如抗旱、产量性状), 隐性性状 (如抗病性状), 表型鉴定困难、花费高的性状 (如加工品质), 多个抗性基因聚合的性状的选择具有显著作用^[2-5]。

1 分子标记辅助育种的基础

1.1 作图群体的选择

1.1.1 双亲作图群体 选择合适的作图亲本及构建适当的群体类型是成功和高效进行基因型及 QTL 定位的 2 个关键因素。目前, 已有多种类型的群体用于遗传图谱的构建, 其中最常用的是双亲作图群体。按照其特征进行分类, 可分为暂时性群体和永久性群体。其中, 暂时性群体主要包括 F₂ 群体及其衍生的 F₃ 家系、BC₁ 群体。暂时性分离群体最显著的特点是含有可发生分离的杂合基因型后代, 遗传信息丰富。永久性群体主要包括重组自交系群体、双单倍体群体、F₁ 花粉分化和用于 QTL 精细定位的近等基因系群体、染色体片段代换系群体等。永久性分离群体最显著的特点是不同株系间的基因型存在差异, 而同一个株系内不同个体间的基因型无差异, 永久性分离群体后代稳定, 通过自交其后代不会再次分离, 故其可永久使用。

Khedikar 等利用高产易感晚斑病和锈病的花生材料 TAG24 和高产抗叶斑病和锈病的材料 GPBD4 构建 RIL 群体, 利用 SSR 标记进行遗传图谱构建, 结合田间表型调查数据对晚斑病抗性 QTL 和锈病抗性 QTL 进行定位, 结果获得 1 个锈病抗性主效 QTL, 可解释 6.90%~55.20% 的表型变异, 标记 IPAHM 103 与该 QTL 紧密关联, 可解释 55.20% 的表型变异^[6]。利用多份抗病材料和易感材料对 IPAHM 103 进行验证, 结果表明, 该标记可靠, 可进一步用于花生抗锈病的分子标记辅助育种中。针对分子标记育种效率还不太高等问题,

收稿日期: 2016-11-03

基金项目: 河南省重大科技专项 (编号: 141100110600); 国家花生产业技术体系建设专项 (编号: CARS-14); 河南省现代农业产业技术体系建设专项 (编号: S2012-5)。

作者简介: 王亚琦 (1990—), 女, 河南济源人, 硕士研究生, 研究方向为作物遗传育种。E-mail: 2335738351@qq.com。

通信作者: 张新友, 博士, 研究员, 长期从事花生育种工作。E-mail: haasz@126.com。

研究人员提出了一系列的对策。Tanksley 等提出,利用高代回交导入系同时进行 QTL 分析和遗传改良,其主要是将 QTL 的定位分析推迟到如 BC₂、BC₃ 等高世代,使 QTL 的检测和有利等位变异的导入相结合^[7]。同时在该类群体的基础上可以进一步构建高级群体,即在 QTL 初定位基础上选择目标 QTL 区间杂合的单株再自交或回交 1~2 次,结合分子标记辅助选择即可得到 QTL 的近等基因系(QTL - NILs)。利用此近等基因系一方面可以验证 QTL 的真实性,另一方面又可以获得经过改良特定数量性状的稳定品系,直接应用于大田生产。

研究表明,基于双亲群体的 QTL 作图是定位数量性状的有效方法之一^[8-9],但双亲作图群体仍存在一些弊端:(1)由于这些群体建立在双亲的遗传基础之上,导致群体遗传差异仅限于双亲之间;(2)双亲作图群体只能对目标性状或相关性状进行定位;(3)由于双亲作图群体有限的重组使得定位区间偏大,若想进一步对目标性状进行精细定位,发掘目标基因,进行分子标记辅助选择育种和群体改良,需要进一步缩小定位区间。此外,对于某些自花授粉作物来说,双亲之间的多态性更低,利用双亲构建群体进行 QTL 定位,效果不太理想。

1.1.2 多亲本作图群体 巢式关联作图群体是由一个共同亲本和多个亲本杂交获得一系列重组自交系群体的集合。Yu 等采用 25 个具有遗传多样性的玉米自交系与参考系进行杂交,构建了包含约 250 个自交系的 25 个 RIL 群体^[10]。Kump 等利用包含 25 个家系的 NAM 群体对玉米小斑病进行全基因组关联分析,鉴定获得 32 个 QTLs,其中包含多个微效基因,认为 NAM 可显著提高作图分辨率^[11]。Bergelson 等系统比较了巢式关联作图与巢式连锁作图,发现将二者结合起来有利于图谱的精细定位^[12]。另外,多亲本高级世代互交系是另外一个多亲本作图群体,它是利用多个亲本杂交、然后互交,再随机交配,最后自交产生重组自交家系。

由于多个亲本的互交使其多态性增加,可以提高 QTL 定位精度。Kover 等在拟南芥上利用 19 个材料互交获得包含 572 个 RILs 的 MAGIC 群体,并利用 1 026 个 SNPs 标记对该多亲本高级世代互交系群体进行了全基因组扫描,分析结果表明,利用该群体定位的 QTL 与已知的 QTL 极为接近,并缩小了定位区间。说明多亲作图群体进行全基因组扫描功效更高且更加精确^[13]。

1.1.3 自然群体 自然群体进化历史的各种变异和交换重组,大大增加了 QTL 定位精度,甚至可以直接定位到基因本身,因此,更有利于分子标记辅助育种的应用及 QTL 克隆。但是,自然群体的群体结构往往会产生假阳性结果。高通量基因分型技术的不断提高,DNA 测序技术的不断更新,以及关联作图中群体结构统计方法的发展促进了作物自然群体在全基因组关联作图中的应用。作为关联分析在植物应用上的里程碑,Atwell 等对拟南芥 107 表型进行全基因组关联分析,获得多个相关性状的主效基因,GWAS 在拟南芥上的成功应用为其他作物进行 GWAS 分析提供了良好的思路^[14]。Zhao 等利用 44 100 个 SNPs 对来自 82 个国家的具有遗传多样性的 413 份水稻材料进行全基因组关联分析,就形态、发育、农艺性状等 34 个性状展开了系统的研究,发现大量复杂性状的常见变异基因,并将遗传变异和代谢途径有效的结合在一起,

为品种发展和作物改良提供了平台^[15]。Morris 等对收集的 971 份全球各地适应多种农业气候的高粱品种进行测序,获得约 265 000 个 SNPs 标记用于图谱构建,并发掘经典株高位点,花序结构候选基因^[16]。该图谱可用于分子标记辅助育种和全基因组选择。于志远等收集了 327 大豆材料,这些材料由优异亲本材料和东北主推品种组成;利用 186 对 SSR 标记对所有材料基因型进行检测,并进一步对所收集大豆材料的含油量和蛋白质含量进行关联分析;最终检测到 12 个与含油量极显著关联的位点,11 个与蛋白质含量极显著关联的位点^[17]。

1.2 标记类型和基因分型技术

分子标记及其在基因组中的位置对分子育种策略的选择起着至关重要的作用。分子标记种类较多,大致可分为以下几类:(1)基于 DNA - DNA 分子杂交的分子标记,如 RFLP (restriction fragment length polymorphism) 标记、CAPS (cleaved amplified polymorphic Sequence) 标记、VNTR (variable number of tandem repeats) 标记等。(2)基于 PCR (polymerase chain reaction) 的分子标记,如 SSR (simple sequence repeat) 标记、ISSR (inter - simple sequence repeat) 标记、ISTR (inverse sequence - tagged repeat) 标记、STS (sequence tag site) 标记等。(3)基于 PCR 与限制性酶切技术结合的分子标记,如 AFLP (amplified fragments length polymorphism) 标记等。(4)基于测序的单核苷酸多态分子标记,如 SNP (single Nucleotide polymorphisms) 标记。分子育种技术也在不断地发展与更新。相对于第一代基于分子杂交的 RFLP 标记,基于 PCR 的 SSR 标记大大减少了基因分型的费用和时间,从而促进分子育种的快速发展与应用。相对于 SSR 标记,SNP 标记在基因组中覆盖率更广,近年来,随着测序技术的快速发展,多种作物中大量的 SNP 标记被开发出来,并构建了高密度分子图谱,从而为分子育种奠定了坚实的基础^[18-20]。

此外,单倍型也被认为是一种分子标记,它常常被定义为来自一个特异基因区段或单个染色体上多个标记的组合,还可以是来自不同染色体上控制同一目标性状的标记的组合。试验表明,利用单倍型替代 SNP 进行目标性状的多样性分析和关联作图效果更加理想。植物中第 1 个单倍型图谱来自于拟南芥,研究人员利用包含 341 602 个非单态型 SNPs 的基因芯片对 19 个样本进行分析,并获得了全基因组 LD^[21]。Gore 等利用 27 份多样化的玉米自交系进行单倍型作图,最终获得图谱包含 300 万以上个 SNP 标记,平均每 40 bp 就可以检测到 1 个 SNP^[22]。我国研究人员已在玉米中发现了 5 000 万以上个 SNPs,并利用这些标记构建玉米的第 2 个单倍型图谱^[23]。Huang 等在水稻中,利用新一代测序技术对 517 份水稻品种进行测序,识别了 300 万以上个 SNPs,并利用这些标记构建水稻单倍型图谱^[24]。

与其他标记相比,功能标记不会与基因发生重组,因此,功能标记是分子育种的理想标记。目前,小麦的抗白粉病功能标记^[25-28]、抗条锈病功能标记^[29]、玉米高直链淀粉功能标记^[30]、花生高油酸功能标记^[31-32]陆续被开发出来,从而为相关目标性状的分子标记育种奠定了坚实基础。

基因分型大致分为 3 个阶段,分别是凝胶电泳、基因芯片和测序。伴随着基因分型技术的发展,基因分型通量大幅度

提高——由原来的少数几个标记增加到成千上万个标记同时对样品进行基因分型。相比于凝胶电泳,基因芯片技术(DArT)在增加基因分型通量的同时降低了基因分型的费用^[33],该技术已在小麦、大麦、高粱等作物上得到广泛应用^[34-37]。还有其他类型芯片,如 affymetrix 基因芯片、Sequenom 质谱分型芯片等。随着高通量测序成本的不断降低,测序分型迅速成为进行高通量基因分型的一种方法,越来越受到广大科研工作者的青睐。Wang 等利用新一代测序技术对包含 100 个单株的 F₁ 群体进行测序,共开发出 1 814 个高质量的 SNP 标记,并构建一张包含 1 646 个标记的葡萄遗传图谱^[38]。Zhou 等利用新一代测序技术对包含 166 个单株的 RIL 群体及双亲进行测序,构建了一张包含 1 621 个 SNPs 的花生遗传图谱^[39]。基因分型技术的发展为分子标记育种大规模应用创造了条件。

1.3 作图方法

常用的作图方法有连锁作图、关联作图。连锁作图常以双亲本群体或多亲本群体为作图群体,它利用标记和标记、标记和基因之间的重组率来确定其在染色体上的相对位置,进而构建遗传图谱。随着测序技术的发展和 SNP 标记的应用,以连锁不平衡为基础的关联作图逐渐成为一种可信赖的作图方法。相比于连锁作图,关联作图有以下几点优势:(1)关联作图可以直接利用自然群体,不需要构建群体,因此节省大量的人力物力。(2)关联作图利用自然群体广泛的遗传背景和长期积累的历史变异及大量的重组信息,大大增加了图谱的分辨率,有效提高了 QTL 的定位精度。(3)关联作图可以同时检测多个性状,而连锁作图仅能检测目标性状及相关少数几个性状。(4)关联作图可以直接利用已有的历史数据进行关联分析。Erena 等通过关联作图发掘出稳定遗传的小麦籽粒产量 QTL,并获得与之紧密关联的分子标记^[40]。Su 等对 200 个美国冬小麦进行关联分析,鉴定出 3 个同时与小麦千粒质量和粒长紧密关联的 SNP 标记^[41]。但关联作图也存在一些弊端,如自然群体存在一定的群体结构和亲缘关系,容易引起假阳性的关联,关联分析对效应值较大的稀有等位基因定位效果较差,而连锁分析对于该类等位基因定位效果较好,由此可见,连锁作图与关联作图是存在互补的。

为了有效地结合连锁作图与关联作图的优势,Myles 等提出了连锁图谱与关联作图图谱整合的策略^[42]。Lu 等利用 3 个 RIL 群体构建连锁图谱,并利用包含 305 个自交系的自然群体进行关联作图,最终将连锁图谱与关联图谱进行整合,构建了玉米的第 1 张整合图谱,并因此发现了单独作图未能定位的 18 个玉米抗旱相关 QTL^[43]。Li 等利用双亲群体和自然群体对玉米株高和穗高进行连锁作图和关联作图,共定位 21 个 QTLs、41 个 SNPs,同时结合连锁和关联分析,获得 1 个株高候选基因和 1 个穗高候选基因,为玉米株高、穗高遗传研究提供了基础^[44]。

2 分子选择育种发展

随着分子生物学和基因组学等学科的飞速发展,分子选择育种也在不断发展。近年来,测序技术的不断发展及测序成本的降低使得全基因组大规模测序成为可能,从而形成了基于全基因组策略的分子选择育种^[1]。全基因组策略可以

被定义为在全基因组水平上进行分子选择育种的工具包和方法论。全基因组策略的发展包括对所有登记在册的种质资源进行全基因组测序,重要基因组区域、基因及功能基因区域分子标记的开发,多种目标性状高效的表型鉴定系统(在多个环境中进行测量)的建立,基因、基因型、表现型及相关环境因素的整合等。

2.1 分子标记辅助选择

第一阶段的分子标记辅助选择包括分子标记辅助回交选择和分子标记辅助轮回选择。这个阶段可利用的分子标记数量有限,并且主要依赖于主效 QTL 的定位精度。在具体的育种过程中一般会选取 2~10 个与目标 QTL 紧密连锁的分子标记,同时对目标 QTL 进行检测^[45-48]。Peng 等认为,当分子标记与目标基因之间的距离小于 2 cM 时,单个标记检测的准确率可 ≥90%^[49]。李进波等研究结果表明,当分子标记与目标基因间的距离在 1.0 cM 以内时,可用单个分子标记进行辅助育种^[50],Peng 等利用分子标记选择目标基因的准确率均接近 100%。

2.1.1 分子标记辅助回交选择 在分子标记辅助回交选择中,除了需要对目标 QTL 进行选择外,还需要至少 200 个覆盖全基因组的标记进行背景选择,以保证所选植株与受体材料保持高度一致^[1]。目前,分子标记辅助回交选择已经在农作物抗病性育种^[51-54]、品质育种^[55-56]得到广泛应用。Zhao 等将玉米抗黑穗病基因经过多代回交逐渐渗入到 10 个玉米自交系中,最终 10 个自交系均抗黑穗病且其他大部分性状均未改变^[54]。Barloy 等利用 RAPD 和 SCAR 标记,将野生牧草中的 2 个谷物孢囊线虫抗性基因分别渗入到小麦中,并同时 2 个抗性基因聚合到小麦中^[57]。结果表明,聚合 2 个抗性基因的小麦抗性高于渗入单个抗性基因的小麦,但低于野生牧草抗性。朱映东等以纯和香型水稻与非香型巨胚水稻杂交, F₁ 自交得 F₂, 选取 F₂ 中巨胚特性种子继续种植自交。提取 F₂ 叶片基因组 DNA^[55], 采用可以识别香味基因特异位点的限制性内切酶选择纯和香型基因,连续自交到 F₈ 获得 3 个香型巨胚水稻品系。Zhou 等利用 SSR 与 RFLP 标记辅助选择,通过 3 次回交 1 次自交将恢复系中控制加工品质和口感的基因渗入到品质不理想的珍汕 97 中(常用作杂交稻母本),改良后的珍汕 97 直链淀粉含量下降,黏稠度与糊化温度均有所提高^[56]。然而,由于遗传图谱无法定位效应较小的 QTL,导致分子标记回交育种只能聚合效应较大的 QTL,对于效应较小的 QTL 及一些稀有变异则无法应用到回交育种中。Varshney 等利用分子标记辅助回交选择将栽培花生 GPBD4 中抗锈病 QTL(可解释 82.62% 的表型变异)分别导入 3 个易感锈病的材料中,通过 2~3 代的回交及自交,获得 81 个抗性提高的基因渐渗系^[58]。利用与抗锈病 QTL 连锁的 13 个分子标记对 43 个表现良好的基因渐渗系进行筛选,最终在 11 份基因渐渗系中发现目标 QTL。因此,这些分子标记可用于花生抗锈病的育种过程中。Chu 等以栽培花生 Tifguard(抗线虫)为轮回亲本,分别和栽培花生 Georgia-02C 和 Florida-07(油酸含量较高)杂交,利用分子标记辅助回交选择,将抗线虫性状与高油酸性状聚合在了一起^[59]。

2.1.2 分子标记辅助轮回选择 分子标记辅助轮回选择在 90 年代被提出^[60-61]。该方法在每一代选择中利用分子标记

对所有需要改良性状或重要性状进行选择,并将每一代中获得的优良单株互相杂交,然后杂交后代中通过上述标记选出优良单株。经过多代连续杂交与分子辅助选择,最终获得具备理想基因型的植株^[62-63]。这个方法适用于多个亲本优良基因的聚合。在没有 QTL 信息的情况下,可通过标记与性状的相关性进行选择。模拟研究发现,分子标记辅助轮回选择聚合有利基因的效率要比表型选择高出 3% ~ 20%^[64]。

2.2 全基因组选择

分子标记辅助育种在质量性状或由单个主效基因控制的数量性状改良中得到成功应用,已经成为辅助作物遗传育种改良的有力工具。但由于数量性状遗传基础的复杂性和 QTL 定位的限制性,导致分子标记辅助育种在复杂数量性状改良特别是由多个微效基因控制的数量性状改良中的应用受到诸多限制^[65]。全基因组选择有效避免了以上限制,它利用覆盖全基因组的分子标记同时关联主效和微效基因,并进行复杂性状育种值预测^[66]。全基因组选择简单来讲就是全基因组的分子标记辅助选择,不需要进行传统意义上标记和性状的关联。它是基于连锁不平衡理论,即假设标记与相邻的 QTL 处于连锁不平衡状态,利用覆盖全基因组的分子标记将染色体分成若干区段,然后通过模拟群体的基因型和表型估算每个染色体片段的效应值。由于标记和 QTL 的连锁不平衡,所以利用相同标记估算的染色体片段的效应在不同群体中是相同的,基于以上结果,研究人员可以利用相同标记进行其他个体或群体的育种值的预测。与传统育种相比,全基因组选择分子育种技术选择更准确,更高效,可使育种周期缩短 $Y_3 \sim Y_2$ 。以全基因组序列为指导的新一代育种技术效率更高,成本大大降低。全基因组选择分为 3 个步骤:(1) 基于参考群体全基因组分子标记与表型鉴定数据建立 BLUP (best linear unbiased prediction) 模型,估计每一标记的育种值;(2) 利用同样的分子标记对其他材料育种值进行预测;(3) 根据预测育种值进行选择^[67]。

相比于 MAS 依赖于 QTL 定位的准确性及其附近标记,仅选用少量分子标记进行预测少量的 QTL 效应,全基因组选择采用覆盖整个基因组的分子标记来捕获整个基因组上的变异并对育种值进行有效预测。显著性差异不明显的分子标记无法应用到 MAS 中,但可以用于全基因组选择对育种值进行估计,这点对由大量微效基因控制的数量性状尤为重要^[68]。对于植物育种来说,全基因组选择具有革命性意义。因为,全基因组选择可以减少表型鉴定的频率,并且其选择依赖于基因型而不是表型,因此选择结果更加可靠。全基因组选择育种的方法可以同时多个性状进行选择,并显著提高选择的效率,同时全基因组选择选用覆盖整个基因组的分子标记,这样可以将每个起作用基因的效应包括在内,增加了选择的准确性。模拟研究^[69]与实际研究^[70]均发现,在相同时间内,全基因组选择比传统的表型选择拥有更大收益。VanRaden 等首次在动物上进行全基因组选择^[70-71],随后,全基因组选择在其他动物上成功应用^[72]。Daetwyler 等阐述了如何利用全基因组选择分析人类遗传病^[73]。Bernardo 等展示了全基因组选择在植物中的应用前景^[74-76]。

目前,相关研究表明,科研人员利用育种数据(个体基因型,中等密度到高等密度的标记)可以在不同环境下对植物

的一些性状(如籽粒产量、生物学产量、抗病性、开花习性)进行预测,预测的准确性依赖于性状的遗传力、群体大小、标记数目、参考群体与检验群体之间的关系以及基因与环境互作^[77-81]。Ornella 等利用 5 个抗秆锈病小麦品种(Juchi、Pavon76、Muu、Kingbird、K - Nyangumi)分别和中度感病品种(PBW343)杂交构建 RIL 群体,然后在不同环境下检测群体对秆锈病的抗性,并利用 1 400 个 DArT (diversity arrays technology) 标记进行基因分型。利用 1 个群体对另 1 个群体进行预测的准确度相对较高,利用其他 4 个群体对另 1 个群体预测的准确性也较高。有趣的是,在群体结构存在的情况下,群体之间秆锈病预测的准确度依然很高,这说明性状结构对预测结果起着重要作用。一些复杂性状(如籽粒产量)预测结果不太理想。Windhausen 等收集了 255 份多样化的玉米系,通过 6 次试验将其分成 8 个群体,在其他群体标记效应估计基础之上分别对每个群体进行预测,预测能力几乎为零,结果说明群体结构对预测结果的准确率起重要作用^[82]。Zhang 等利用 30 000 多个 SNP 标记对 309 份大豆材料进行种子质量的全基因组关联分析,同时以该群体为模拟群体估算种子质量的育种值,并与分子标记辅助选择作比较发现,随着 SNP 标记的数目和模拟群体大小的变化,GS 和 MAS 的预测准确率分别是 0.75 ~ 0.87 和 0.62 ~ 0.75^[83]。表明 GS 拥有更加准确的预测值及更高的育种效率。Arruda 等分别利用全基因组选择和分子标记辅助选择对 273 份软红冬麦进行赤霉病抗性的预测和选择,MAS 统计模型以研究最多的 Fhb - 1 为基础进行选择,而全基因组选择模型利用分布于整个基因组的 19 992 个 SNPs 进行预测选择,分别利用普通最小二乘法和最佳无偏线性回归预测进行标记效应值的估算,结果表明,基因组选择预测精度在 0.4 ~ 0.9 之间,而 MAS 精度低于 0.3,由此可知,在小麦赤霉病抗性育种中,GS 为最优选择^[84]。

全基因组选择育种需要考虑以下几点因素:(1) 标记密度要有效地覆盖整个基因组,使得每个基因区域至少有 1 个标记与之处于连锁不平衡状态。覆盖整个基因组所需要的标记数取决于 LD 衰减程度^[85]。(2) 模拟群体数量要足够大并具有代表性,包括了待预测群体的亲本及与其亲缘关系很近的材料,同时模拟群体要进行多代模拟,从而保证了全基因组选择的高度准确性。(3) 估计标记效应。模拟全基因组选择模型统计方法的准确性取决于标记效应的强度。(4) 性状的遗传力。全基因组选择中遗传力低的性状需要大的模拟群体提高育种估计值的准确性,因为遗传力的降低会导致育种值估计准确性降低^[86]。

近年来,随着拟南芥、水稻、玉米、芝麻等植物全基因组测序的完成,植物基因组学研究已经由简单的质量性状转向复杂的数量性状,特别是大量丰富且廉价 SNP 标记的开发以及生物信息学的迅猛发展,利用全基因组选择方法进行植物数量性状改良成为遗传学研究的热点之一。Zhang 等利用不同标记密度对不同环境中的 19 个热带玉米双亲群体的不同性状进行全基因组选择精度的对比,发现对于中高度遗传力的简单性状,低标记密度(约 200 个标记)可以达到较高精度的全基因组选择,而对于复杂数量性状,则需要高密度标记才可以达到较高精度的全基因组选择预测;环境表型互作的全基因组预测模型精度高于非环境表型互作的全基因组预测模

型^[87]。Spindel 等利用 70 000 多个标记对 363 个水稻自交系进行全基因组关联分析,并进行基因组预测,发现株高和产量的预测精度在 0.31 ~ 0.34 之间,而开花期的预测精度高达 0.63^[88]。

3 展望

MAS(marker assisted selection)建立在 QTL 的精细定位基础之上,目前,仍有许多作物,尤其是自花授粉作物 QTL 定位精度不够,导致分子标记难以用于实际的育种过程当中。某些已经完成精细定位并成功用于实际育种当中的 QTLs 大部分都是主效 QTL,而一些微效 QTLs 则很少被利用。复杂性状受环境影响较大,在实际定位过程中往往会遗漏一些 QTLs,并且即便通过 MAS 将目标 QTLs 渗入受体中,在环境影响下,目标 QTLs 的表达可能会受影响。随着基因分型成本的快速降低和统计方法的快速发展,基因组标记密度越来越大,复杂性状的基因分析和操控更有效,具有更多优势的全基因组选择的育种方法应运而生,并快速应用于植物遗传育种。相信随着人们对植物基因组水平的认识不断深入,标记密度不断增加以及演算方法不断完善,全基因组选择将会成为作物遗传育种的最有效方法。

参考文献:

- [1] Xu Y B, Lu Y L, Xie C X, et al. Whole - genome strategies for marker - assisted plant breeding[J]. Molecular Breeding, 2012, 29 (4): 833 - 854.
- [2] Hill C B, Taylor J D, Edwards J, et al. Whole - genome mapping of agronomic and metabolic traits to identify novel quantitative trait loci in bread wheat grown in a water - limited environment[J]. Plant Physiology, 2013, 162(3): 1266 - 1281.
- [3] Tyrka M, Perovic D, Wardynska A, et al. A new diagnostic SSR marker for selection of the *Rym4/Rym5* locus in barley breeding[J]. Journal of Applied Genetics, 2008, 49(2): 127 - 134.
- [4] Cavanagh C R, Taylor J, Larroque O, et al. Sponge and dough bread making: genetic and phenotypic relationships with wheat quality traits[J]. Theoretical and Applied Genetics, 2010, 121(5): 815 - 828.
- [5] Tester M, Langridge P. Breeding technologies to increase crop production in a changing world[J]. Science, 2010, 327 (5967): 818 - 822.
- [6] Khedikar Y P, Gowda M C, Sarvamangala C, et al. A QTL study on late leaf spot and rust revealed one major QTL for molecular breeding for rust resistance in groundnut *Arachis hypogaea* L. [J]. Theoretical and Applied Genetics, 2010, 121(5): 971 - 984.
- [7] Tanksley S D, Nelson J C. Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines[J]. Theoretical and Applied Genetics, 1996, 92(2): 191 - 203.
- [8] Moncada M D P, Tovar E, Montoya J C, et al. A genetic linkage map of coffee (*Coffea arabica* L.) and QTL for yield, plant height, and bean size[J]. Tree Genetics & Genomes, 2015, 12(1): 1 - 17.
- [9] Lee S, Freewalt K R, Mchale L K, et al. A high - resolution genetic linkage map of soybean based on 357 recombinant inbred lines genotyped with BARCSoySNP6K[J]. Molecular Breeding, 2015, 35 (2): 58.
- [10] Yu J M, Holland J B, McMullen M D, et al. Genetic design and statistical power of nested association mapping in maize[J]. Genetics, 2008, 178(1): 539 - 551.
- [11] Kump K L, Bradbury P J, Wissner R J, et al. Genome - wide association study of quantitative resistance to southern leaf blight in the maize nested association[J]. Nature Genetic, 2011, 43 (2): 163 - 168.
- [12] Bergelson J, Roux F. Towards identifying genes underlying ecologically relevant traits in *Arabidopsis thaliana* [J]. Nature Reviews Genetics, 2010, 11(12): 867 - 879.
- [13] Kover P X, Valdar W, Trakalo J, et al. A multiparent advanced generation inter - cross to fine - map quantitative traits in *Arabidopsis thaliana* [J]. PLoS Genetics, 2009, 5(7): e1000551.
- [14] Atwell S, Huang Y S, Vilhjalmsen B J, et al. Genome - wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines[J]. Nature, 2010, 465 (7298): 627 - 631.
- [15] Zhao K Y, Tung C W, Eizenga G C, et al. Genome - wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa* [J]. Nature Communications, 2011, 2(1): 1 - 10.
- [16] Morris G P, Ramu P, Deshpande S P, et al. Population genomic and genome - wide association studies of agroclimatic traits in sorghum [J]. Proceedings of the National Academy of Sciences of the United States of America, 2013, 110(2): 453 - 458.
- [17] 于志远, 王伟威, 魏 嵘, 等. 利用关联分析方法挖掘自然群体中大豆油分和蛋白质含量相关 SSR 标记[J]. 大豆科学, 2015, 34(6): 977 - 981.
- [18] Spindel J, Wright M, Chen C, et al. Bridging the genotyping gap: using genotyping by sequencing (GBS) to add high - density SNP markers and new value to traditional bi - parental mapping and breeding populations[J]. Theoretical and Applied Genetics, 2013, 126(11): 2699 - 2716.
- [19] Zhou X J, Xia Y L, Ren X P, et al. Construction of a SNP - based genetic linkage map in cultivated peanut based on large scale marker development using next - generation double - digest restriction - site - associated DNA sequencing (ddRADseq) [J]. BMC Genomics, 2014, 15(351): 1 - 14.
- [20] Liu L Z, Qu C M, Wittkop B, et al. A high - density SNP map for accurate mapping of seed fibre QTL in *Brassica napus* L. [J]. PLoS One, 2013, 8(12): e83052.
- [21] Kim S, Plagnol V, Hu T T, et al. Recombination and linkage disequilibrium in *Arabidopsis thaliana* [J]. Nature Genetics, 2007, 39(9): 1151 - 1155.
- [22] Gore M A, Chia J M, Elshire R J, et al. A first - generation haplotype map of maize[J]. Science, 2009, 326 (5956): 1115 - 1117.
- [23] Chia J M, Song C, Bradbury P J, et al. Maize HapMap2 identifies extant variation from a genome in flux[J]. Nature Genetics, 2012, 44(7): 803 - 807.
- [24] 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.
- [25] 吴金华, 张西平, 胡言光, 等. 小麦抗白粉病相关基因 *GST* 克隆与表达[J]. 西北植物学报, 2013, 33(1): 34 - 38.
- [26] 刘新颖, 王晓杰, 薛 杰, 等. 小麦钙调素新亚型 *TaCaM5* 的克隆及表达分析[J]. 作物学报, 2010, 36(6): 953 - 960.

- [27] 王 坤, 王海燕, 刘大群. 叶锈菌与“*TcLr19*”小麦互作体系中 *PR1* 基因的克隆及分析[J]. 河北农业大学学报, 2012, 35(2): 1–6.
- [28] Wang C R, Sheng C, Yu S B. Functional markers developed from multiple loci in GS3 for fine marker – assisted selection of grain length in rice [J]. Theoretical and Applied Genetics, 2011, 122(5): 905–913.
- [29] Fu D L, Uauy C, Assaf D, et al. A kinase – START gene confers temperature – dependent resistance to wheat stripe rust[J]. Science Express, 2009, 323(5919): 1357–1360.
- [30] 宁丽华, 陈亭亭, 刘怀华, 等. 高直链淀粉玉米 *amylose – extender* 基因功能标记的开发及应用[J]. 分子植物育种, 2011, 9(2): 185–189.
- [31] Lopez Y, Nadaf H L, Smith O D, et al. Isolation and characterization of the $\Delta 12$ – fatty acid desaturase in peanut (*Arachis hypogaea* L.) and search for polymorphisms for the high oleate trait in Spanish marker – type lines[J]. Theoretical and Applied Genetics, 2000, 10(17): 1131–1138.
- [32] Chen Z, Wang M L, Barkley N A, et al. A simple allele – specific PCR assay for detecting *FAD2* alleles in both A and B genomes of the cultivated peanut for high – oleate trait selection [J]. Plant Molecular Biology Reporter, 2010, 28(3): 542–548.
- [33] Jaccoud D, Peng K, Feinstein D, et al. Diversity arrays: a solid state technology for sequence information independent genotyping [J]. Nucl Acids Res, 2001, 29(4): e25.
- [34] Akbari M, Wenzl P, Caig V, et al. Diversity arrays technology (DArT) for high – throughput profiling of the hexaploid wheat genome [J]. Theoretical and Applied Genetics, 2006, 113(8): 1409–1420.
- [35] Mace E S, Xia L, Jordan D R, et al. DArT markers: diversity analyses and mapping in Sorghum bicolor [J]. BMC Genomics, 2008, 9(1): 26.
- [36] Mace E S, Rami J F, Bouchet S, et al. A consensus genetic map of sorghum that integrates multiple component maps and high – throughput Diversity Array Technology (DArT) markers [J]. BMC Plant Biology, 2009, 9(1): 13.
- [37] Wenzl P, Li H, Carling J, et al. A high – density consensus map of barley linking DArT markers to SSR and RFLP loci and agronomic traits [J]. BMC Genomics, 2006, 7(1): 206–228.
- [38] Wand N, Fang L C, Xin H P, et al. Construction of a high – density genetic map for grape using next generation restriction – site associated DNA sequencing [J]. BMC Plant Biology, 2012, 12(1): 148.
- [39] Xiao J Z, Xia Y L, Ren X P, et al. Construction of a SNP – based genetic linkage map in cultivated peanut based on large scale marker development using next – generation double – digest restriction – site – associated DNA sequencing (ddRADseq) [J]. BMC Genomics, 2014, 15(1): 351.
- [40] Erena A, Patrick F B, Scott D H, et al. Genome wide association mapping of yield and yield components of spring wheat under contrasting moisture regimes [J]. Theoretical and Applied Genetics, 2014, 127(4): 791–807.
- [41] Su Z Q, Jin S J, Lu Y E, et al. Single nucleotide polymorphism tightly linked to a major QTL on chromosome 7A for both kernel length and kernel weight in wheat [J]. Molecular Breeding, 2016, 36(2): 15–25.
- [42] Myles S, Peiffer J, Brown P J, et al. Association mapping: critical considerations shift from genotyping to experimental design [J]. Plant Cell, 2009, 21(8): 2194–2202.
- [43] Lu Y, Zhang S H, Shah T, et al. Joint linkage – linkage disequilibrium mapping is a powerful approach to detecting quantitative trait loci underlying drought tolerance in maize [J]. Proceedings of the National Academy of Sciences of the United States of America, 2010, 107(45): 19585–19590.
- [44] Li X P, Zhou Z J, Ding J Q, et al. Combined linkage and association mapping reveals QTL and candidate genes for plant and ear height in maize [J]. Front Plant Sci, 2016, 7: 833.
- [45] Shamsudin N A, Swamy B P, Ratnam W, et al. Marker assisted pyramiding of drought yield QTLs into a popular Malaysian rice cultivar, MR219 [J]. BMC Genetics, 2016, 17(1): 30.
- [46] Salameh A, Buerstmayr M, Steiner B, et al. Effects of introgression of two QTL for fusarium head blight resistance from Asian spring wheat by marker – assisted backcrossing into European winter wheat on fusarium head blight resistance, yield and quality traits [J]. Molecular Breeding, 2011, 28(4): 485–494.
- [47] Yang L Q, Wang W, Peng Y, et al. Marker – assisted selection for pyramiding the waxy and opaque – 16 genes in maize using cross and backcross schemes [J]. Molecular Breeding, 2013, 31(4): 767–775.
- [48] Hao X M, Li X W, Yang X H, et al. Transferring a major QTL for oil content using marker – assisted backcrossing into an elite hybrid to increase the oil content in maize [J]. Molecular Breeding, 2014, 34(2): 739–748.
- [49] Peng J H, Fahima T, Roder M S, et al. Microsatellite high – density mapping of the stripe rust resistance gene YrH52 region on chromosome 1B and evaluation of its marker – assisted selection in the F – 2 generation in wild emmer wheat [J]. New Phytologist, 2000, 146(1): 141–154.
- [50] 李进波, 王春连, 夏明元, 等. 分子标记辅助选择 *Xa23* 基因培育杂交稻抗白叶枯病恢复系 [J]. 作物学报, 2006, 32(10): 1423–1429.
- [51] Vida G, Gál M, Uhrin A, et al. Molecular markers for the identification of resistance genes and marker – assisted selection in breeding wheat for leaf rust resistance [J]. Euphytica, 2009, 170(1/2): 67–76.
- [52] 董 娜, 张亚娟, 张军刚, 等. 分子标记辅助小麦抗白粉病基因 *Pm21* 和 *Pm13* 聚合育种 [J]. 麦类作物学报, 2014, 34(12): 1639–1644.
- [53] Singh A K, Singh V K, Singh A, et al. Introgression of multiple disease resistance into a maintainer of Basmati rice CMS line by marker assisted backcross breeding [J]. Euphytica, 2015, 203(1): 97–107.
- [54] Zhao X R, Tan G Q, Xing Y E, et al. Marker – assisted introgression of *qHRSRI* to improve maize resistance to head smut [J]. Molecular Breeding, 2012, 30(2): 1077–1088.
- [55] 朱映东, 时亚琼, 周锋利, 等. 分子标记辅助选育香型巨胚水稻 [J]. 上海师范大学学报(自然科学版), 2013, 42(6): 623–628.
- [56] Zhou P H, Tan Y F, He Y Q, et al. Simultaneous improvement for four quality traits of Zhenshan 97, an elite parent of hybrid rice, by

- molecular marker – assisted selection[J]. Theoretical and Applied Genetics,2003,106(2):326–331.
- [57] Barloy D, Lemoine J, Abelard P, et al. Marker – assisted pyramiding of two cereal cyst nematode resistance genes from *Aegilops variabilis* in wheat[J]. Molecular Breeding,2007,20(1):31–40.
- [58] Varshney R K, Manish K P, Pasupuleti Janila, et al. Marker assisted introgression of a QTL region to improve rustresistance in three elite and popular varieties of peanut (*Arachis hypogaea* L.) [J]. Theoretical and Applied Genetics,2014,127(8):1771–1781.
- [59] Chu Y, Wu C L, Holbrook C C, et al. Marker – assisted selection to pyramid nematode resistance and the high oleic trait in peanut[J]. The Plant Genome,2011,4(2):110–117.
- [60] Edwards M, Johnson L. Proceedings of symposium on analysis of molecular marker data [M]. Corvallis: American Society of Horticultural Science and Crop Science Society of America, 1994: 33–40.
- [61] Stam P. Proceedings of the 9th meeting of EUCARPIA section on biometrics in plant breeding (1994) centre for plant breeding and reproduction research[M]. Wageningen; the Netherland, 1995:32–44.
- [62] Peleman J D, vander V J. Breeding by design[J]. Trends in Plant Science,2003,8(7):330–334.
- [63] van Berloo R, Stam P. Simultaneous marker – assisted selection for multiple traits in autogamous crops [J]. Theoretical and Applied Genetics,2001,102(6):1107–1112.
- [64] Bernardo R. Molecular markers and selection for complex traits in plants; learning from the last 20 years[J]. Crop Science,2008,48(5):1649–1664.
- [65] Meuwissen T H, Hayes B J, Goddard M E. Prediction of total genetic value using genome – wide dense marker maps[J]. Genetics,2001,157(4):1819–1829.
- [66] 吴永升, 邵俊明, 周瑞阳, 等. 植物数量性状全基因组选择研究进展[J]. 西南农业学报,2012,25(4):1510–1514.
- [67] Rutkoski J E, Heffner E L, Sorrells M E. Genomic selection for durable stem rust resistance in wheat [J]. Euphytica, 2011, 179(1):161–173.
- [68] Wong C K, Bernardo R. Genomewide selection in oil palm: increasing selection gain per unit time and cost with small populations[J]. Theoretical and Applied Genetics,2008,116(6):815–824.
- [69] Lorenzana R E, Bernardo R. Accuracy of genotypic value predictions for marker – based selection in biparental plant populations [J]. Theoretical and Applied Genetics,2009,120(1):151–161.
- [70] VanRaden P M. Genomic measures of relationship and inbreeding [J]. Interbull Bulletin,2007,37:33–36.
- [71] VanRaden P M, Van Tassell C P, Wiggins G R, et al. Invited review: reliability of genomic predictions for North American Holstein bulls[J]. Dairy Science,2008,92(1):16–24.
- [72] Hayes B J, Bowman P J, Chamberlain A J, et al. Invited review: genomic selection in dairy cattle: progress and challenges [J]. Journal of Dairy Science,2009,92(2):433–443.
- [73] Daetwyler H D, Villanueva B, Woolliams J A. Accuracy of predicting the genetic risk of disease using a genome – wide approach[J]. PLoS One,2008,3(10):e3395.
- [74] Bernardo R, Yu Y. Prospects for genome – wide selection for quantitative traits in maize[J]. Crop Science,2007,47(3):1082–1090.
- [75] Lorenzana R E, Bernardo R. Accuracy of genotypic value predictions for marker – based selection in biparental plant populations [J]. Theoretical and Applied Genetics,2009,120(1):151–161.
- [76] Heffner E L, Jannink J L, Sorrells M. Genomic selection accuracy using multifamily prediction models in a wheat breeding program [J]. The Plant Genome,2011,4:65–75.
- [77] de los Campos G, Naya H, Gianola D, et al. Predicting quantitative traits with regression models for dense molecular markers and pedigree[J]. Genetics,2009,182(1):375–385.
- [78] Crossa J, de los Campos G, Perez P, et al. Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers[J]. Genetics,2010,186(2):713–724.
- [79] Pérez P, de los Campos G, Crossa J, et al. Genomic – enabled prediction based on molecular markers and pedigree using the Bayesian Linear Regression Package in R[J]. The Plant Genome, 2010,3(2):106–116.
- [80] Pérez – Rpdriquez P, Gianola D, González – Camacho J M, et al. Comparison between linear and non – parametric regression models for genome – enabled prediction in wheat [J]. Genes Genomes Genet,2012,2(12):1595–1605.
- [81] Ornella L, Sukhwinder – Singh S, Perez P, et al. Genomic prediction of genetic values for resistance to wheat rusts [J]. The Plant Genome,2012,5(3):136–148.
- [82] Windhausen V S, Atlin G N, Crossa J, et al. Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments [J]. Genes Genomes Genet,2012,2(11):1427–1436.
- [83] Zhang J P, Song Q J, Cregan P B, et al. Genome – wide association study, genomic prediction and marker – assisted selection for seed weight in soybean (*Glycine max*) [J]. Theoretical and Applied Genetics,2016,129(1):117–130.
- [84] Arruda M P, Lipka A E, Brown P J, et al. Comparing genomic selection and marker – assisted selection for Fusarium head blight resistance in wheat (*Triticum aestivum* L.) [J]. Molecular Breeding,2016,36(7):84.
- [85] Flint – Garcia S, Thornsberry J M, Buckler E S. Structure of linkage disequilibrium in plants[J]. Annual Review of Plant Biology,2003,54(4):357–374.
- [86] Hayes B J, Bowman P J, Chamberlain A J, et al. Genomic selection in dairy cattle: progress and challenges [J]. Journal of Dairy Science,2009,92(2):433–443.
- [87] Zhang X, Pérez – Rodríguez P, Semagn K, et al. Genomic prediction in biparental tropical maize populations in water – stressed and well – watered environments using low – density and GBS SNPs[J]. Heredity,2015,114(3):291–299.
- [88] Spindel J, Begum H, Akdemir D, et al. Genomic selection and association mapping in rice (*Oryza sativa*): effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines [J]. PLoS Genetics, 2015, 11(2):e1004982.