

高通量测序技术在转基因作物分子特征评价中的应用： 机遇和挑战

赵圣博, 罗军玲, 曾新华, 袁荣, 吴刚*, 闫晓红*

(中国农业科学院油料作物研究所, 农业农村部油料作物生物学与遗传育种重点实验室, 农业农村部转基因植物环境安全监督检验测试中心, 湖北 武汉, 430062)

摘要: 转基因作物的分子特征为转基因的安全性评估和后续监测奠定了基础。由于靶点的特异性, 以聚合酶链反应(PCR)为基础的传统分子特征鉴定方法不能全面检测外来基因的非预期插入、载体骨架残留及未知的转基因事件。更多新的育种技术产生的基因修饰作物也对传统分子特征鉴定方法提出了挑战。下一代测序技术, 可以克服基于PCR的方法的某些局限性, 提供快速、全面分子特征数据。本文综述了T-DNA整合的复杂机制, 传统分子特征鉴定方法的局限性, 高通量测序方法在转基因分子特征安全评价中的应用、遇到的挑战及应用策略, 为转基因的安全评价工作提供借鉴。

关键词: 下一代测序技术; 转基因作物; 分子特征; 安全评价

中图分类号: Q78 **文献标识码:** A **文章编号:** 1007-9084(2021)01-0031-09

Application of high-throughput sequencing technology in evaluation of molecular characteristics of transgenic crops: opportunities and challenges

ZHAO Sheng-bo, LUO Jun-ling, ZENG Xin-hua, YUAN Rong, WU Gang*, YAN Xiao-hong*

(Key Laboratory of Biology and Geneics Improvement of Oil Crops, Ministry of Agriculture and Rural Affairs, Oil Crops Research Institute, Chinese Academy of Agricultural Sciences, Wuhan 430062, China)

Abstract: Molecular characterization lays a foundation for safety assessment and subsequent monitoring of transgenic plants. Due to the target-specific nature, conventional polymerase chain reaction (PCR)-based methods cannot comprehensively detect unintended gene insertions, let alone detect unknown GM events. More and more newly developed crops by new plant breeding technology challenged classical PCR-based methods. Next-generation sequencing is expected overcome these limitations, providing rapid and comprehensive molecular feature data. This article summarized complex mechanism of T-DNA integration, limitations of traditional molecular identification methods, and application of high-throughput sequencing methods in safety evaluation of genetically modified molecular characteristics. It also discussed challenges encountered and application strategies, which might provide alternatives for safety evaluation of genetic modification.

Key words: next-generation sequencing (NGS); transgenic crops; molecular characterization; safety assessment

使用人工选择和杂交可以改良驯化动植物, 满足人类需求。但是, 驯化过程限制了现代农作物和动物的遗传多样性, 使其易受生物和非生物胁迫的影响^[1]。通过农杆菌介导的转化和基因枪等遗传修

饰技术, 将重组DNA随机插入基因组, 可对作物进行遗传改良^[2]。自从1990年代中期美国首次引入转基因(genetically modified, GM)作物以来, 转基因作物已被世界许多国家广泛采用^[3]。到目前为止, 全

收稿日期: 2020-11-09

基金项目: 国家自然科学基金面上项目(31671733); 转基因生物新品种培育重大专项(2018ZX0801105B)

作者简介: 赵圣博(1989-), 男, 博士, 主要从事基因工程与转基因安全评价研究, E-mail: shengbo_ab@163.com

* 通讯作者: 吴刚(1976-), 男, 研究员, 博士生导师, 主要从事基因工程与转基因安全评价研究, E-mail: wugang@caas.cn;

闫晓红(1975-), 女, 副研究员, 硕士生导师, 主要从事基因工程与转基因安全评价研究, E-mail: yanxiaohong@caas.cn

球转基因作物的累计种植面积已超过20亿公顷^[4],越来越多的转基因植物衍生的食品和饲料进入供应链。此外,越来越多的基因或调控元件被转移到作物基因组中以改善农艺性状^[5]。但是,与转基因作物的生产有关的几个问题仍在争论中,这些问题包括潜在的生态影响,对食品安全的关注以及作物的遗传稳定性^[6,7]。国际公认,转基因作物的安全性评估在产品开发阶段及投放市场之前至关重要,对于具有优良性状的转基因品系只有进行广泛的安全测试和综合分析,才可以商业化。分子特征是转基因作物特异性检测核心内容,是转基因作物安全评价和政府监管的基础。转基因作物的分子特征通常包括基因组特征(例如插入位点、侧翼序列和拷贝数),转录组特征,蛋白质组学特征和代谢组学特征,这些特征信息对于转基因作物的研发、安全评估、检测和监测至关重要。根据分子特征,可以识别转基因的预期和非预期作用^[8,9],促进转基因作物及其衍生产品——食品/饲料的质量和安全风险评估^[8]。

T-DNA插入位点和相关侧翼序列的详细分子特征在转基因作物的安全性评估和追踪单个转基因事件中尤其重要。根据T-DNA序列信息,可以通过基于PCR的方法分离和鉴定T-DNA的确定整合位点及其旁侧序列^[10-13]。传统转基因植物的T-DNA侧翼序列是通过基于PCR的方法鉴定的,这些方法包括热不对称交错PCR(thermal asymmetric interlaced PCR, TAIL-PCR)^[13]、接头连接PCR(restriction site extension PCR, RSE-PCR)^[14]、反向PCR(inverse PCR, IPCR)^[15]或限制性位点延伸PCR^[11]、基因组步移(genome walking)^[11,13]。低拷贝插入是从转基因品系中选择最佳转基因事件最有用的分子特征,Southern印迹常用于载体插入和整合的拷贝数分析^[16]。

尽管基于PCR的方法,已经在转基因玉米^[17]、大豆^[18]、棉花^[19]和苜蓿^[20]中成功鉴定了多个T-DNA侧翼序列,然而,这些基于PCR的方法存在一定局限性。如果在外源插入序列中发生缺失、修饰或重排,基于PCR的方法不足以识别所有插入位点及其侧翼序列。另一方面,如果宿主基因组复杂,存在高度重复序列,可能会增加侧翼序列的鉴定难度^[5,21]。例如,与大豆、棉花和水稻相比,玉米由于基因组比较大,序列复杂,富含高度重复序列,使用基于PCR的方法很难鉴定玉米转基因品系插入基因的侧翼序列^[22]。

下一代测序(next-generation sequencing, NGS)技术可用于探测基因组结构变异和重排^[23],具有高通量,不需要大量DNA,省时省力的优势^[24]。随着NGS技术的快速发展和成本降低,可以借助于NGS技术,以可接受的成本,相对快速鉴定转基因农作物的插入位置,拷贝数,完整性和稳定性等分子特征^[25]。特别是NGS可以使用市售试剂盒,以标准流程执行大多数步骤,获得高通量数据,因此减少了实验的成本和工作量^[16]。基于NGS的分子特征鉴定可以克服基于PCR的方法的某些局限性,包括需要大量DNA,人工干预以及无法识别复杂遗传变化的情况^[24],迄今为止,基于NGS技术已成功在拟南芥^[26]、水稻^[27]、大豆^[21]和玉米^[28]等转基因植物中鉴定出许多外源基因的侧翼序列。此外,NGS可以进一步揭示包括SNP和小的InDel在内的核酸序列变异,可检测到小的序列修饰^[24],甚至可以通过与毒素基因,毒素靶标基因,过敏原和抗营养因子数据库中的序列进行比对,查找潜在的相似性,将NGS鉴定的准确序列信息直接用于评估转基因植物的潜在毒性或致敏性。本文将对T-DNA整合的复杂性,传统分子特征鉴定方法及其局限性,NGS在转基因分子特征鉴定中的应用、遇到的挑战及应用策略进行综述,为转基因的安全评价工作提供借鉴。

1 复杂的T-DNA整合机制导致不同的外源DNA插入

根癌农杆菌一直是植物基因工程中的主力军。DNA双链断裂末端(double strand break, DSB)修复机制是实现T-DNA整合的主要途径。在农杆菌转化过程中,T-DNA可以利用宿主DNA修复机制,随机或诱导地(例如通过核酸酶活性)插入染色体双链断裂的位点。在许多情况下,T-DNA并非作为完整单元简单地插入植物基因组,根癌农杆菌介导的T-DNA整合会涉及到:超出LB和RB边界的载体骨架经常与T-DNA一起转移^[29],其它细菌DNA也可能进入植物细胞,包括质粒DNA^[30,31]和染色体DNA^[32-34];T-DNA整合到植物基因组中遵循“非法”模型,即整合是随机的,不是由序列同一性而是由边界与基因组之间的序列微观相似性指导的^[35,36];填充DNA的存在,即通常在T-DNA的串联重复拷贝之间或边界与gDNA之间经常发现的未知来源的DNA序列,表明了DSB修复机制在T-DNA整合中的作用^[37];转化方法以及细胞类型和/或发育阶段,染色质修饰也是T-DNA的整合影响因素。T-DNA

整合进基因组复杂机制使插入位点、旁侧序列及拷贝数的识别复杂化。

2 传统分子特征鉴定方法

从科学和监管的角度,了解基因插入引起的结构变异,包括插入位点,拷贝数和潜在的骨架插入,以及宿主基因组表观遗传变化,对转基因作物的安全性评估和监测起着至关重要的作用。

2.1 基于PCR的插入位点和旁侧序列的分子特征鉴定方法

转基因作物中外源T-DNA插入位点、旁侧序列及拷贝数的鉴定,是转基因作物分子特征鉴定的重要内容,为转基因事件身份鉴定提供确凿证据。基于PCR的方法,例如热不对称交错PCR^[38]、连接介导PCR (ligation-mediated PCR, LM-PCR)^[39]和反向PCR^[15]用于基因组插入位点及旁侧序列分析。

基于PCR的方法均依赖于外源插入序列的已知信息,无法检测到转基因作物的未被记录的分子特征,因此在许多情况下,T-DNA插入信息常常被低估了^[40]。以商业化的大豆事件GTS40-3-2为例,最初基于PCR的方法仅鉴定了宿主基因组中*EPSPS* (5-enolpyruvylshikimate 3-phosphate synthase, 5-烯醇丙酮酸莽草酸3-磷酸合酶)基因表达框的一个插入拷贝^[41]。但是,进一步的研究显示存在另外两个*CP4*、*EPSPS*基因的部分片段的插入(长度分别为72和250 bp),以及3'-NOS连接处非预期DNA重排,导致GTS40-3-2的分子特征被修订数次^[41, 42]。

2.2 基于DNA杂交和PCR的插入拷贝数测定方法

传统上,Southern印迹通过转基因植株DNA插入序列同源的核酸特异性探针杂交来检测拷贝数,目前已广泛用于转基因事件的分子特征鉴定,以确定转基因的存在和拷贝数^[43, 44]。Southern印迹分析涉及探针的正确设计和限制性内切酶的选择,并且需要预先知道基因或插入序列的信息,非常耗时且费力,其结果可能无法准确反映重排GM拷贝的存在,并且受限于限制性内切酶的选择^[45, 46]。

实时定量PCR (real time quantitative PCR, qPCR)通过与内源参考基因相比,可以对转基因拷贝数进行相对定量,由于qPCR简单、准确、通量高、重复性好和低成本,因此被确定为转基因定量分析的标准检测方法。和Southern印迹相比,qPCR更多应用在转基因作物鉴定和定量分析中,鲜有关于其在转基因拷贝数确定中的应用报道^[46, 47]。另外,qPCR通常只针对基因序列,有时无法检测到GM插入的

截短或突变拷贝的存在^[47]。

数字PCR (digital PCR, dPCR)技术,无需绘制标准曲线,可对样品中基因拷贝数进行绝对定量。鉴于dPCR绝对定量的性质,和相对定量的qPCR相比,从dPCR分析得出的拷贝数值更可靠和准确。最近的一项使用微滴 (droplet) dPCR对12个欧盟授权转基因玉米事件进行多重定量分析的报告显示,与qPCR相比,其通量和成本效益更高^[48],这表明微滴dPCR在确定转基因拷贝数中的巨大潜力。但与qPCR相似,尽管dPCR已经在几种转基因作物中进行了拷贝数测定^[49, 50],但每个数字PCR平台存在测量偏差和不确定性,需要对dPCR平台在拷贝数测定方面进行验证或标准化。

3 高通量测序方法

3.1 序列捕获技术

基于NGS技术的序列捕获技术 (sequence capture technology),首先筛选出含有插入片段的基因组DNA片段,然后使用Illumina测序来确定插入片段的位置。序列捕获技术是一种快速,经济高效的方法,可用于鉴定插入侧翼序列和插入结构^[26]。Inagaki等^[26]用此方法分析了29个拟南芥转基因品系,并成功确定了22个品系中T-DNA的插入位点,其中4个品系具有多个插入位点。由于该方法筛选的DNA片段大小范围250~450 bp,并且受限于探针设计,序列捕获技术难以鉴定同一基因座位含有两个或多个拷贝的转基因品系的侧翼序列和外源序列的分子特征结构,并且该方法难以适用于经过复杂加工的食品和饲料样品。到目前为止,没有一种可用的富集方法能够以非常敏感的方式富集足够长的已知插入片段的上下游DNA片段。此外,为了获得可靠的测序结果,需要高而均匀的富集序列覆盖范围,这对于基因组中的某些区域(尤其是那些富含GC的区域)存在挑战^[51]。

3.2 全基因组测序技术

基于NGS技术的全基因组测序 (whole genome sequencing, WGS)技术不仅能提供T-DNA插入片段及其侧翼DNA的信息,而且还能提供全基因组序列的其它信息。转基因植物基因组的突变不仅来源于转化过程本身,也可能来自体细胞自发突变,及转基因植物的组织培养,再生和繁殖过程。WGS可用于全面检测宿主基因组中的外来DNA残留及变异^[23, 24, 52]。Schouten等^[53]人应用NGS筛选了全基因组的突变,其中一些起源于转化过程本身。然而,在

五种转基因植物的基因组中仅鉴定出一些小突变,且这些突变与T-DNA插入片段的位置或数量无关。目前已经开发出了大规模提交序列数据的指南,用于对遗传修饰生物(GMO)的安全性评估^[54]。

3.3 高通量测序技术在分子特征鉴定中的优势

3.3.1 非预期的外源基因插入 NGS技术基于以下过程鉴定外源基因。首先,将所有测序reads分别比对到参考基因组和T-DNA/载体序列;其次,根据比对结果将测序reads分为三个亚组:1)仅比对到参考基因组的测序reads,2)仅比对到载体序列的测序reads,和3)同时比对到参考基因组和载体序列的测序reads。同时比对到参考序列和载体序列的测序reads用于鉴定T-DNA插入。与Southern印迹法相比,NGS检测转基因植物中T-DNA插入片段更加灵敏。NGS已被证明是检测转基因植物中T-DNA插入片段的一种有效的方法,同时也可以检测转基因植物中整合的非预期的载体骨架片段。在水稻中,Yang等^[55]利用NGS研究三个不同独立的转基因株系,与PCR和Southern杂交结果相比,在其中的两个转基因株系中,检测到了非预期插入。NGS可以发现非预期T-DNA插入,这不仅为对转基因事件的风险评估提供了至关重要的直接分子证据,还证明了NGS在该领域的高分辨率。

3.3.2 NGS用于探测碎片 转基因植物中的外源DNA碎片源自转化过程中使用的T-DNA。Schouten等^[53]人应用NGS鉴定出一个额外的50 bp微小插入片段,该碎片来自T-DNA中的*gfp*基因,与两个边界的距离都超过2 kb,没有任何T-DNA边界序列,该插入片段以前无法通过常规PCR或Southern印迹检测到。由于碎片不是一个完整的基因,50 bp插入片段不会产生功能性肽并产生相应的表型效应。目前尚不知道是什么原因导致T-DNA碎片的插入,是否与所用的浸花转化方法有关,并且很难估计转基因植物中碎片的发生频率。

3.3.3 NGS用于未知的转基因事件 未授权转基因作物(unauthorized GM, UGM)包括经批准但滥用的转基因作物(异步授权作物:在一个辖区批准,另一个辖区未批准),未批准但被释放(混杂)或未经批准但逃脱了田间试验(特别是在种子生产阶段)的转基因作物。关于UGM事件的报道越来越多,这些事件极大地影响了国际贸易和当地经济^[56]。因此需要在转基因作物的同步授权方面开展国际合作,建立交流渠道以在田间试验中分享有关转基因作物的信息。除此外,还需要开发高通量和高效的

UGM作物的检测筛选方法。针对UGM的分子特征鉴定方法,我们应该从高度依赖分子特征和靶标特异的追溯策略转变为更通用,更广泛的全面筛选策略,并与生物信息学相集成,关注同步授权和共享数据库的开发。Yang等^[40]使用NGS技术,开发了不依赖于插入序列和载体序列先验知识的TranSeq方法,用于检测UGM作物未知的插入序列、旁侧序列和插入拷贝数^[40],证明NGS在鉴定UGM作物方面的巨大潜力。在同一年,基于NGS全基因组测序方法对欧盟未经授权的转基因水稻事件LLRice62进行快速分子特征研究^[5]。两项研究都证明,基于NGS的方法可以在一个实验中,无论是否具有先验插入信息,都可以在UGM作物找出DNA样品中存在的单个、多个或复杂插入。

3.3.4 NGS用于复杂基因组的T-DNA插入的鉴定 NGS特别适合解决由于T-DNA整合到基因组DNA的重复区域而引起的外源序列难以鉴定问题^[55,57]。NGS方法已被证明是发现不同动植物物种中基因融合、序列重排、DNA插入和结构变异的强大工具^[16,23,58-60]。NGS甚至可以检测T-DNA内外的微小DNA重排(插入/缺失和倒位)。在转基因水稻事件TT51-1和T1c-19中,传统的基于PCR的方法仅显示了一个完整插入,NGS则可揭示额外的T-DNA插入^[55]。在转基因玉米MIR152中,NGS鉴定了3'端的279 bp玉米插入序列,这是以前基于PCR的基因组步移无法实现的^[57]。

4 高通量测序用于转基因作物分子特征鉴定的挑战及策略

4.1 高通量测序深度、测序读长及污染问题

足够数量的高质量原始数据是利用NGS技术全面分析研究转基因作物分子特征的前提。NGS需要一定的覆盖深度才能组装包含整个T-DNA和侧翼区域的重叠群^[24,57]。目前尚未确定全面检测插入片段是否存在所需要的测序深度,但理论上测序深度越高,对GM事件进行精确分子特征鉴定的效果就越好^[61,62]。然而,测序深度越高,成本也越高;因此,需要确定合适的测序深度以用于GM事件的分子特征鉴定^[25,63]。T-DNA的整合如果发生在宿主基因组的重复区域内,对于在比对步骤中检测插入位点和旁侧序列存在挑战^[25]。NGS在转基因作物分子特征鉴定中的大多数应用中,假定的插入位点是基于结合测序读长鉴定的,即一端与载体序列匹配,另一端与参考基因组匹配。短的测序读长,对

含有复杂基因组及富含重复序列的宿主而言,确定插入位点存在挑战。通过使用 Pacific BioSciences sequencing platform (PacBio) 测序平台或 DNA 单分子实时测序(读取长度 10~20 kb),产生更长的测序读长,可以克服此限制^[64]。对于具有复杂基因组或插入序列含有复杂结构的某些转基因农作物而言,长的读长测序方法更可靠。因此,在鉴定转基因作物的侧翼序列时,应根据其基因组特征和插入序列的内部结构灵活选择不同的方法。NGS 的灵敏度高,在样品收集、DNA 提取、PCR 扩增、上机测序过程中均有可能导致污染,在实验过程中越早剔除污染越好,可以通过严格样品收集过程、NGS 实验室进行分区、全自动化操作、生物信息学等方法去除污染。另外测序读长的增加,也有助于排除污染。

4.2 高通量测序在转基因作物分子特征鉴定中的分析流程

鉴于其高通量、不断降低的成本以及各种生物信息学工具的开发,NGS 可以促进转基因生物分子特征解析。但是,当前没有 NGS 用于转基因作物分子特征鉴定的标准化工作流程、必要步骤及模型。造成这种情况的主要原因是,NGS 数据分析需要对生物信息学有深刻的理解,不同测序平台产生不同测序数据,需要不同的数据处理流程。建议制定高通量测序用于转基因作物分子特征鉴定的标准流程,包括测序深度要求、生物信息分析方法及步骤。

4.3 高通量测序应用于新育种技术产品的鉴定

GM 转基因生物的定义是生物本身遗传物质发生了改变,这种改变不是通过交配和/或自然重组自然发生的(第 2001/18 / EC 号指令第 2 条)^[65],而是通过基因工程进行了改变。此外,通过遗传修饰技术产生的生物体,即使基因组中不存在外源 DNA 或 RNA,也被归类为 GM^[65,66]。几种新育种技术(new breeding techniques, NBTs)的出现正在彻底改变农业生产过程,目前已经开发了几种生物技术,例如基因组编辑技术,可有效地生产新型农作物和牲畜^[67,68]。目前,NGS 已经成功用于基因编辑植物的脱靶检测,但 NGS 应用于基因编辑植物的脱靶检测往往需要设置严格的对照,来排除自发突变、组织培养导致的突变^[69-73]。同样在设置严格的对照及足够的测序深度前提下,NGS 也是可以辨别基因编辑作物与已有品种的。另外,只要在生产过程中使用了外来 DNA, NBT 产品就可能存在外来 DNA 的意外残留。因此与基因组编辑技术的临床应用要求对目

标突变进行全基因组范围的研究^[74]不同,在农产品方面,重点应放在检测外来 DNA 的意外残留,仅应在宿主基因组中检查在利用 NBT 技术育种过程中使用的载体序列是否存在。例如,为了应对生物技术的新时代,日本环境部门和日本其它相关部委就其 NBT 法规政策发表了声明,其中要求对意外残留的外部 DNA 进行评估([https://www. env. go. jp/press/2_2_%20genome%20editing_En. pdf](https://www.env.go.jp/press/2_2_%20genome%20editing_En.pdf))。由于预计 NBT 技术不仅在发达国家而且在发展中国家都将发挥重要作用,因此,针对 NBT 产品的安全评估方法必须具有成本效益、简单易行、且不牺牲准确性。当前的大规模 DNA 测序方法在检测外来 DNA 片段的意外残留方面起关键作用,最适合此类评估。欧盟关于序列数据提交的指南要求、程序及其用法进行了详细说明^[54]。但 NGS 数据分析的结果因所使用的生物信息学分析程序而异^[75],需要对其分析流程标准化。

4.4 多种组学联合应用

与基于基因组高通量测序技术的方法在转基因作物分子特征鉴定中的应用相比,在过去的 10 年中,使用其它组学工具对转基因作物进行鉴定的研究进展甚微。其主要原因不是因为缺乏成熟的平台^[76],而是按照实质等同原则,应用此类分析方法进行转基因作物分子特征鉴定存在技术难度且成本高昂^[77-79]。与基因组不同,转录组、蛋白质组和代谢组的变化不仅受到遗传因素(包括遗传修饰)的影响,还受到各种内部(例如发育阶段、组织和细胞类型)和外部(例如环境)因素的影响。在许多情况下,外部因素引起的变化可能大于仅由基因改造引起的变化^[78,80]。利用转录组、蛋白质组和代谢组学研究转基因作物分子特征将引入自然变异的概念^[81-83]。根据实质等同原则,如果转基因作物的变化与非转基因作物的变化相当^[9],则该转基因作物至少可以视为与非转基因作物一样安全。但是,如果转基因作物的变化与非转基因作物的变化不相当,我们是否裁定这种转基因作物不安全? 答案显然不是简单的是或否,因为在任何情况下都不能忽略自然突变因素。先前的组学研究表明,自然变异可以引起转基因作物的许多变化^[81-84]。为了提供令人信服的比较结果,以正确理解实质等同原则的本质,在应用大规模的比较组学研究转基因作物时,除了选择正确的比较参数,确保所有作物均在相同条件下以相同的发育阶段生长外,还必须考虑自然变异。

组学研究中观察到的变化也可能来自同一组学研究的平台。由于单一平台的局限性,通常建议将平台组合应用于转基因作物的分子特征,以更好地覆盖RNA、蛋白质和代谢物的整体变化(包括预期和非预期)。虽然已经报道了在不同的转基因作物上使用单个组学方法进行的研究,但是在基因组、转录组、蛋白质组和代谢组水平上同时研究给定转基因作物的研究仍然很少。为了降低单个研究机构进行广泛且重复性的组学研究的高昂成本,需要公共研究中心或第三方服务中心介入,并在未来与研究机构合作,以利用经过验证核实的组学平台数据,建立作物自然变异数据库。

5 展望

转基因作物的分子特征提供了T-DNA插入结构、表达信息以及预期性状的稳定性信息,对于转基因作物的研发、安全评估、检测和监测至关重要。根据分子特征,可以轻松地识别转基因的预期和非预期作用,极大地促进后续的风险评估工作。基于PCR的方法在DNA层次的分子特征鉴定中起着不可或缺的作用。但由于越来越多的转基因作物以及源自新植物育种技术(NPBT)的新型基因工程作物正在准备商业化并在不久的将来释放到环境和/或市场中,这对传统分子特征鉴定提出了巨大挑战。为了向监管机构、零售商和消费者提供经授权的和未经授权的转基因(UGM)作物的详细分子特征,迫切需要开发准确、全面和具有成本效益的分子特征鉴定方法。

随着测序平台的发展和进步,越来越多的证据表明,NGS的高分辨率可以提供转基因作物更准确和充分的分子特征信息。比如,NGS可以识别基因组高度重复和复杂宿主的插入位点和旁侧序列;可以检测转基因植物中非预期T-DNA插入,包括T-DNA碎片和小的载体骨架片段;可以检测到T-DNA插入位点的缺失和易位。另外可将基因组分析与其它针对RNA、蛋白质和代谢产物的分析结合起来鉴定转基因作物的分子特征。未来,我们应该作出更多的努力来制定应用高通量测序技术进行分子特征鉴定的标准流程,建立信息共享数据库,其中包含已被批准、正在审查或在田间试验中的转基因作物的分子特征数据,开发适用于UGM作物、基因叠加GM作物和NPBT产生的作物的合适分子特征检测方法。

参考文献:

- [1] Breseghello F, Coelho A S G. Traditional and modern plant breeding methods with examples in rice (*Oryza sativa* L.) [J]. J Agric Food Chem, 2013, 61(35): 8277-8286. DOI: 10.1021/jf305531j.
- [2] Klümper W, Qaim M. A meta-analysis of the impacts of genetically modified crops [J]. PLoS One, 2014, 9(11): e111629. DOI: 10.1371/journal.pone.0111629.
- [3] Brookes G, Barfoot P. Environmental impacts of genetically modified (GM) crop use 1996-2015: Impacts on pesticide use and carbon emissions [J]. GM Crops Food, 2017, 8(2): 117-147. DOI: 10.1080/21645698.2017.1309490.
- [4] James C. 20th Anniversary (1996 to 2015) of the Global Commercialization of Biotech Crops and Biotech Crop Highlights in 2015. ISAAA Brief No. 51 [M]. Ithaca, NY: ISAAA, 2015.
- [5] Wahler D, Schauer L, Bendiek J, et al. Next-generation sequencing as a tool for detailed molecular characterisation of genomic insertions and flanking regions in genetically modified plants: a pilot study using a rice event unauthorised in the EU [J]. Food Anal Methods, 2013, 6(6): 1718-1727. DOI: 10.1007/s12161-013-9673-x.
- [6] Maghari B M, Ardekani A M. Genetically modified foods and social concerns [J]. Avicenna J Med Biotechnol, 2011, 3(3): 109-117.
- [7] Nicolia A, Manzo A, Veronesi F, et al. An overview of the last 10 years of genetically engineered crop safety research [J]. Crit Rev Biotechnol, 2014, 34(1): 77-88. DOI: 10.3109/07388551.2013.823595.
- [8] Schnell J, Steele M, Bean J, et al. A comparative analysis of insertional effects in genetically engineered plants: considerations for pre-market assessments [J]. Transgenic Res, 2015, 24(1): 1-17. DOI: 10.1007/s11248-014-9843-7.
- [9] EFSA Panel on Genetically Modified Organisms. Guidance for risk assessment of food and feed from genetically modified plants [J]. EFSA J, 2011, 9(5): 2150. DOI: 10.2903/j.efsa.2011.2150.
- [10] Triglia T, Peterson M G, Kemp D J. A procedure for *in vitro* amplification of DNA segments that lie outside the boundaries of known sequences [J]. Nucleic Acids Res, 1988, 16(16): 8186. DOI: 10.1093/nar/16.16.8186.
- [11] Ji J, Braam J. Restriction site extension PCR: a novel method for high-throughput characterization of tagged DNA fragments and genome walking [J]. PLoS One, 2010, 5(5): e10577. DOI: 10.1371/journal.pone.0010577.
- [12] Jones D H, Winistorfer S C. Sequence specific generation of a DNA panhandle permits PCR amplification of unknown flanking DNA [J]. Nucleic Acids Res, 1992, 20(3): 595-600. DOI: 10.1093/nar/20.3.595.
- [13] Liu Y G, Mitsukawa N, Oosumi T, et al. Efficient isolation and mapping of *Arabidopsis thaliana* T-DNA insert junctions by thermal asymmetric interlaced PCR [J].

- Plant J, 1995, 8(3): 457–463. DOI: 10.1046/j.1365-313x.1995.08030457.x.
- [14] O'Malley R C, Ecker J R. Linking genotype to phenotype using the *Arabidopsis* unimutant collection[J]. Plant J, 2010, 61(6): 928–940. DOI: 10.1111/j.1365-313x.2010.04119.x.
- [15] Ochman H, Gerber A S, Hartl D L. Genetic applications of an inverse polymerase chain reaction [J]. Genetics, 1988, 120(3): 621–623.
- [16] Kovalic D, Garnaat C, Guo L, et al. The use of next generation sequencing and junction sequence analysis bioinformatics to achieve molecular characterization of crops improved through modern biotechnology [J]. Plant Genome, 2012, 5(3). DOI: 10.3835/plantgenome2012.10.0026.
- [17] Yang L T, Xu S C, Pan A H, et al. Event specific qualitative and quantitative polymerase chain reaction detection of genetically modified MON863 maize based on the 5' -transgene integration sequence [J]. J Agric Food Chem, 2005, 53(24): 9312–9318. DOI: 10.1021/jf051782o.
- [18] Wang X B, Jiang L X, Wei L, et al. Integration and insertion site of *EPSPs* gene on the soybean genome in genetically modified glyphosate-resistant soybean [J]. Acta Agronomica Sinica, 2010, 36(3): 365–375. DOI: 10.3724/sp.j.1006.2010.00365.
- [19] Akritidis P, Pasentsis K, Tsaftaris A S, et al. Identification of unknown genetically modified material admixed in conventional cotton seed and development of an event-specific detection method [J]. Electron J Biotechnol, 2008, 11(2): 0. DOI: 10.2225/vol11-issue2-fulltext-11
- [20] Sun L, Gill U S, Nandety R S, et al. Genome-wide analysis of flanking sequences reveals that *Tnt1* insertion is positively correlated with gene methylation in *Medicago truncatula* [J]. Plant J, 2019, 98(6): 1106–1119. DOI: 10.1111/tpj.14291.
- [21] Guo B, Guo Y, Hong H, et al. Identification of genomic insertion and flanking sequence of *G2-EPSPS* and *GAT* transgenes in soybean using whole genome sequencing method [J]. Front Plant Sci, 2016, 7: 1009. DOI: 10.3389/fpls.2016.01009.
- [22] Wei F S, Coe E, Nelson W, et al. Physical and genetic structure of the maize genome reflects its complex evolutionary history [J]. PLoS Genet, 2007, 3(7): e123. DOI: 10.1371/journal.pgen.0030123.
- [23] Dubose A J, Lichtenstein S T, Narisu N, et al. Use of microarray hybrid capture and next-generation sequencing to identify the anatomy of a transgene [J]. Nucleic Acids Res, 2013, 41(6): e70. DOI: 10.1093/nar/gks1463.
- [24] Pauwels K, de Keersmaecker S C J, de Schrijver A, et al. Next-generation sequencing as a tool for the molecular characterisation and risk assessment of genetically modified plants: Added value or not? [J]. Trends Food Sci Technol, 2015, 45(2): 319–326. DOI: 10.1016/j.tifs.2015.07.009.
- [25] Guttikonda S K, Marri P, Mammadov J, et al. Molecular characterization of transgenic events using next generation sequencing approach [J]. PLoS One, 2016, 11(2): e0149515. DOI: 10.1371/journal.pone.0149515.
- [26] Inagaki S, Henry I M, Lieberman M C, et al. High-throughput analysis of T-DNA location and structure using sequence capture [J]. PLoS One, 2015, 10(10): e0139672. DOI: 10.1371/journal.pone.0139672.
- [27] Park D, Kim D, Jang G, et al. Efficiency to discovery transgenic loci in GM rice using next generation sequencing whole genome re-sequencing [J]. Genomics Inform, 2015, 13(3): 81–85. DOI: 10.5808/gi.2015.13.3.81.
- [28] Siddique K, Wei J J, Li R, et al. Identification of T-DNA insertion site and flanking sequence of a genetically modified maize event IE09S034 using next-generation sequencing technology [J]. Mol Biotechnol, 2019, 61(9): 694–702. DOI: 10.1007/s12033-019-00196-0.
- [29] Podevin N, De Buck S, De Wilde C, et al. Insights into recognition of the T-DNA border repeats as termination sites for T-strand synthesis by *Agrobacterium tumefaciens* [J]. Transgenic Res, 2006, 15(5): 557–571. DOI: 10.1007/s11248-006-9003-9.
- [30] Kononov M E, Bassuner B, Gelvin S B. Integration of T-DNA binary vector 'backbone' sequences into the tobacco genome: evidence for multiple complex patterns of integration [J]. Plant J, 1997, 11(5): 945–957. DOI: 10.1046/j.1365-313x.1997.11050945.x.
- [31] Wenck A, Czako M, Kanevski I, et al. Frequent collateral long transfer of DNA inclusive of the whole binary vector during *Agrobacterium*-mediated transformation [J]. Plant Mol Biol, 1997, 34(6): 913–922. DOI: 10.1023/a: 1005849303333.
- [32] Ülker B, Li Y, Rosso M G, et al. T-DNA-mediated transfer of *Agrobacterium tumefaciens* chromosomal DNA into plants [J]. Nat Biotechnol, 2008, 26(9): 1015–1017. DOI: 10.1038/nbt.1491.
- [33] Kim S R, An G. Bacterial transposons are co-transferred with T-DNA to rice chromosomes during *Agrobacterium*-mediated transformation [J]. Mol Cells, 2012, 33(6): 583–589. DOI: 10.1007/s10059-012-0010-9.
- [34] Philips J G, Naim F, Lorenc M T, et al. The widely used *Nicotiana benthamiana* 16c line has an unusual T-DNA integration pattern including a transposon sequence [J]. PLoS One, 2017, 12(2): e0171311. DOI: 10.1371/journal.pone.0171311.
- [35] Gheysen G, Villarroel R, Van Montagu M. Illegitimate recombination in plants: a model for T-DNA integration [J]. Genes Dev, 1991, 5(2): 287–297. DOI: 10.1101/gad.5.2.287.
- [36] Singer K, Shibolet Y M, Li J M, et al. Formation of complex extrachromosomal T-DNA structures in *Agrobacterium tumefaciens*-infected plants [J]. Plant Physiol,

- 2012, 160(1): 511–522. DOI:10.1104/pp.112.200212.
- [37] Tzfira T, Li J, Lacroix B, et al. *Agrobacterium* T-DNA integration: molecules and models[J]. Trends Genet, 2004, 20(8): 375–383. DOI:10.1016/j.tig.2004.06.004.
- [38] Liu Y G, Whittier R F. Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking[J]. Genomics, 1995, 25(3): 674–681. DOI:10.1016/0888-7543(95)80010-j.
- [39] Mueller P, Wold B. Ligation-mediated PCR: Applications to genomic footprinting[J]. Methods, 1991, 2(1): 20–31. DOI:10.1016/s1046-2023(05)80122-7.
- [40] Yang H M, Moon S H, Choi Y S, et al. Therapeutic efficacy of human embryonic stem cell-derived endothelial cells in humanized mouse models harboring a human immune system[J]. Arterioscler Thromb Vasc Biol, 2013, 33(12): 2839–2849. DOI:10.1161/atvbaha.113.302462.
- [41] Padgett S R, Kolacz K H, Delannay X, et al. Development, identification, and characterization of a glyphosate-tolerant soybean line[J]. Crop Sci, 1995, 35(5): 1451–1461. DOI:10.2135/cropsci1995.0011183x003500050032x.
- [42] Windels P, Taverniers I, Depicker A, et al. Characterisation of the Roundup Ready soybean insert[J]. Eur Food Res Technol, 2001, 213(2): 107–112. DOI:10.1007/s002170100336.
- [43] Zastrow-Hayes G M, Lin H, Sigmund A L, et al. Southern-by-sequencing: a robust screening approach for molecular characterization of genetically modified crops[J]. Plant Genome, 2015, 8(1): eplantgenome2014.08.0037. DOI:10.3835/plantgenome2014.08.0037.
- [44] Southern E M. Detection of specific sequences among DNA fragments separated by gel electrophoresis. 1975[J]. Biotechnology, 1992, 24: 122–139.
- [45] Pérez Urquiza M, Acatzi Silva A I. Copy number ratios determined by two digital polymerase chain reaction systems in genetically modified grains[J]. Metrologia, 2014, 51(1): 61–66. DOI:10.1088/0026-1394/51/1/61.
- [46] Yang L T, Ding J Y, Zhang C M, et al. Estimating the copy number of transgenes in transformed rice by real-time quantitative PCR[J]. Plant Cell Rep, 2005, 23(10/11): 759–763. DOI:10.1007/s00299-004-0881-0.
- [47] Ingham D J, Beer S, Money S, et al. Quantitative real-time PCR assay for determining transgene copy number in transformed plants[J]. Biotechniques, 2001, 31(1): 132–4, 136–40. DOI:10.2144/01311rr04.
- [48] Dobnik D, Spilberg B, Bogožalec Košir A, et al. Multiplex quantification of 12 European union authorized genetically modified maize lines with droplet digital polymerase chain reaction[J]. Anal Chem, 2015, 87(16): 8218–8226. DOI:10.1021/acs.analchem.5b01208.
- [49] Dong L, Meng Y, Sui Z, et al. Comparison of four digital PCR platforms for accurate quantification of DNA copy number of a certified plasmid DNA reference material[J]. Sci Rep, 2015, 5: 13174. DOI:10.1038/srep13174.
- [50] Félix-Urquidez D, Pérez-Urquiza M, Valdez Torres J B, et al. Development, optimization, and evaluation of a duplex droplet digital PCR assay to quantify the T-nos/hmg copy number ratio in genetically modified maize[J]. Anal Chem, 2016, 88(1): 812–819. DOI:10.1021/acs.analchem.5b03238.
- [51] Bodi K, Perera A G, Adams P S, et al. Comparison of commercially available target enrichment methods for next-generation sequencing[J]. J Biomol Tech, 2013, 24(2): 73–86. DOI:10.7171/jbt.13-2402-002.
- [52] Fraiture M A, Herman P, Papazova N, et al. An integrated strategy combining DNA walking and NGS to detect GMOs[J]. Food Chem, 2017, 232: 351–358. DOI:10.1016/j.foodchem.2017.03.067.
- [53] Schouten H J, Vande Geest H, Papadimitriou S, et al. Re-sequencing transgenic plants revealed rearrangements at T-DNA inserts, and integration of a short T-DNA fragment, but no increase of small mutations elsewhere[J]. Plant Cell Rep, 2017, 36(3): 493–504. DOI:10.1007/s00299-017-2098-z.
- [54] European Union Reference Laboratory for Genetically Modified Food and Feed. Guideline for the submission of DNA sequences and associated annotations within the framework of Directive 2001/18/EC and Regulation (EC) No. 1829/2003. [S].
- [55] Yang L, Wang C, Holst-Jensen A, et al. Characterization of GM events by insert knowledge adapted re-sequencing approaches[J]. Sci Rep, 2013, 3: 2839. DOI:10.1038/srep02839.
- [56] Holst-Jensen A, Bertheau Y, de Loose M, et al. Detecting un-authorized genetically modified organisms (GMOs) and derived materials[J]. Biotechnol Adv, 2012, 30(6): 1318–1335. DOI:10.1016/j.biotechadv.2012.01.024.
- [57] Cade R, Burgin K, Schilling K, et al. Regulatory science evaluation of whole genome sequencing and an insertion site characterization method for molecular characterization of GM maize[J]. J Regul Sci, 2018, 6(1): 1–14.
- [58] Campbell P J, Stephens P J, Pleasance E D, et al. Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing[J]. Nat Genet, 2008, 40(6): 722–729. DOI:10.1038/ng.128.
- [59] Fullwood M J, Wei C L, Liu E T, et al. Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses[J]. Genome Res, 2009, 19(4): 521–532. DOI:10.1101/gr.074906.107.
- [60] Hormozdiari F, Hajirasouliha I, McPherson A, et al. Simultaneous structural variation discovery among multiple paired-end sequenced genomes[J]. Genome Res, 2011,

- 21(12): 2203–2212. DOI:10.1101/gr.120501.111.
- [61] Wang J, Wang W, Li R Q, et al. The diploid genome sequence of an Asian individual [J]. *Nature*, 2008, 456 (7218): 60–65. DOI:10.1038/nature07484.
- [62] Ajay S S, Parker S C, Abaan H O, et al. Accurate and comprehensive sequencing of personal genomes [J]. *Genome Res*, 2011, 21 (9): 1498–1505. DOI: 10.1101/gr.123638.111.
- [63] Sims D, Sudbery I, Hott N E, et al. Sequencing depth and coverage: key considerations in genomic analyses [J]. *Nat Rev Genet*, 2014, 15 (2): 121–132. DOI: 10.1038/nrg3642.
- [64] Liang C J, Dijk J P, Scholtens I M J, et al. Detecting authorized and unauthorized genetically modified organisms containing vip3A by real-time PCR and next-generation sequencing [J]. *Anal Bioanal Chem*, 2014, 406 (11): 2603–2611. DOI:10.1007/s00216-014-7667-1.
- [65] New Techniques Working Group (2011) Final report. http://www.keine-gentechnik.de/fileadmin/files/Infodienst/Dokumente/2011_Report_of_the_Working_Group_on_New_Techniques_Final.pdf. Accessed 30 July 2018 [R].
- [66] New Techniques Working Group, Poudelet E. Multilateral meeting on “The future of plant breeding techniques in the European Union”. Steering note [R]. 2014.
- [67] Lusser M, Parisi C, Plan D, et al. Deployment of new biotechnologies in plant breeding [J]. *Nat Biotechnol*, 2012, 30(3): 231–239. DOI:10.1038/nbt.2142.
- [68] Schaart J G, van de Wiel C C M, Lotz L A P, et al. Opportunities for products of new plant breeding techniques [J]. *Trends Plant Sci*, 2016, 21 (5): 438–449. DOI: 10.1016/j.tplants.2015.11.006.
- [69] Zhang H, Zhang J, Wei P, et al. The CRISPR/Cas9 system produces specific and homozygous targeted gene editing in rice in one generation [J]. *Plant Biotechnol J*, 2014, 12(6): 797–807. DOI:10.1111/pbi.12200.
- [70] Feng Z Y, Mao Y F, Xu N F, et al. Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in *Arabidopsis* [J]. *Proc Natl Acad Sci USA*, 2014, 111(12): 4632–4637. DOI:10.1073/pnas.1400822111.
- [71] Nekrasov V, Wang C, Win J, et al. Rapid generation of a transgene-free powdery mildew resistant tomato by genome deletion [J]. *Sci Rep*, 2017, 7 (1): 482. DOI: 10.1038/s41598-017-00578-x.
- [72] Tang X, Liu G, Zhou J, et al. A large-scale whole-genome sequencing analysis reveals highly specific genome editing by both Cas9 and Cpf1 (Cas12a) nucleases in rice [J]. *Genome Biol*, 2018, 19(1): 84. DOI:10.1186/s13059-018-1458-5.
- [73] Li J, Manghwar H, Sun L, et al. Whole genome sequencing reveals rare off-target mutations and considerable inherent genetic or/and somaclonal variations in CRISPR/Cas9-edited cotton plants [J]. *Plant Biotechnol J*, 2019, 17(5): 858–868. DOI:10.1111/pbi.13020.
- [74] Tsai S Q, Joung J K. Defining and improving the genome-wide specificities of CRISPR–Cas9 nucleases [J]. *Nat Rev Genet*, 2016, 17(5): 300–312. DOI:10.1038/nrg.2016.28.
- [75] Pabinger S, Dander A, Fischer M, et al. A survey of tools for variant analysis of next-generation genome sequencing data [J]. *Brief Bioinform*, 2014, 15(2): 256–278. DOI:10.1093/bib/bbs086.
- [76] Hall R D, de Maagd R A. Plant metabolomics is not ripe for environmental risk assessment [J]. *Trends Biotechnol*, 2014, 32 (8): 391–392. DOI: 10.1016/j.tibtech.2014.05.002.
- [77] Gong C Y, Wang T. Proteomic evaluation of genetically modified crops: current status and challenges [J]. *Front Plant Sci*, 2013, 4: 41. DOI:10.3389/fpls.2013.00041.
- [78] Ricroch A E, Bergé J B, Kuntz M. Evaluation of genetically engineered crops using transcriptomic, proteomic, and metabolomic profiling techniques [J]. *Plant Physiol*, 2011, 155 (4): 1752–1761. DOI:10.1104/pp.111.173609.
- [79] Stewart D, Shepherd L V T. Metabolomics for the safety assessment of genetically modified (GM) crops [M]//Metabolomics in Food and Nutrition. Amsterdam: Elsevier, 2013: 192–216. DOI:10.1533/9780857098818.2.192.
- [80] Zhou J, Ma C, Xu H, et al. Metabolic profiling of transgenic rice with cryIac and sck genes: an evaluation of unintended effects at metabolic level by using GC–FID and GC–MS [J]. *J Chromatogr B Analyt Technol Biomed Life Sci*, 2009, 877 (8/9): 725–732. DOI: 10.1016/j.jchromb.2009.01.040.
- [81] Coll A, Nadal A, Collado R, et al. Natural variation explains most transcriptomic changes among maize plants of MON810 and comparable non-GM varieties subjected to two N-fertilization farming practices [J]. *Plant Mol Biol*, 2010, 73(3): 349–362. DOI:10.1007/s11103-010-9624-5.
- [82] Clarke J D, Alexander D C, Ward D P, et al. Assessment of genetically modified soybean in relation to natural variation in the soybean seed metabolome [J]. *Sci Rep*, 2013, 3: 3082. DOI:10.1038/srep03082.
- [83] Rao J, Yang L, Guo J, et al. Metabolic changes in transgenic maize mature seeds over-expressing the *Aspergillus niger phyA2* [J]. *Plant Cell Rep*, 2016, 35(2): 429–437. DOI:10.1007/s00299-015-1894-6.
- [84] Harrigan G G, Lundry D, Drury S, et al. Natural variation in crop composition and the impact of transgenesis [J]. *Nat Biotechnol*, 2010, 28 (5): 402–404. DOI: 10.1038/nbt0510-402.