

综述 Reviews

Mapping-by-sequencing在作物农艺性状相关基因定位的应用

杨琳, 杨剑飞, 王欢欢, 王宇*, 李玉花*

东北林业大学生命科学学院, 哈尔滨150040

摘要: 通过挖掘控制重要农艺性状的基因以提高农作物产量是育种学家一直关注的问题。然而由于农作物遗传背景复杂且基因组庞大, 该过程往往步骤繁琐, 耗时长且成本较高。Mapping-by-sequencing (利用高通量测序进行基因定位, MBS) 利用二代测序技术和遗传分离群体混池建库分析的方法成功解决了上述问题, 加快了农作物重要农艺性状的基因定位同时也降低了研究成本。该方法已经成功应用于重要农艺性状相关联的突变位点鉴别和基因定位。本文综述了Mapping-by-sequencing的实验原理, 并总结了该方法在水稻、大麦、小麦及玉米等农作物中与目标农艺性状相关基因定位的应用。

关键词: Mapping-by-sequencing; Bulk segregant analysis; 基因定位; 作物

为了应对人口增长给粮食生产带来的巨大压力, 科学家们致力于利用多种新兴、高效的育种方法如: 分子标记技术、细胞工程技术、转基因技术、自然和人为诱发的遗传变异等, 来提高作物的产量及质量。运用正向遗传学筛选定位出具有优良农艺性状的候选基因是研究遗传变异育种的重要手段之一。然而, 对于基因组庞大且复杂的农作物来说, 尽管传统方法已经广泛应用于农作物的目标基因定位, 但仍存在局限性: 传统图位克隆基因定位步骤繁琐, 定位区间宽泛且精确度低, 而且不同农作物中开发的分子标记数量具有显著差别(Fazio等2003; Bernardo 2008; Ishihara等2014)。

随着第二代高通量测序技术的迅猛发展, 基因组从头测序、重测序和转录组测序能够快速而准确地为研究者提供相应作物的基因组、转录组信息, 加快了农作物相关农艺性状的基因定位研究(Xu等2013; Ogura和Busch 2014)。目前, 许多作物已经获得全基因组序列信息草图可供后续功能基因组学研究, 各类种质资源群体也完成了全基因组重测序(whole genome resequencing, WGR), 这都为农作物重要农艺性状的基因定位研究奠定了良好的基础(崔晓峰2013)。虽然高通量测序加快了突变位点的筛选过程, 但是面对农作物的庞大定位群体, 对每个个体逐一进行全基因组重测序, 需要构建多个DNA文库, 仍存在实验周期长、人力耗费量大、资金花费多和数据分析复杂等难题。

近几年, 一种名为Mapping-by-sequencing (MBS)的策略可以解决以上诸多难题, 从而快速鉴定出目标基因。Mapping-by-sequencing是一种将

群体分离分析法与全基因组重测序、外显子测序及RNA测序等第二代高通量测序技术相结合进行基因定位分析的方法, 其中包括SHOREMap (Schneeberger等2009)以及MutMap (Abe等2012)等方法。MBS利用一对目标性状存在相对差异的亲本所产生的分离群体, 将分离群体按照目标性状表型混合成DNA池(DNA pool), 随后运用二代测序技术对DNA池进行全基因组重测序, 将测序结果比对到亲本之一的参考序列上, 分析差异SNP, 筛选与目标性状紧密连锁的SNPs (单核苷酸多态性, single nucleotide polymorphism)。它可以同时定位突变位点和鉴定候选突变基因, 能够精细定位到点突变, 极大地提高了定位候选基因的效率, 缩短了从建立作图群体到精细定位目标基因的步骤和时间。该方法已经成功应用于模式生物拟南芥和其他动植物中(Bigelow等2009; Nijman等2010; Zuryn等2010; Austin等2011; Uchida等2011)。在农作物中, MBS方法目前已经应用于水稻、大麦、小麦、玉米等重要作物的农艺性状的基因定位(Abe等2012; Liu等2012; Mascher等2014; Ramirez-Gonzalez等2015)。本文总结了近年来MBS在农作物的农艺性状基因定位中的应用, 以期为相关研究者提供参考。

收稿 2016-11-21 修定 2017-02-22

资助 国家自然科学基金(31471911、31272200和31401907)和黑龙江省自然科学基金(C2015052)。

* 共同通讯作者(E-mail: vonov@sina.com; lyhshen@126.com)。

1 Mapping-by-sequencing技术原理及分类

1.1 MBS技术原理

首先,通过正向遗传学技术筛选所得的目标性状突变体要与非突变亲本进行杂交获得分离群体。分离群体的构建方法通常分为两种:第一种是将纯合的目标突变体与其远缘亲本品系杂交(Schneeberger等2009);第二种就是纯合的目标突变体与野生型亲本杂交(Abe等2012)。然后,通过第二代高通量测序技术对野生型亲本及突变体池进行全基因组重测序,将突变体池的序列比对到野生型亲本的参考基因组上,分析它们之间的单核苷酸多态性(SNP)和删除或者插入(Indel)序列,随后利用SAMTools(Li等2009)、GATK(DePristo等2011)、MAQ(Bigelow等2009)、BamBam(Page等2014)及Sliding Window Analysis等SNP分析软件计算SNP频率(SNP-Index),就可以将候选突变位点定位于染色体上一定区域内。SNP频率即为与野生型相比,突变池中的某一位点发生突变的reads数量占定位到该位点全部reads个数的比值。根据基因的连锁互换定律,目标突变位点附近的SNP频率会形成一个峰值区域,通常位于峰值区域中间且突变的SNP频率接近于1的SNP就是真正的候选突变位点。具体过程如图1所示。

MBS技术解决了农作物群体数量庞大及研究成本高的问题。通常只需要对20~50株F₂代个体进行混池测序就可以筛选出目标SNP位点进行后续定位分析。与传统的基因定位方法相比,MBS不需要开发新的分子标记,而是直接分析突变体池与野生型亲本的SNP差异就能够完成目标基因的精细定位,缩短了筛选候选突变位点的时间,便于快速鉴定出候选突变基因。

1.2 MBS分类

1.2.1 SHOREMap

SHOREMap主要适用于目标性状突变体与远缘亲本的杂交分离群体分析。该方法将F₂代群体中突变表型个体提取DNA后混池,建立测序文库进行全基因组重测序,利用一种名为SHOREMap的软件分析突变亲本与远缘亲本等位基因频率的偏差,最终确定突变基因的位置(Schneeberger等2009)。

1.2.2 MutMap

MutMap是将目标性状突变体与未诱变的野生型亲本杂交,再自交构建F₂代定位群体($n > 100$),

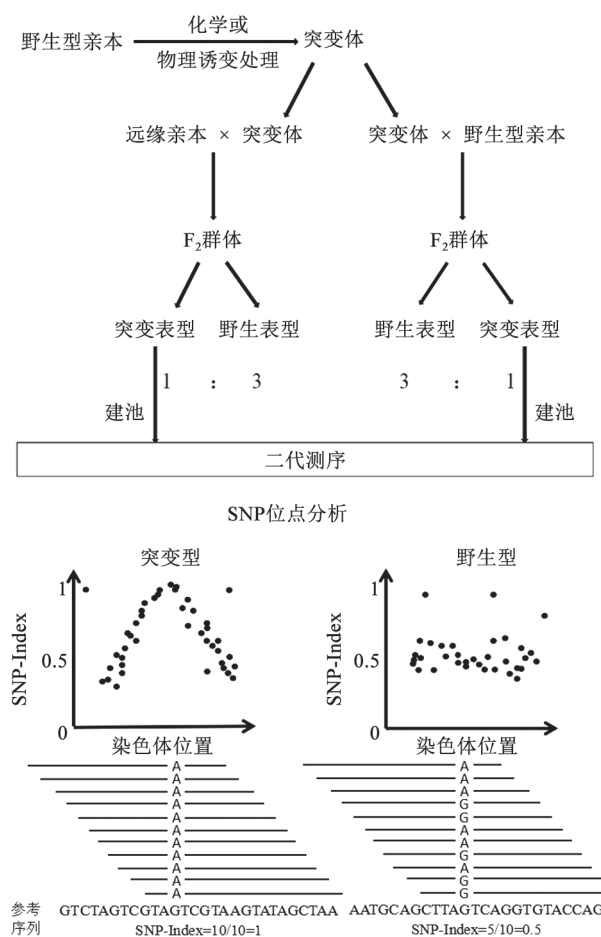


图1 MBS技术分析目标基因定位流程

Fig.1 The flowchart of MBS

参照James等(2013)文献进行修改。

在F₂代中至少筛选出具有极端突变表型的20株个体进行DNA混池(Birkeland等2010; Abe等2012; Zhu等2012; Nordström等2013)。利用全基因组测序方法对突变体池测序并筛选候选SNP位点,根据与野生型的差异位点将目标基因定位到染色体上。该方法减少了遗传背景的差异,群体构建数量较小。

2 基于Mapping-by-sequencing技术的改进及其应用

2.1 MutMap特殊突变类型

自然界中,部分农作物存在生长早期致死或杂交不育的问题,无法通过杂交来构建定位群体,以至于难以进行基因定位研究。Fekih等(2013)针对这一现象,设计出一种基于MutMap改进的方法,名为MutMap+。与MutMap相比,该方法无需将成年突变纯合体与野生型亲本杂交,能够应对一些致病、致死基因或不育变异株的研究。它利用

EMS诱变野生型亲本获得含有突变位点的杂合 M_1 代,通过 M_1 代自交获得 M_2 代。而在 M_2 代中,当目标性状由隐性单基因控制时,野生型与突变型的分离比为3:1,其中突变型由于基因功能缺失无法得到后代,而杂合的 M_2 代野生型个体可以获得子代。因此通常选取10株野生型表型的 M_2 代个体进行自交获得大量 M_3 代个体, M_3 发生性状分离,挑选 M_2 代杂合的野生型亲本和 M_3 代中野生型和突变型各20~40株个体,混池测序。通过与亲本基因组进行比对,筛选出野生型池和突变型池中含有差异SNP频率接近于1的位点,将这些位点进行二次筛选,将仅存在于突变型池中的SNP位点作为候选突变位点。

2.2 MutMap-Gap

当栽培变种的重测序结果与参考序列比对时,有时部分测序reads无法比对到参考序列上,可能导致该变种基因组序列拼接结果中出现一个间隙(gaps)。由于参考基因组中并不包含这段序列因此无法通过简单的比对鉴别出这些区域。根据以上的问题,Takagi等(2013b)提出MutMap-Gap这一方法。它是一种结合了MutMap和*de novo*组装的基因定位方法。首先,运用MutMap定位出含有候选突变位点的区域,随后将定位到该区域附近的序列与未定位到目标基因组上的序列进行组合拼接,获得目标突变区域的基因组序列,定位出拼接序列中的候选突变位点。

该方法适用于发掘出某些农作物品系中特有的,但在该作物的参考基因组中不存在的且与性状相关联的突变基因。因此,当研究者利用MutMap或MutMap+无法鉴别出突变基因时,可以尝试用MutMap-Gap进一步挖掘可能存在于某些品系特有区域中的候选基因。

2.3 Cloud-Map

Cloud-Map是一种建立于MBS基础上,充分地利用云计算技术分析突变基因组序列的方法,其网址为<http://wiki.g2.bx.psu.edu/>(Minevich等2012)。它将MutMap技术与Galaxy可视化平台结合,使用者仅需点击网页上传数据并设定相应参数,就可以精确地确认目标基因的定位区间,并且可以快速检索候选突变位点,为许多不具备高性能服务器的实验室提供了便利。

Cloud-Map利用BWA、Bowtie等不同类型的比对工具将上传到服务器的原始数据装配到参考基因组上,通过GATK和SAMTools等SNP筛选工具将突变体序列与野生型序列进行比对,筛选出野生型亲本与突变体之间的差异SNP位点。除此之外,用户可以在Cloud-Map中的候选基因列表里检索到由氨基酸变异导致性状发生变化的候选基因和尚未注释的区域。

该方法简化了基因定位数据分析的过程并成功解决部分没有生物信息背景的研究者不能够熟练运用计算机语言处理二代测序数据的难题。

2.4 QTL-seq

农作物中重要的农艺性状大多数以数量性状(quantitative trait loci, QTL)为主,因此研究作物数量性状遗传对农作物育种具有十分重要的意义。传统QTL定位存在大量的数量性状,逐个筛选非常费时,需要通过多次与野生型亲本杂交来纯化遗传背景并不断开发新的分子标记才能够精细定位出QTLs。Takagi等(2013a)结合第二代高通量测序技术,基于MBS的原理,开发了一种快速完成QTL定位的方法,被称为QTL-Seq。QTL-seq一般采用一对相对性状存在差异的亲本杂交 F_2 群体或重组自交系作为定位群体,筛选群体中目标性状存在相对差异的极端表型的个体并分为两个极端DNA池。利用全基因组重测序对两个极端池测序并将结果比对到亲本的序列上进行分析,筛选出与性状相关联的SNP位点并计算SNP在不同表型亲本中出现的频率,通过分析SNP频率就可以定位候选的变异位点,获得候选目的基因。

QTL-Seq只需要通过分析杂交后代表型并分别混池测序分析即可快速鉴定出与表型相关联的QTLs,提高了QTLs定位的精确性,甚至可以找到真正调控目的性状的候选基因。

3 高复杂度基因组作物的MBS分析策略

农作物基因组庞大、遗传背景复杂,如大麦(*Hordeum vulgare* L.)本身具有相对较大的基因组(约5 Gb)和约有80%重复DNA序列,而小麦(*Triticum aestivum* L.)是天然的六倍体植株,基因组更加庞大约有17 Gb。对这些农作物进行全基因组重测序,会产生基因组测序成本高以及后期数据计算难度大等问题。

为了解决这一难题,部分研究者基于MBS的原理,利用二代测序技术的延展性提出简化基因组的方法来完成基因定位,如:全外显子组捕获测序BSA (Bulked segregant analysis) (Hodges等2007; Mascher等2013b, 2014; Pankin等2014)、BSR-Seq (Bulked Segregant RNA-Seq) (Liu等2012)等。全外显子组捕获测序BSA是一种利用全外显子组捕获测序技术进行候选突变位点定位的方法。该方法将外显子捕获芯片与片段化的DNA进行杂交,对富集到芯片上的DNA片段进行洗脱扩增,最后通过高通量测序并筛选SNP。全外显子组捕获测序BSA能够降低测序成本,简化数据分析过程。而BSR-Seq技术则是将RNA-Seq与BSA相结合,分析F₂中突变表型池与野生表型池之间的转录本SNP来筛选候选突变SNP。通常仅出现于突变池中,而野生池中不存在的SNP即为候选SNP。BSR-Seq不仅可以用于候选突变基因定位还可以明确突变体的基因表达模式,更快速地鉴定出与目标性状相关联的基因。

除此之外,在水稻研究中,一种新颖的方法NIKS (needle in the *k*-stack)也可以在参考序列缺失的条件下对重要的农艺性状进行基因定位。NIKS利用亲本及突变体的全基因组测序数据,将两者分别拼接成长度为 k 的序列,利用Jellyfish软件分别计算出亲本及突变体的 k 长度序列出现于全基因组序列的频率。由于突变位点只存在于突变体中,而其他位点亲本和突变体都有,因此当利用突变体与亲本比较时,突变体中 k -mer值相对较低的区域极可能为突变位点所在区域。若野生型与突变体相比较,某区域 k -mer差值小且位置相同就可以确定是突变区域(Marçais和Kingsford 2011)。

对于高复杂度基因组的农作物来说,单个个体分析是不现实的。因此NIKS需要利用分离群体进行分析。在水稻亚种‘Hitomebore’中,利用NIKS对突变体池12~17倍覆盖度的测序结果分析,成功鉴定出3~11个候选突变位点(Nordström等2013)。

4 Mapping-by-sequencing技术在不同作物中的应用

4.1 MBS技术在水稻中的应用

水稻(*Oryza sativa* L)是世界主要粮食作物之一,为世界近一半人口提供了充足的食物资源。

Abe等(2012)利用MutMap的方法,EMS诱变1 200株水稻,自交纯合3~4代后(M₃~M₄),获得叶绿素显著减少的突变体。突变体与野生型杂交获得F₂代(>200株),选取F₂代中20株突变型个体混池,以12倍水稻基因组的覆盖度进行测序,鉴定出7个位于10号染色体的SNP位点,5个位于1号染色体上的SNP位点,最终通过实验验证候选SNP定位在*Os-CAOI*基因上;该研究者同时还利用该方法鉴定出几个水稻重要的农艺性状,如半矮生、雄性不育等。陈竹锋等(2014)利用MutMap以20~30倍的覆盖度对*osms55*雄性不育突变体F₂群体中的331株极端突变个体进行混池测序,成功鉴定并克隆出与雄性不育突变表型相关联的基因*LOC_Os02g40450*,其突变位于第4个内含子的剪切识别位点GT-AG上,其中AG的G突变为A,导致突变体蛋白比野生型少5个氨基酸,引起减数分裂染色体交叉(Cross-Over, CO)数目显著下降,而存在的少数CO也为随机分布,从而导致水稻产生雄性不育。Yang等(2016)利用MutMap的方法,从砷酸盐敏感的野生型粳稻*Osphf1-7*的EMS诱变群体中筛选出具有砷酸盐耐受性的突变体*OsPHF1*,其在含有30 μmol·L⁻¹的砷酸盐土壤中存活率高于野生稻。对*OsPHF1*突变体F₂代群体中具有砷酸盐耐受能力的52株突变型子代以30倍基因组的覆盖度进行测序,将SNP位点定位在*LOC_Os01g72570*基因的第4个外显子内成功鉴别出候选突变基因*OsCLT1*。

Fekih等(2013)采用Mut-Map+对一个早期生长发育不良的水稻突变体Hit9188进行研究,该突变体呈现出植株矮小,叶色变浅的表型且在发芽3周后死亡。通过从M₃中选取40株野生型和40株突变型植株分别混合建池并进行测序比对,将突变位点定位于1号染色体上的1.0~5.23 Mb处。进一步分析发现在该区域内的*OsNAP6*基因编码区存在1个SNP位点,导致其编码的蛋白质第378位的丙氨酸突变为苏氨酸,使该基因功能丧失表现为植株矮小的表型。Mut-Map+还被用于研究具有白化致死突变表型的Hit11440水稻中,研究者从M₃代中选取20株突变体和20株野生型个体混合建池后以基因组12倍的覆盖度进行全基因组重测序,最终成功筛选出突变基因*LOC_Os08g-0139100*,该基因位于8号染色体的0.32~4.66 Mb

处。Takagi等(2013b)利用MutMap-Gap成功鉴定出水稻品种‘Hitomebore’的抗病基因*Pii*。

Takagi等(2013a)以水稻‘Nortai’和‘Hitomebore’的241个重组体为材料,利用QTL-seq的方法筛选20个抗稻瘟病与20个易感染的RIL,以大于6倍基因组覆盖度的测序深度测序,将候选位点定位于6号染色体的2.39~4.39 Mb处。Yang等(2013)利用QTL-seq,以水稻‘Nipponbare’与耐冷品种‘LPBG’的10 800个F₃杂交后代个体为材料,根据耐寒性筛选出430株极端不耐寒个体和385株极端耐寒个体分别混池测序。利用统计学算法对450 000个SNP位点进行过滤并定位到水稻‘Nipponbare’的参考序列上,最终获得6个QTLs位点与水稻耐寒性相关,分别位于第1、2、5、8和10号染色体上。

MBS方法应用于水稻功能基因定位时,一般以突变体与野生型亲本杂交的后代群体作为定位群体(>200株),选取20~50株的目标性状突变体混池测序即可(James等2013)。大多数研究表明以10倍左右的基因组覆盖度对突变体池测序就可以筛选出候选突变位点,然而由于品种差异,有时测序深度需要提高到30倍以上才能获得精确的候选基因定位信息。

4.2 MBS技术在大麦中的应用

大麦(*Hordeum vulgare* L.)是继玉米、水稻、小麦之后排名世界第4的重要粮食类作物。大麦是自花授粉作物,受到天然筛选及人工育种的影响,其群体结构非常复杂而且基因组复杂(约5 Gb),约有80%为重复DNA序列,这使得测序的成本增高。

Mascher等(2014)对通过X-ray诱变得到的的一种茎节点增多且矮化的突变体*mnd*进行突变位点筛选,分别对F₂代中18株突变表型个体及30株野生型个体混池,利用外显子组测序的方法,将测序结果与野生型品种‘Barke’的基因组序列进行比对,筛选SNP位点发现在染色体5H处有一个明显的SNP频率峰值。随后在该峰值相邻的80~100 Mb区域内,利用pop-seq (Mascher等2013a)筛选出含有由X-ray引起部分缺失的突变位点,并最终定位于染色体5H的96 cM处,从而鉴定出突变基因*Mnd*,该基因属于细胞色素P₄₅₀中的CRP78A家族,可以缩短叶原基间隔期,加速叶片萌发。Pankin等(2014)以春大麦早花突变体*eam5*与野生型亲本

‘Bowman’的864株F₂群体为研究对象,对其中204株表现出早花表型个体及突变亲本、野生亲本进行全外显子组捕获测序,分别从‘Bowman’、*eam5*及突变池中获得93、80和211 Mb的序列信息并比对到大麦参考序列(IBSC 2012)上,定位出候选突变位点位于染色体5H的122.6~129.9 cM处,鉴定出与早花表型相关联的候选基因*HvPHYC*。

与水稻不同,由于大麦自身庞大的基因组和复杂的遗传背景,无法利用全基因组重测序进行突变基因筛选,因此采用全外显子组捕获测序来鉴定突变位点。早期研究者Mascher等(2013b)成功构建了61.6 Mb的大麦外显子芯片,并利用该芯片鉴别出了许多与重要农艺性状相关联的候选基因。

4.3 MBS技术在小麦中的应用

小麦是单子叶多倍体植物,四倍体小麦(*Triticum turgidum* ssp. *durum*)和六倍体小麦(*T. aestivum*)为两大主要食用小麦品种,其单倍体基因组大小为5.3 Gb左右,包含近85%~90%的重复序列且没有完整的基因组序列信息。Ramirez-Gonzalez等(2015)通过将抗黄锈病的小麦突变体*Yr15*与‘AVS’小麦杂交获得F₂分离群体,对232个F₂代植株中抗病和易感病表型个体分别混池。利用RNA-Seq以20倍覆盖度分别对其和亲本进行测序,将结果比对到单基因簇(Unigenes)和UCW基因模型上(Katoh和Standley 2013)。同时以‘中国春’小麦基因组为参考,筛选出只存在于野生型亲本中的1 582个SNP位点,其中23个SNPs与*Yr15*性状相关联,并将*Yr15*定位于1号染色体上的0.26~0.51 cM间隔处。Trick等(2012)利用BSR-Seq,以粮食蛋白浓度存在相对差异的2种小麦品种‘LDN’和‘RSL65’为材料,将候选SNP定位于染色体组6BS上,鉴定出控制粮食蛋白浓度的基因*GPC* (Grain Protein Content)。

MBS技术在小麦基因定位研究中,常用BSR-seq来筛选与表型性状相关联的候选突变位点,并分析突变型与野生型转录组中差异表达基因群,高效地定位出候选突变基因,该方法不仅解决了小麦高度复杂性的基因组问题,还节约了测序的成本。

4.4 MBS技术在玉米中的应用

在玉米的研究中,Liu等(2012)构建一个非‘B73’玉米背景下叶片表皮粗糙的突变体与叶片表面光滑‘B37’玉米的F₂群体,从中各选30株突变表

型个体与野生表型个体分别混池。利用转录组测序的方法,将结果比对到亲本之一的‘B37’玉米参考基因组序列上,成功将突变位点定位于4号染色体183.5~185.2 Mb的间隔中,鉴定突变基因为*GL3*,该基因控制叶片表皮蜡质积累。Haase等(2014)将早花矮株‘B73’与晚花高株‘Mo17’的杂交群体以及两者的重组自交系作为研究群体,利用QTL-seq将控制玉米株高的2个QTLs定位于3号和9号染色体上,并定位出控制玉米开花时间的6个QTLs,分别位于1、4、5、8和9号染色体上。Chen等(2014)以‘Chang7-2’与‘787’杂交获得的F₂群体为材料,鉴定出位于第10号染色体上控制玉米须颜色的基因*r1*及第3号染色体上控制穗分支数量的基因*bal*。

4.5 MBS技术在其他作物中的应用

除以上作物之外, MBS技术也被应用于其他作物中。Thyssen等(2015)选取野生型陆地棉品种‘DP5690’与无絮陆地棉突变体Ligon (*Li₁*)为试

验材料,将杂交获得2 567株F₂分离群体中的100株突变型个体和100株野生型个体分别进行混池测序,利用BSA-seq将*Li₁*位点定位于12号染色体的255 kb间隔中。在黄瓜中,以早花系‘Muromskij’ (CGN23617)和晚花系‘9930’ (P₁)的F₂群体为试验材料,利用QTL-seq筛选出控制黄瓜早花性状的基因*Efl.1* (Lu等2014)。在番茄中,利用QTL-Seq成功筛选出3个与果实重量相关联的QTLs,分别位于1、3和11号染色体上(Illa-Berenguer等2015)。Das等(2015)以鹰嘴豆F₄群体(cv. CC7184×cv. ICC15061)中的221株个体为材料,利用SNP和SSR分子标记将控制种子重量的QTL定位于1号染色体。

本文对MBS成功应用于以上几种重要农作物的农艺性状的基因定位进行总结(表1)。由于农作物的不同,构建群体和测序的策略也不同,我们需要考虑诸多因素,例如群体构建的大小、测序方法的选择等,具体情况具体分析。

表1 作物中Mapping-by-sequencing技术的应用

Table 1 Application of Mapping-by-sequencing in crops

作物	构建定位群体方法	突变体建池群体数/个	测序深度/倍	MBS应用方法	基因	参考文献
水稻	F ₂	20	12×	MutMap	<i>OsCAO1</i>	Abe等2012
	M ₃	40	12×	MutMap+	<i>OsNAP6</i>	Fekih等2013
	M ₃	20	12×	MutMap+	<i>LOC_Os08g0139100</i>	Fekih等2013
	F ₂	24	8×	MutMap-Gap	<i>Pii</i>	Takagi等2013b
	F ₂	331	20~30×	MutMap	<i>LOC_Os02g40450</i>	陈竹锋等2014
	F ₂	52	30×	MutMap	<i>OsCLT1</i>	Yang等2016
	RIL	20	>6×	QTL-seq	<i>qCTSS-1</i> 、 <i>qCTSS-2</i> 、 <i>qCTSS-5</i> 、 <i>qCTSS-8</i> 和 <i>qCTSS-10</i>	Takagi等2013a
大麦	F ₂	18	-	全外显子捕获 测序BSA	<i>Mnd</i>	Mascher等2014
	F ₂	204	-	全外显子捕获 测序BSA	<i>HvPHYC</i>	Pankin等2014
小麦	F ₂	20	20×	BSR-seq	<i>Yr15</i>	Ramirez-Gonzalez等2015
	F ₂	20	20×	BSR-seq	<i>GPC-B1</i>	Trick等2012
玉米	F ₂	30	-	BSR-seq	<i>GL3</i>	Liu等2012
	F ₂	30	-	BSR-seq	<i>r1</i> 、 <i>bal</i>	Chen等2014

5 注意事项及展望

Mapping-by-sequencing是一种崭新、快速有效地完成遗传作图及候选突变位点鉴别的方法。MBS灵活运用分离群体分析法和二代测序技术,

从而大幅度降低了测序成本,缩短构建群体时间,成功发掘出许多与重要农艺性状相关联的基因。然而MBS应用于农作物时,应该注意以下这些问题:

对于遗传背景复杂的农作物来说,群体构建尤为重要。MBS采用分离群体分析法,将野生型亲本与突变体杂交构建F₂代群体,并筛选突变表型个体混池建库进行测序分析。其中应选取多代自交的纯合野生型作为亲本进行诱变处理,同时值得注意的是,选择目标突变体建池时,一定要关注极端表型,尽量避免混入野生型或不清晰的表型,否则将对后续SNP-index的计算产生很大的干扰。另外,候选突变SNP定位时可能会存在假阳性,而群体建池的大小和基因组测序的平均覆盖率将会影响SNP假阳性的比率。一般情况下,杂交群体应大于200株,突变体建池至少30株。基因组测序时,建议测序覆盖度要达到30倍以上。

MBS目前主要致力于隐性遗传突变的研究,显性突变的研究结果尚未取得显著成果。而且针对农作物的数量性状研究,利用MBS只可以定位出其主效的QTL,微效QTL则无能为力。但是随着高通量测序技术的日趋成熟,测序成本的降低,相信MBS可以克服这些弊端,替代传统遗传图谱,广泛地应用于高复杂度基因组的农作物研究。MBS将为农作物基因组学研究提供强有力的工具,推动作物育种快速发展,带领作物科学研究跨入新的时代。

参考文献

- Abe A, Kosugi S, Yoshida K, Natsume S, Takagi H, Kanzaki H, Matsumura H, Yoshida K, Mitsuoka C, Tamiru M, et al (2012). Genome sequencing reveals agronomically important loci in rice using MutMap. *Nat Biotechnol*, 30 (2): 174–178
- Austin RS, Danielle V, George S, Robert B, Provart NJ, Dario B, Zhang J, Pauline F, Gong Y, Wang PW, et al (2011). Next-generation mapping of *Arabidopsis* genes. *Plant J*, 67 (4): 715–725
- Bernardo R (2008). Molecular markers and selection for complex traits in plants: learning from the last 20 years. *Crop Sci*, 48 (5): 1649–1664
- Bigelow H, Doitsidou M, Sarin S, Hobert O (2009). MAQGene: software to facilitate *C. elegans* mutant genome sequence analysis. *Nat Methods*, 6 (8): 549
- Birkeland SR, Jin N, Ozdemir AC, Jr LR, Weisman LS, Wilson TE (2010). Discovery of mutations in *Saccharomyces cerevisiae* by pooled linkage analysis and whole-genome sequencing. *Genetics*, 186 (4): 1127–1137
- Chen Z, Wang B, Dong X, Liu H, Ren L, Chen J, Hauck A, Song W, Lai J (2014). An ultra-high density bin-map for rapid QTL mapping for tassel and ear architecture in a large F₂ maize population. *BMC Genomics*, 15: 433
- Chen ZF, Yan W, Wang N, Zhang WH, Xie G, Lu J, Jian ZH, Liu DF, Tang XY (2014). Cloning of a rice male sterility gene by a modified MutMap method. *Hereditas*, 36 (1): 85–93 (in Chinese with English abstract) [陈竹锋, 严维, 王娜, 张文辉, 谢刚, 卢嘉威, 简智华, 刘东风, 唐晓艳(2014). 利用改进的MutMap方法克隆水稻雄性不育基因. *遗传*, 36 (1): 85–93]
- Cui XF (2013). Comparison of the major advances on plant biology between china domestic and abroad studies in 2008–2013. *Plant Physiol J*, 49 (6): 515–539 (in Chinese with English abstract) [崔晓峰(2013). 2008–2013年间植物生物学领域国内外重要研究进展比较. *植物生理学报*, 49 (6): 515–539]
- Das S, Upadhyaya HD, Bajaj D, Kujur A, Badoni S, Laxmi, Kumar V, Tripathi S, Gowda CLL, Sharma S, et al (2015). Deploying QTL-seq for rapid delineation of a potential candidate gene underlying major trait-associated QTL in chickpea. *DNA Res*, 22 (3): 4116–4118
- DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, Angel GD, Rivas MA, Hanna M, et al (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*, 43 (5): 491–498
- Fazio G, Staub JE, Stevens MR (2003). Genetic mapping and QTL analysis of horticultural traits in cucumber (*Cucumis sativus* L.) using recombinant inbred lines. *Theor Appl Genet*, 107 (5): 864–874
- Fekih R, Takagi H, Tamiru M, Abe A, Natsume S, Yaegashi H, Sharma S, Sharma S, Kanzaki H, Matsumura H, et al (2013). MutMap+: genetic mapping and mutant identification without crossing in rice. *PLoS One*, 8 (7): e68529
- Haase NJ, Beissinger T, Hirsch CN, Vaillancourt B, Deshpande S, Barry K, Buell CR, Kaeppeler SM, De LN (2014). Shared genomic regions between derivatives of a large segregating population of maize identified using bulked segregant analysis sequencing and traditional linkage analysis. *G3 (Bethesda)*, 5 (8): 1413–1428
- Hodges E, Xuan Z, Balija V, Kramer M, Molla MN, Smith SW, Middle CM, Rodesch MJ, Albert TJ, Hannon GJ, et al (2007). Genome-wide *in situ* exon capture for selective resequencing. *Nat Genet*, 39 (12): 1522–1527
- James GV, Patel V, Nordström KJ, Klasen JR, Salomé PA, Weigel D, Schneeberger K (2013). User guide for mapping-by-sequencing in *Arabidopsis*. *Genome Biol*, 14 (6): R61
- Illa-Berenguer E, Van Houten J, Huang Z, van der Knaap E (2015). Rapid and reliable identification of tomato fruit weight and locule number loci by QTL-seq. *Theor Appl Genet*, 128 (7): 1329–1342
- Ishihara T, Hayano-Saito Y, Oide S, Ebana K, La NT, Hayashi K, Ashizawa T, Suzuki F, Koizumi S (2014). Quantitative trait locus analysis of resistance to panicle blast in the rice cultivar *Miyazakimochi*. *Rice*, 7 (1): 2
- Katoh K, Standley DM (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*, 30 (4): 772–780
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment-Map format and

- SAMtools. *Bioinformatics*, 25 (16): 2078–2079
- Liu S, Yeh CT, Tang HM, Dan N, Schnable PS (2012). Gene mapping via bulked segregant RNA-Seq (BSR-Seq). *PLoS One*, 7 (5): e36406
- Lu H, Lin T, Klein J, Wang S, Qi J, Zhou Q, Sun J, Zhang Z, Weng Y, Huang S (2014). QTL-seq identifies an early flowering QTL located near *Flowering Locus T* in cucumber. *Theor Appl Genet*, 127 (7): 1491–1499
- Marçais G, Kingsford C (2011). A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. *Bioinformatics*, 27 (6): 764–770
- Mascher M, Jost M, Kuon JE, Himmelbach A, Abfalğ A, Beier S, Scholz U, Graner A, Stein N (2014). Mapping-by-sequencing accelerates forward genetics in barley. *Genome Biol*, 15 (6): 29–42
- Mascher M, Muehlbauer GJ, Rokhsar DS, Chapman J, Schmutz J, Barry K, Muñoz-Amatriain M, Close TJ, Wise RP, Schulman AH, et al (2013a). Anchoring and ordering NGS contig assemblies by population sequencing (POPSEQ). *Plant J*, 76 (4): 718–727
- Mascher M, Richmond TA, Gerhardt DJ, Himmelbach A, Clissold L, Sampath D, Ayling S, Steuernagel B, Pfeifer M, D’Ascenzo M, et al (2013b). Barley whole exome capture: a tool for genomic research in the genus *Hordeum* and beyond. *Plant J*, 76 (3): 494–505
- Minevich G, Park DS, Blankenberg D, Poole RJ, Hobert O (2012). CloudMap: a cloud-based pipeline for analysis of mutant genome sequences. *Genetics*, 192 (4): 1249–1269
- Nijman IJ, Mokry M, Van BR, Toonen P, De BE, Cuppen E (2010). Mutation discovery by targeted genomic enrichment of multiplexed barcoded samples. *Nat Methods*, 7 (11): 913–915
- Nordström KJ, Albani MC, James GV, Gutjahr C, Hartwig B, Turck F, Paszkowski U, Coupland G, Schneeberger K (2013). Mutation identification by direct comparison of whole-genome sequencing data from mutant and wild-type individuals using *k*-mers. *Nat Biotechnol*, 31 (4): 325–330
- Ogura T, Busch W (2014). From phenotypes to causal sequences: using genome wide association studies to dissect the sequence basis for variation of plant development. *Curr Opin Plant Biol*, 23 (2): 98–108
- Page JT, Liechty ZS, Huynh MD, Udall JA (2014). BamBam: genome sequence analysis tools for biologists. *BMC Res Notes*, 7: 829
- Pankin A, Campoli C, Dong X, Kilian B, Sharma R, Himmelbach A, Saini R, Davis SJ, Stein N, Schneeberger K, et al (2014). Mapping-by-sequencing identifies *HvPHYTOCHROME C* as a candidate gene for the early maturity 5 locus modulating the circadian clock and photoperiodic flowering in barley. *Genetics*, 198 (1): 383–396
- Ramirez-Gonzalez RH, Segovia V, Bird N, Fenwick P, Holdgate S, Berry S, Jack P, Caccamo M, Uauy C (2015). RNA-Seq bulked segregant analysis enables the identification of high-resolution genetic markers for breeding in hexaploid wheat. *Plant Biotechnol J*, 13 (5): 613–624
- Schneeberger K, Ossowski S, Lanz C, Juul T, Petersen AH, Nielsen KL, Jørgensen JE, Weigel D, Andersen SU (2009). SHOREmap: simultaneous mapping and mutation identification by deep sequencing. *Nat Methods*, 6 (8): 550–551
- Takagi H, Abe A, Yoshida K, Kosugi S, Natsume S, Mitsuoka C, Uemura A, Utsushi H, Tamiru M, Takuno S (2013a). QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *Plant J*, 74 (1): 174–183
- Takagi H, Uemura A, Yaegashi H, Tamiru M, Abe A, Mitsuoka C, Utsushi H, Natsume S, Kanzaki H, Matsumura H, et al (2013b). MutMap-Gap: whole-genome resequencing of mutant F₂ progeny bulk combined with *de novo* assembly of gap regions identifies the rice blast resistance gene *Pii*. *New Phytol*, 200 (1): 276–283
- Thyssen GN, Fang DD, Turley RB, Florane C, Li P, Naoumkina M (2015). Mapping-by-sequencing of *Ligon-lintless-1 (Li₁)* reveals a cluster of neighboring genes with correlated expression in developing fibers of Upland cotton (*Gossypium hirsutum* L.). *Theor Appl Genet*, 128 (9): 1703–1712
- Trick M, Adamski NM, Mugford SG, Jiang CC, Febrer M, Uauy C (2012). Combining SNP discovery from next-generation sequencing data with bulked segregant analysis (BSA) to fine-map genes in polyploid wheat. *BMC Plant Biol*, 12: 14
- Uchida N, Sakamoto T, Kurata T, Tasaka M (2011). Identification of EMS-induced causal mutations in a non-reference *Arabidopsis thaliana* accession by whole genome sequencing. *Plant Cell Physiol*, 52 (4): 716–722
- Xu X, Zeng L, Tao Y, Vuong T, Wan J, Boerma R, Noe J, Li Z, Finnerty S, Pathan SM, et al (2013). Pinpointing genes underlying the quantitative trait loci for root-knot nematode resistance in palaeopolyploid soybean by whole genome resequencing. *Proc Natl Acad Sci USA*, 110 (33): 13469–13474
- Yang J, Gao MX, Hu H, Ding XM, Lin HW, Wang L, Xu JM, Mao CZ, Zhao FJ, Wu ZC (2016). *OsCLT1*, a CRT-like transporter 1, is required for glutathione homeostasis and arsenic tolerance in rice. *New Phytol*, 211 (2): 658–670
- Yang Z, Huang D, Tang W, Zheng Y, Liang K, Cutler AJ, Wu W (2013). Mapping of quantitative trait loci underlying cold tolerance in rice seedlings via high-throughput sequencing of pooled extremes. *PLoS One*, 8 (7): e68433
- Zhu Y, Mang HG, Sun Q, Qian J, Hipps A, Hua J (2012). Gene discovery using mutagen-induced polymorphisms and deep sequencing: application to plant disease resistance. *Genetics*, 192 (1): 139–146
- Zuryn S, Le GS, Jamet K, Jarriault S (2010). A strategy for direct mapping and identification of mutations by whole-genome sequencing. *Genetics*, 186 (1): 427–430

The application of Mapping-by-sequencing applies in crops

YANG Lin, YANG Jian-Fei, WANG Huan-Huan, WANG Yu*, LI Yu-Hua*

College of Life Science, Northeast Forestry University, Harbin 150040, China

Abstract: In order to design better crops, forward genetic approach was usually used to screen loss-of-function genes related to important agronomic traits. With the repetitive DNA, multiple gene copies, chromosomal rearrangements and large complex polyploidy crop genomes can considerably complicate to discover mutated gene by conventional forward genetics approaches. With the next-generation sequencing platforms, MBS technology (Mapping-by-sequencing) is a powerful tools for rapid identification of causal mutations in plants. MBS technology achieves this by sequencing genomic DNA of plants pooled from an F₂ population based on Bulk segregant analysis (BSA). In this review we summarized the principle of Mapping-by-sequencing and the applications in rice, barley, wheat and maize for gene screening.

Key words: Mapping-by-sequencing (MBS); Bulk segregant analysis (BSA); genetic mapping; crops

Received 2016-11-21 Accepted 2017-02-22

This work was supported by the National Natural Science Foundation of China (Grant Nos. 31471911, 31272200 and 31401907) and Natural Science Foundation of Heilongjiang Province of China (Grant No. C2015052).

*Co-corresponding authors (E-mail: vonov@sina.com; lyhshen@126.com).