

高通量测序技术结合正向遗传学手段在基因定位研究中的应用

陆才瑞, 邹长松, 宋国立

中国农业科学院棉花研究所, 棉花生物学国家重点实验室, 安阳 455000

摘要: 传统的利用正向遗传学方法的基因定位一般是通过构建遗传连锁图谱进行的, 该过程步骤繁琐、耗时耗力, 很多情形下定位精确度低、区间大。随着高通量测序技术的快速发展以及测序成本的不断降低, 多种简单快捷的利用测序手段定位基因的方法被开发出来, 包括对突变体基因组直接测序定位、突变体材料构建混池测序定位和遗传分离群体测序构建图谱定位等, 还可以对转录组和部分基因组进行测序定位。这些方法可以在核苷酸水平鉴定突变位点, 并已推广到复杂的遗传背景中。近期报道的一些测序定位甚至是在不依赖于参考基因组序列、遗传杂交和连锁信息的情况下完成的, 这使得很多非模式物种也能开展正向遗传学研究。本文就这些新技术及其在基因定位中的应用进行了综述。

关键词: 高通量测序; 基因定位; 正向遗传学

Recent progress in gene mapping through high-throughput sequencing technology and forward genetic approaches

Cairui Lu, Changsong Zou, Guoli Song

State Key Laboratory of Cotton Biology, Institute of Cotton Research of Chinese Academy of Agricultural Sciences, Anyang 455000, China

Abstract: Traditional gene mapping using forward genetic approaches is conducted primarily through construction of a genetic linkage map, the process of which is tedious and time-consuming, and often results in low accuracy of mapping and large mapping intervals. With the rapid development of high-throughput sequencing technology and decreasing cost of sequencing, a variety of simple and quick methods of gene mapping through sequencing have been developed, including direct sequencing of the mutant genome, sequencing of selective mutant DNA pooling, genetic map construction through sequencing of individuals in population, as well as sequencing of transcriptome and partial genome. These methods can be used to identify mutations at the nucleotide level and has been applied in complex genetic background. Recent reports have shown that sequencing mapping could be even done without the reference of genome sequence, hybridization, and genetic linkage information, which made it possible to perform forward genetic study in many non-model species. In this review, we summarized these new technologies and their application in gene

收稿日期: 2015-02-04; 修回日期: 2015-04-27

基金项目: 国家自然科学基金项目(编号: 31271768, 31301369, 31401425)资助

作者简介: 陆才瑞, 博士, 研究方向: 棉花基因的定位与功能分析。E-mail: lucr99@126.com

通讯作者: 宋国立, 博士, 研究员, 研究方向: 棉花结构与功能基因组学研究。E-mail: sglzms@163.com

DOI: 10.16288/j.ycz.15-068

网络出版时间: 2015-5-8 9:15:16

URL: <http://www.cnki.net/kcms/detail/11.1913.R.20150508.0915.002.html>

mapping.

Keywords: high-throughput sequencing; gene mapping; forward genetic approaches

DNA 测序技术是现代生物学研究中重要的手段之一, 它能快速准确地获取生物体基因组上的遗传信息, 因而在生命科学研究中具有极其重要的作用。一代测序技术的标志是 1977 年 Sanger 等发明的双脱氧核苷酸末端终止法和 Gilbert 等发明的化学降解法, 它们具有简单方便、准确可靠、测序片段长等优点^[1]。此后的几十年中, 陆续发展了第二代测序技术, 其核心思想是边合成边测序(Sequencing by Synthesis), 即通过捕捉新合成的末端的标记来确定 DNA 的序列, 它具有一代测序技术不可比拟的通量高和成本低等特点, 主要技术平台有: 罗氏公司的 454 技术、Illumina 公司的 Solexa 技术、ABI 公司的 SOLID 技术^[2]。最近几年, 诞生了以单分子测序为特点的第三代测序技术, 它比二代测序技术的读长更长, 但也存在错误率较高和成本高的缺点, 因此未能得到广泛的应用。它的主要技术平台有: Helicos 公司的单分子 SMS(Single molecule sequencing)测序技术、Pacific Biosciences 公司的单分子实时(Single molecule real time, SMRT)测序技术和 Oxford Nanopore Technologies 公司的纳米孔单分子测序技术^[2]。目前测序领域应用最广泛的仍然是二代高通量测序技术。最近 10 年间, 它促进了基因组学、转录组学、比较基因组学、进化基因组学和反向遗传学等学科的飞速发展, 尤其在基因多态性、遗传定位、表达和调控等方面的作用更加突出。利用高通量测序技术对基因组测序、重测序和表达序列测序可为基因定位提供快速而准确的信息^[3]。

几十年来, 采用正向遗传学方法对突变体基因进行定位已成为日益强大的工具^[4,5]。该方法的基本原则是通过物理或化学诱变的方式随机产生各类突变体, 从中选择感兴趣的突变表型, 通过杂交构建重组作图群体, 然后利用遗传作图的方法将目标基因定位到一个较大的染色体区间, 随后在此区间内利用突变位点和遗传标记之间的重组关系精细定位候选基因。为进一步确认该基因的功能, 还需要鉴定该位点在突变体和野生型个体中基因型与表型的对应关系, 以及开展功能互补实验和相关的分子作用研究等^[6,7]。从理论上讲, 正向遗传学的研究方法

可适用于所有的生物种类, 但到目前为止却主要应用在模式生物的研究中。这主要是因为模式生物易于构建大规模的突变体库, 搭建用于检测和鉴定候选基因的基因组操作平台。相对于模式生物而言, 一般生物的正向遗传学研究仍然是一个步骤繁琐且耗时费力的过程^[8]。

遗传作图的分辨率是利用正向遗传学手段进行基因定位的关键因素, 突变位点所在区域的分辨率越高, 则定位的精度也越高。遗传图谱的分辨率主要取决于多态性标记的密度^[8], 随着植物基因组参考序列的释放^[9,10]以及全基因组范围分子标记的开发, 遗传图谱的构建已经得到了长足的发展^[11], 但对于精细定位而言依然存在很大的困难。二代测序技术的出现则使得区间定位与候选基因识别可直接通过测序完成^[12]。对于变异较少的材料而言, 突变位点自身提供了定位的标记, 可以通过测序直接获得候选基因^[13~16]。而通过构建遗传图谱定位突变位点的方法适用于几乎所有植物, 虽然大规模的作图群体逐个进行基因分型成本依然很高, 但在较小的基因组模式生物中已获得了一定的进展^[17~20]。BSA(Bulk segregant analysis)技术也已经能在节省成本的情况下, 对单基因控制的性状进行解析^[12,21], 甚至包括一些复杂性状^[22]。另外, 目标区域的靶向测序也能快速找到突变基因^[23~26]。本文将对这些研究方法的进展及其在一些物种中的应用进行综述。

1 突变体的创制

进行基因定位的首要条件是筛选含有目标基因的材料及其突变体, 在这些材料之间, 目标基因和遗传背景方面的差异对基因定位的复杂程度起着至关重要的作用。因而, 突变体库是功能基因组学研究中不可或缺的重要材料。大多数生物的突变体库构建工作都在其模式材料中开展, 主要通过辐射诱变、化学诱变和转座子插入等方法实现。快中子(Fast neutron)是一种高能量辐射源, 其诱变效果大多为多个碱基的缺失, 已应用于多种植物的突变体创制。在大豆的研究中, 利用快中子技术对 120 000 粒大豆种子进行快中子射线辐射诱导缺失突变体, 得到

了23 000个独立的M2突变株系^[27]。在花生的研究中,结合快中子和体外培养技术也建立了突变体资源库。研究者先将花生的干种子利用快中子辐射诱变,再分离胚芽嫩叶置于2,4-D和BAP中培养,诱导细胞胚胎发生和植株再生,通过进一步的实验验证该技术高效地创制了花生突变资源^[28]。N-甲基-N-亚硝基脲(MNU)是一种烷化剂,可以诱导生物体产生随机点突变。利用MNU对水稻开花过程中的合子细胞进行诱变后,整个基因组的突变率为每135 kb内含有1个突变碱基,在检测的基因区中突变率为每1 kb内含有7.4个突变碱基,且分布均匀^[29]。在百脉根的基因组研究中,LORE1反转座子系统被用于创制插入突变体,并通过二代测序技术对插入位置进行快速检测^[30]。

在突变体创制过程中,首先需要考虑的是突变的剂量效应。高剂量突变剂的使用可以在单个个体中得到更多的突变位点,但也增加了个体基因组中的突变载荷,因为靶基因可能与非等位的突变基因紧密连锁,这将导致突变基因鉴定过程的复杂化。其次,突变密度决定了获得目标表型的概率,它与突变类型息息相关。大片段的缺失或易位通常要比单碱基突变产生的密度低,但大范围的片段丢失或重排可能会破坏该区域内基因的连锁关系。在这种情况下,即便能确定突变基因的位置,也不清楚是哪些基因引起的表型变异。因此,筛选突变体的策略取决于诱变方法、实验物种以及其繁育系统和目标表型突变的显隐性等。

2 突变体基因组直接测序定位

与对照材料相比,自然变异和诱变产生的变异材料一般都含有成千上万的差异位点,这使得分离目标突变位点极其困难。然而在特定的条件下,如果产生的突变位点数目足够少,则可以直接对突变体基因组进行深度测序检测差异位点以找到对应突变表型的位点^[31~33]。这种方法主要用于基因组较小、复杂程度较低的低等生物中,如在枯草杆菌中,*relA*基因的缺失导致生长缺陷。研究人员对一个*relA*基因缺失却生长正常的突变菌株进行全基因组测序,以寻找相关突变位点^[13]。由于细菌的基因组很小,因此一共只找到10处突变位点。同时发现,稳定的突变表型需要其中两个基因同时发生突变,这样的

三基因互作在传统的作图工作中很难被鉴定。在酿酒酵母的乙醇耐受性研究中,突变体的全基因组序列与参考基因组相比也只有17个差异位点,去除参考序列的3个错误位点,最终得到14个候选突变位点^[14]。尽管直接对突变体基因组测序分析突变位点展示了其强大的功能,但是在大量的突变位点中直接识别突变依然很复杂,尤其是在化学诱变剂产生成千上万的突变位点情况下,因此最好采用适当的策略降低突变位点数量。

如果突变基因被定位在某一区域,那么就可以只考虑突变体基因组在这一区域的突变情况。在秀丽线虫的研究中,由于基因组较大,突变体基因组测序后与参考基因组序列比对会得到过多差异位点,因此研究者利用群体将突变基因定位到4 Mb区间内,在此区间内找到了16个能改变氨基酸的SNP,经RNAi实验验证发现,其中一个基因的突变对应表型变异^[15]。通过全基因组测序对绿藻的16个材料进行突变位点定位研究,基因组序列相互比对去除非突变相关SNP后,仍得到7092个差异位点,其中954个位于外显子区,即便去掉154个同义突变,数量仍然很庞大。用传统方法将突变位点定位于6.6 Mb的区间后,仅发现两处碱基突变,进一步的功能实验证明其中一处为突变基因^[16]。

对材料的处理有时也能极大地降低突变位点数目。拟南芥是具有相对较大基因组的模式植物,直接对原始突变系测序很难找到预期突变位点,研究者将突变体与野生型材料回交4代,从而去除了很多背景突变,降低了变异位点的数目。对野生型和回交后的突变体材料测序后,分别与参考基因组序列比对,发现突变体特异的SNP大多位于5号染色体,其中32个SNP为非同义突变,经T-DNA插入实验确定了候选基因^[34]。另外,如果突变体基因组自身具有较低的突变率,也可以通过对突变个体的直接测序快速鉴定突变位点。比如,由快中子辐射诱变产生的拟南芥突变体具有较少的突变位点,研究者通过对其M3代的基因组测序,找到108个突变位点,其中就包含导致突变表型的突变基因^[35]。

到目前为止,在所有直接对突变体测序而成功定位的报道中,绝大多数都是在低等生物中进行的,高等生物中只有在对拟南芥进行多代回交处理后获得成功,而在复杂的高等生物中应用这种方法很难

获得成功。另外,突变体与基因的编码区都有直接的联系,调控方面的突变仍然较少。如果已知突变位点的定位信息以及突变表型的遗传规律,直接对突变体进行基因组测序仍是一种明智的策略。

3 突变体混池测序定位

用于正向遗传学研究的重组作图群体中,具有突变体表型的个体携带突变基因。突变基因和其附近的遗传标记在重组群体中常常共同分离,距离突变基因越远则发生重组的概率越大,筛选具有突变体表型的个体意味着也同时选择了一系列与突变基因连锁的标记。利用基因分型技术可以通过检测重组群体中的等位基因频率,从而鉴定含有突变基因的区域并确定候选基因。这使得应用 DNA 混池(BSA)进行遗传作图可以大大简化基因分型劳动量(图 1)。

3.1 参考基因组材料为亲本的测序定位

全基因组测序数据在混池中进行作图和候选基因鉴定的方法在拟南芥中率先得到应用^[9],并且以

参考基因组材料直接作为亲本带来了极大的便利。研究者构建了由 EMS 诱导的隐性生长缺陷型突变株和参考基因组材料 Col-0 为亲本的 F₂ 作图群体,并将 500 个突变表型个体的 DNA 混合并测序。测序 reads 与参考基因组比对后,突变基因被定位于第 4 号染色体上一个很小的区间。距离该区间中心约 4 kb 的位置有一个错义突变,而第二个突变位点距离约 200 kb。因此,4 kb 位置的突变是无可争议的候选基因,实验证实该基因即为对应突变表型的调控基因^[12]。采用较少的突变体构建混池也能达到定位的目的。在拟南芥的另一项研究中,对 93 个 F₂ 突变体构建的混池进行测序,通过比较混池与两个亲本间基因组序列的 SNP 比率差异,在第 2 号染色体上得到一个候选区间,实验证实其中的一个 SNP 为突变基因^[21]。这种简单的方法迅速被应用到其他隐性突变体中^[36~38],包括拟南芥的突变体^[39]和其他物种,如酵母细胞、秀丽隐杆线虫、粗糙脉孢菌、小鼠和斑马鱼等^[40~46]。但用突变体与其他材料杂交也存在一定的困难,因为不同的遗传背景导入会在群体中改变或影响突变表型^[47~49]。

为了克服亲本材料之间存在的大量差异带来的复杂性问题,可以用同样遗传背景的材料作为亲本构建群体^[48,50~53]。在水稻的研究中开发了一种 MutMap 的方法^[48],利用参考基因组材料作为亲本,另一个亲本则是由上述亲本通过诱变筛选后多代自交而来。因此,两个亲本间遗传背景极相似,避免了其他背景带来的影响。研究者对两个叶片表现为苍白绿的突变体进行定位,各取 20 个 F₂ 突变株构建混池测序,与参考基因组序列比对,分别在第 10 号和 1 号染色体上发现了突变体特异的 SNP 区域,进一步实验鉴定了其中对应突变表型的 SNP^[48]。这种方法也被迅速应用到其他突变体的研究中^[54,55]。为了更进一步去除非突变位点的影响,用突变体与参考基因组材料多代回交后,选择具有突变表型的后代作为亲本材料,这样可以在保留突变位点的基础上,极大地减少两个亲本之间其他变异位点的数目^[56]。值得注意的是,用这种类似于近等基因系的材料进行遗传定位必须使用高通量测序技术,而无法使用以自然多态性为标记的传统基因分型技术。

3.2 非参考基因组材料为亲本的测序定位

尽管以参考基因组材料作为亲本提供了优质的

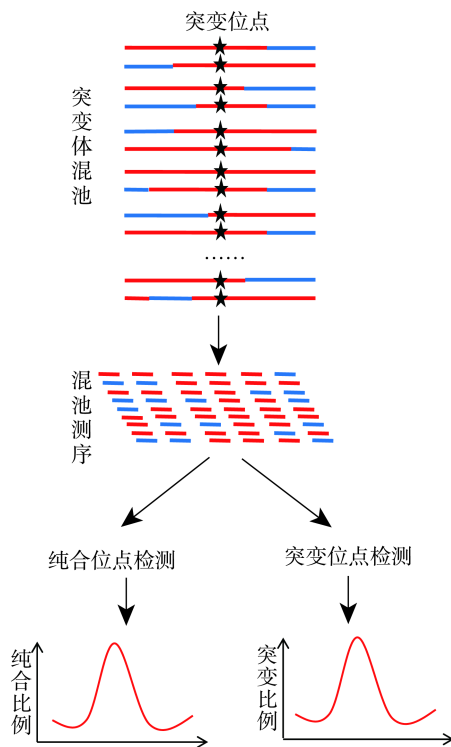


图 1 利用混池测序定位的基本原理

突变体混池中,由于选择的原因突变位点及其附近的位点理论上是纯合的,其他位点则是杂合的,因而纯合度或者突变率极高的区域含有突变位点。

参考序列,但由于遗传背景的原因,很多突变位点无法在这些材料中进行研究,而必须使用非参考基因组材料作为亲本进行测序定位的研究。

在拟南芥研究中率先开发了 NGM(Next-generation mapping)方法^[57]。该方法把突变体与非参考基因组材料杂交,挑选 F_2 群体中的突变体构建混池测序,得到的 reads 与参考基因组序列比对。理论上突变位点附近区域在混池中只来自于其中一个亲本,属于相对纯合区域,与参考序列相比只包含一个亲本的差异位点,而其他区域则包含两个亲本与参考序列的所有差异。因此,检测 SNP 最少的区域很可能对应着突变基因所在的区域^[57-59]。利用全基因组数据的 K-mer 值也可以计算纯合突变位点^[53]。它借用了基因组拼接过程中的 K-mer 原理,即:测序错误序列的 K-mer 值一般较低,因为这些 reads 匹配到参考序列上的比例较少,据此可以把它们过滤掉。在测序定位过程中,如果两个材料或 DNA 混池的遗传背景极其相似,当它们相互比对时,可以找到那些 K-mer 值很低的区域,它们在 2 个混池中位置相同、差异很小,这即是样品特异的突变位点区域。在高山南芥的研究中,应用此方法鉴定了 16 个 *fde1* 和 13 个 *pep1-1* 的突变位点,进一步实验证实突变基因包含在这 29 个位点中^[53]。

杂合 SNP 与纯合 SNP 之间的比例在利用非参考基因组材料定位突变基因过程中也起着重要的作用。研究者利用拟南芥及其突变体构建了 F_2 群体,对包含 88 个 F_2 突变个体的混池进行深度测序并与参考序列比对,发现 1、2、3、5 染色体中大多为杂合 SNP,只有 4 号染色体含有大量的突变体基因型纯合 SNP,这种连锁不平衡意味着突变基因很可能在 4 号染色体上。通过检测每 50 kb 区间内杂合 SNP 与纯合 SNP 的比例,发现在 6850~9100 kb 区间内存在一个纯合 SNP 峰,突变基因被认为存在于该区间内。去除背景相关的 SNP 后,筛选那些存在于该区间的基因区,并且表现为 G-A 或 C-T 碱基突变的 SNP,共得到 42 个。这个数字依然很大,又对另外一个表型一致的突变体采用同样的流程进行筛选,得到 52 个 SNP。而二者共同位于同一个基因内的 SNP 只有 1 个,经实验验证该基因即为突变基因^[39]。

另外,将野生型和突变体混池与参考序列比较,可以剔除大量的背景变异位点。在拟南芥中,将野

生型亲本和 F_3 的矮秆突变体混池分别测 $25\times$,二者都与拟南芥的参考序列进行比对,在混池中发现 4023 个高质量的碱基差异,涉及 63 个基因的 531 个 SNP^[60]。除了 1 个 SNP 以外,其他 SNP 都能在野生型的亲本材料与 Col-0 中检测到。这个 1 bp 的丢失位于 *Atlg58440* 的第 7 个外显子上,T-DNA 插入实验证实该基因突变导致类似的表型。利用 BC_1F_2 群体的 110 株突变体混池测序对突变位点进行定位的实验中,对非突变亲本的深度测序也排除了大量非突变相关的变异位点,最终得到 6 个 SNP,其中 1 个在 *HASTY* 基因中产生了提前终止子^[61]。

4 基于群体测序构建图谱定位

突变体及其混池的测序在单个基因控制性状的定位中表现出强大的功能,不仅降低了成本和工作量,而且提高了效率和准确性。然而对于多个基因控制的复杂数量性状来说,由于其表型在群体中呈连续分布,没有明显的表型分离比例,因而使用这些技术很难达到对 QTL 准确高效的定位。数量性状基因的定位需要对作图群体的每一个个体分别进行测序,逐个检测其基因型,进而通过构建连锁图谱进行 QTL 定位^[62-68]。

在水稻研究中,利用具有参考基因组序列的日本晴为亲本构建重组自交系,对 150 个重组自交系单株通过 bar-coded 策略测序,每个单株平均测 $0.02\times$ ^[17]。单株序列与亲本比对获得 SNP,频率为每 40 kb 含有 1 个 SNP,继而构建遗传图谱。将临近的 100 kb 范围内基因型相同的位点作为一个 bin,得到了 2334 个 bin,平均长度为 164 kb,其中最长达 5.8 Mb。随后,对株高进行了 QTL 定位,找到 4 个 QTLs,其中解释最大表型变异率的区间位于第 1 号染色体的 40.1~40.2 Mb 之间,其中含有水稻“绿色革命”基因。在利用 LYP9 和 93-11 构建的重组自交系群体中,选择 132 个单株,每个测 $4\times$,两个亲本分别测 $48\times$ 和 $36\times$ ^[18]。与日本晴参考序列比对,过滤后得到 171 847 个高质量 SNP,进而构建高密度遗传图谱。分析获得 43 个产量相关 QTLs,其中 20 个 QTLs 是新发现的。同时,qSN8 和 qSPB1 被精细定位,各自得到一个候选基因并互补验证。

在缺乏父母本基因型的情况下,对群体的测序也可以进行定位。在水稻研究中利用极大似然法构建

亲本的基因型草图,再通过贝叶斯方法获得高质量的 SNPs,最后结合隐马尔可夫模型对群体单株分型,从而构建遗传图谱^[19]。该实验使用了 238 个重组自交系单株,每个单株测 0.05 ×,最终将数量性状谷粒宽度定位在 200 kb 的区间内。在对大豆 246 个重组自交系个体利用高通量测序方法构建遗传图谱时,每个个体测序 0.19 ×,并开发了一个软件流程来推断亲本的基因型^[20]。最终获得了 3 个抗病相关的 QTLs,其中一个主要 QTL 位于第 10 号染色体上,区间大小 29.7 kb,含有 3 个基因。结构和表达差异分析显示其中 2 个基因是抗病 QTL 的候选基因。

鉴于群体测序的成本较高,一些研究者选择使用限制性位点相关的 DNA 分析(Restriction-site-associated DNA, RAD)技术^[69],该技术选择性地对特定酶切位点附近序列测序,进而达到简化基因组的目的。由于 RAD 技术没有明确地富集基因,包括实际突变位点的可能性很低,因而还需要进一步的精确定位工作^[69]。RAD-seq 数据的低复杂性使其可以用于没有参考基因组序列的分析,有利于在非模式生物中开展正向遗传研究。当然,依赖于参考序列的分析通常是更好的,因为它们可以利用 reads 比对到参考序列的顺序所提供的连锁信息。

5 基于转录组测序定位

全基因组的测序定位提供了全面而完整的基因分型信息,但对于一些较大的基因组来说,其复杂性以及较高的测序费用阻碍了这种方法的应用。由于绝大多数突变都与基因有关,因此直接分析 RNA 能避开基因组的复杂性,且相对于基因组数据而言,转录组的测序量显著减少,从而降低了费用。另外,使用转录组数据不仅能识别基因的突变位点,还能确定基因的表达水平和选择性剪接对于突变表型的影响。

在玉米的研究中,使用 RNA-seq 数据为基础对一个能改变嫩叶中蜡累积的突变(*gl3*)进行测序定位^[70]。研究者用 *gl3* 突变体构建了 F₂ 群体,各选 30 株突变表型个体与野生表型个体构建混池,测序后比对到亲本之一的 B73 玉米参考基因组序列上,分析 SNP 位点在基因组上的等位基因频率分布。理论上突变位点附近区域的 SNP 是纯合的,距离该位置越远,则杂合度越高。最终突变基因被定位到~2 Mb 的区

间内。通过转座子插入实验获得了相同表型的突变体,同时得到两个候选基因,其中一个候选基因的表达量在突变体混池中显著下调,证实该基因的突变导致了突变体表型^[70]。

在果蝇研究中,利用转录组测序定位开发了 MMAPPR 方法^[71]。该方法利用 F₂ 群体构建包含 20 个野生型个体和突变体型个体的两个混池,并进行 3 次重复,以确定突变基因的连锁区域、表达情况以及对蛋白翻译的影响。MMAPPR 方法在判断定位区间时,可以不考虑 RNA 取样的组织和时间,甚至在突变基因不表达的情况下也能判定。该研究用 4 个突变基因验证了此方法的可行性。其中, *nkx2.5* 和 *tbx1* 都在 0.5 Mb 范围内鉴定到提前终止子,与该突变体已知的突变基因相符。*zy13* 的突变则是由候选基因 *ctr9* 中的无义突变造成的,注射野生型的 *ctr9* 能修复突变体的表型。*zy14* 被定位在 6.6 Mb 内,其突变是由表达量的降低造成的,注射野生型的 *cds2* 也能修复突变体的表型^[71]。

构建混池所用的个体数目对定位结果有明显的影响。在斑马鱼的研究中,4 个已鉴定的突变基因被用来确定转录组混池测序在基因定位中的效果^[72]。F₂ 群体中选择同等数目的突变体和野生型构建混池,测序后与参考基因组进行比对,得到约 4 万个高质量的 SNP,以 50 个 SNP 为滑动窗口在突变体混池中检测 SNP 频率的变化差异。以 *hoxb1b* 基因为例,由 20、40、80 个个体构建的混池得到的连锁区域大小分别为 11.6 Mb、7.9 Mb 和 6.5 Mb。*vangl2* 和 *egr2b* 基因分别使用 37 个和 30 个个体构建混池,得到 2.3 Mb 和 6.6 Mb 的连锁区域。*nhs11b* 基因只有 8 个突变个体,得到约 40 Mb 的连锁区域。结果表明,混池个体数量越多,定位的区间就越小,同时基因所在的位置对定位效果的影响也很明显。另外,此研究还表明转录组测序可以鉴定多种形式的突变。*hoxb1b* 和 *nhs11b* 基因的突变形式是无义突变,而在 *vangl* 基因中,选择性剪接的改变导致了表型的突变,*egr2b* 基因则是表达水平的差异导致了表型变异^[72]。

值得注意的是,虽然突变体池和野生型池之间的表达差异有助于预测候选基因,但差异表达的基因一般不能作为初选的候选基因^[70](图 2)。即使突变可能影响基因的编码区,但也不一定会改变其表达。相反,突变基因可能对其下游的基因表达产生严重

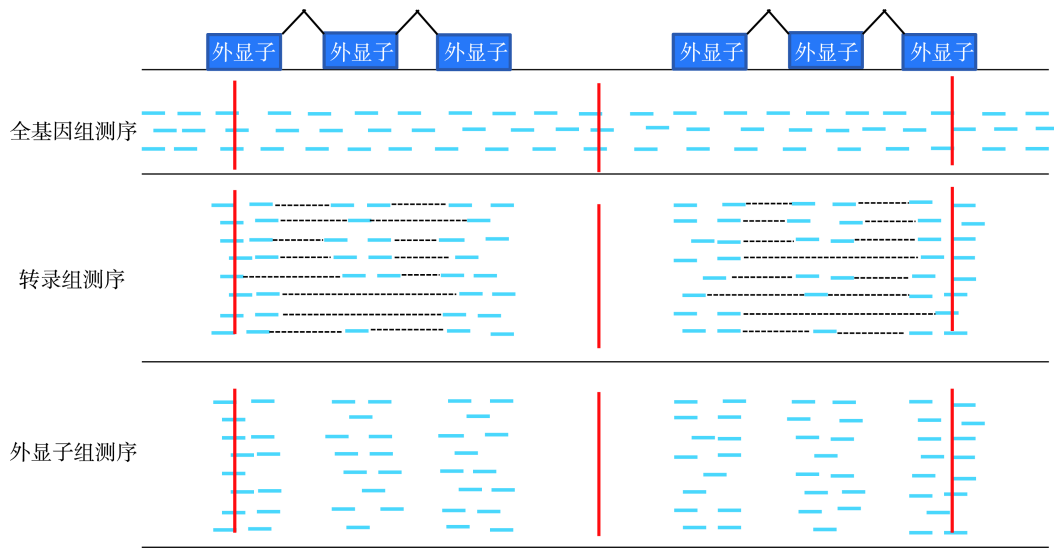


图 2 不同测序数据的利弊

红线表示突变位点。全基因组测序得到整个基因组的随机数据,覆盖度广,缺点是一般深度不高。转录组测序得到表达序列,问题在于突变位点必须表达才能被鉴定。外显子测序得到基因组上的基因序列,缺点是基因间区的突变位点很难被鉴定。

影响。以转录组数据为基础的测序定位的最大问题在于,突变位点可能不存在于转录组中,比如位于调节元件中的突变。并且在有些情况下,无法确定突变基因在什么组织和时间中表达。此时,可以结合参考基因组序列,通过其附近基因的变异指导候选定位区域的分析。

6 部分基因组的测序

到目前为止,全基因组测序仍然是提供完整定位信息的最佳途径,这种方法通常需要一个高质量、较完整的参考基因组序列。然而,许多重要作物的基因组没有测序或不完整。并且,对于较大和重复性高的基因组而言,测序费用也阻碍了该方法的发展。如果后续的分析只集中于基因的突变,可能不需要在全基因组范围内筛选。在这种情况下,外显子组测序^[73~75]或定位区间 DNA 的富集测序^[23~26]是非常有效的方法,富集的混池 DNA 甚至可以降低基因组的复杂度,同时进行测序定位^[76](图 2)。

由于大麦的基因组很大(5 Gb),且重复序列很多,很难将其组装完整。因而,也很难通过对基因组测序的方法来定位基因。研究者利用大麦基因组的注释基因、全长 cDNA 和转录组组装序列作为探针定制了 61.6 Mb 的大麦外显子组芯片^[77]。通过此芯片捕获大麦的外显子组序列并测序,对 X 光诱变产生

的隐性大麦多节矮秆突变体进行基因定位^[78]。在 F_2 代中取 18 个突变体和 30 个野生型个体构建 DNA 混池,分别用大麦外显子组探针富集编码序列并测序。两个混池的测序数据与大麦的参考基因组序列比对,分析 SNP 位点及其等位基因频率,在染色体 5H 长臂发现一个 30 cM 区域,该区域内突变型的 SNP 等位基因频率达 95%,而野生型的 SNP 频率仅为 30%。由于 X 光引发的突变多为缺失突变,在测序数据中寻找 >150 bp 的缺失突变,发现有 18 个缺失突变位于上述区域,其中一个基因与水稻 *pla1* 基因相似,而 *pla1* 基因的突变会使得水稻有类似大麦多节矮秆突变体的表型,后经证实该基因导致了突变表型^[78]。可见,利用突变体资源,结合外显子组测序、合理数据分析方法以及现有的基因组资源,可以通过测序法对大型基因组物种进行经济有效的基因定位^[79, 80]。

有些突变基因已经被定位到某一区间,但由于多态性分子标记的缺乏,使用传统遗传学手段很难精细定位到基因水平。在这种情况下,定位区间 DNA 的富集测序对于开发 SNP 标记和确定候选基因具有重要意义。在果蝇的研究中,基于 *twr* 基因的初定位结果,对 3R 染色体约 460 kb 区间设计探针抓取该位置的突变体基因组序列,与参考基因组序列进行比较,发现只有 26 个位点能导致氨基酸的改变^[24]。其中 24 个位点遗传自其祖先种,实验验证其与突变

无关,而另外两个位点中的一个被确定为突变基因。

7 展望

应用正向遗传学手段筛选突变基因是将表型与其潜在的基因联系起来的最成功的策略之一。应用该方法进行基因定位主要依赖于重组事件,但对于那些重组频率很低甚至完全被抑制的区域则缺乏有效的办法。二代测序技术的出现使得突变位点的识别可以通过直接比较突变体和野生型基因组而实现,从而将突变位点的鉴定从对重组事件的依赖中解放出来。目前,由于混池测序分析可以容易且廉价地处理大量的样本,所以对于很多物种的研究此方法仍然是比较好的选择。在未来几年内,测序成本将逐步降低,这使得对每一个突变个体进行测序而开展正向遗传学的研究成为可能。尽管这种定位结果不会明确地说明哪个候选基因对应突变表型,但随后有针对性的补充实验或转基因互补实验可以快速验证这些基因的功能。鉴于高通量的性质和二代测序的成本效应,基于测序的方法进行突变位点鉴定即将代替重组遗传作图。突变位点的鉴定也将很快脱离模式生物系统的复杂性,并可以在不考虑参考序列、栽培品系和物种的情况下定位基因。

参考文献

- [1] 于军. “人类基因组计划”回顾与展望: 从基因组生物学到精准医学. 自然杂志, 2013, 35(5): 326–331. [DOI]
- [2] 柳延虎, 王璐, 于黎. 单分子实时测序技术的原理与应用. 遗传, 2015, 37(3): 259–268. [DOI]
- [3] Ogura T, Busch W. From phenotypes to causal sequences: using genome wide association studies to dissect the sequence basis for variation of plant development. *Curr Opin Plant Biol*, 2015, 23: 98–108. [DOI]
- [4] Page DR, Grossniklaus U. The art and design of genetic screens: *Arabidopsis thaliana*. *Nat Rev Genet*, 2002, 3(2): 124–136. [DOI]
- [5] Candela H, Hake S. The art and design of genetic screens: maize. *Nat Rev Genet*, 2008, 9(3): 192–203. [DOI]
- [6] Patton EE, Zon LI. The art and design of genetic screens: zebrafish. *Nat Rev Genet*, 2001, 2(12): 956–966. [DOI]
- [7] Kile BT, Hilton DJ. The art and design of genetic screens: mouse. *Nat Rev Genet*, 2005, 6(7): 557–567. [DOI]
- [8] Peters JL, Cnudde F, Gerats T. Forward genetics and map-based cloning approaches. *Trends Plant Sci*, 2003, 8(10): 484–491. [DOI]
- [9] Wang KB, Wang ZW, Li FG, Ye WW, Wang JY, Song GL, Yue Z, Cong L, Shang HH, Zhu SL, Zou CS, Li Q, Yuan YL, Lu CR, Wei HL, Gou CY, Zheng ZQ, Yin Y, Zhang XY, Liu K, Wang B, Song C, Shi N, Kohel RJ, Percy RG, Yu JZ, Zhu YX, Wang J, Yu SX. The draft genome of a diploid cotton *Gossypium raimondii*. *Nat Genet*, 2012, 44(10): 1098–1103. [DOI]
- [10] Li FG, Fan GY, Wang KB, Sun FM, Yuan YL, Song GL, Li Q, Ma ZY, Lu CR, Zou CS, Chen WB, Liang XM, Shang HH, Liu WQ, Shi CC, Xiao GH, Gou CY, Ye WW, Xu X, Zhang XY, Wei HL, Li ZF, Zhang GY, Wang JY, Liu K, Kohel RJ, Percy RG, Yu JZ, Zhu YX, Wang J, Yu SX. Genome sequence of the cultivated cotton *Gossypium arboreum*. *Nat Genet*, 2014, 46(6): 567–572. [DOI]
- [11] Lu CR, Zou CS, Zhang YP, Yu DQ, Cheng HL, Jiang PF, Yang WC, Wang QL, Feng XX, Prosper MA, Guo XP, Song GL. Development of chromosome-specific markers with high polymorphism for allotetraploid cotton based on genome-wide characterization of simple sequence repeats in diploid cottons (*Gossypium arboreum* L. and *Gossypium raimondii* Ulbrich). *BMC Genomics*, 2015, 16(1): 55. [DOI]
- [12] Schneeberger K, Ossowski S, Lanz C, Juul T, Petersen AH, Nielsen KL, Andersen SU. SHOREmap: simultaneous mapping and mutation identification by deep sequencing. *Nat Methods*, 2009, 6(8): 550–551. [DOI]
- [13] Srivatsan A, Han Y, Peng J, Tehranchi AK, Gibbs R, Wang JD, Chen R. High-precision, whole-genome sequencing of laboratory strains facilitates genetic studies. *PLoS Genet*, 2008, 4(8): e1000139. [DOI]
- [14] Smith DR, Quinlan AR, Peckham HE, Makowsky K, Tao W, Woolf B, Shen L, Donahue WF, Tusneem N, Stromberg MP, Stewart DA, Zhang L, Ranade SS, Warner JB, Lee CC, Coleman BE, Zhang Z, McLaughlin SF, Malek JA, Sorenson JM, Blanchard AP, Chapman J, Hillman D, Chen F, Rokhsar DS, McKernan KJ, Jeffries TW, Marth GT, Richardson PM. Rapid whole-genome mutational profiling using next-generation sequencing technologies. *Genome Res*, 2008, 18(10): 1638–1642. [DOI]
- [15] Sarin S, Prabhu S, O'Meara MM, Pe'er I, Hobert O. *Caenorhabditis elegans* mutant allele identification by whole-genome sequencing. *Nat Methods*, 2008, 5(10): 865–867. [DOI]
- [16] Lin H, Miller ML, Granas DM, Dutcher SK. Whole genome sequencing identifies a deletion in protein phosphatase 2A that affects its stability and localization in *Chlamydomonas reinhardtii*. *PLoS Genet*, 2013, 9(9): e1003841. [DOI]

- [17] Huang XH, Feng Q, Qian Q, Zhao Q, Wang L, Wang A, Guan JP, Fan DL, Weng QJ, Huang T, Dong GJ, Sang T, Han B. High-throughput genotyping by whole-genome resequencing. *Genome Res*, 2009, 19(6): 1068–1076. [DOI]
- [18] Gao ZY, Zhao SC, He WM, Guo LB, Peng YL, Wang JJ, Guo XS, Zhang XM, Rao YC, Zhang C, Dong GJ, Zheng FY, Lu CX, Hu J, Zhou Q, Liu HJ, Wu HY, Xu J, Ni PX, Zeng DL, Liu DH, Tian P, Gong LH, Ye C, Zhang GH, Wang J, Tian FK, Xue DW, Liao Y, Zhu L, Chen MS, Li JY, Cheng SH, Zhang GY, Wang J, Qian Q. Dissecting yield-associated loci in super hybrid rice by resequencing recombinant inbred lines and improving parental genome sequences. *Proc Natl Acad Sci USA*, 2013, 110(35): 14492–14497. [DOI]
- [19] Xie WB, Feng Q, Yu HH, Huang XH, Zhao Q, Xing YZ, Yu SB, Han B, Zhang QF. Parent-independent genotyping for constructing an ultrahigh-density linkage map based on population sequencing. *Proc Natl Acad Sci USA*, 2010, 107(23): 10578–10583. [DOI]
- [20] Xu XY, Zeng L, Tao Y, Vuong T, Wan JR, Boerma R, Noe J, Li ZL, Finnerty S, Pathan SM, Shannon JG, Nguyen HT. Pinpointing genes underlying the quantitative trait loci for root-knot nematode resistance in palaeopolyploid soybean by whole genome resequencing. *Proc Natl Acad Sci USA*, 2013, 110(33): 13469–13474. [DOI]
- [21] Cuperus JT, Montgomery TA, Fahlgren N, Burke RT, Townsend T, Sullivan CM, Carrington JC. Identification of *MIR390a* precursor processing-defective mutants in *Arabidopsis* by direct genome sequencing. *Proc Natl Acad Sci USA*, 2010, 107(1): 466–471. [DOI]
- [22] Wolyn DJ, Borevitz JO, Loudet O, Schwartz C, Maloof J, Ecker JR, Berry CC, Chory J. Light-response quantitative trait loci identified with composite interval and eXtreme array mapping in *Arabidopsis thaliana*. *Genetics*, 2004, 167(2): 907–917. [DOI]
- [23] Gupta T, Marlow FL, Ferriola D, Mackiewicz K, Dapprich J, Monos D, Mullins MC. Microtubule actin crosslinking factor 1 regulates the Balbiani body and animal-vegetal polarity of the zebrafish oocyte. *PLoS Genet*, 2010, 6(8): e1001073. [DOI]
- [24] Wang H, Chattopadhyay A, Li Z, Daines B, Li YM, Gao CX, Gibbs R, Zhang K, Chen R. Rapid identification of heterozygous mutations in *Drosophila melanogaster* using genomic capture sequencing. *Genome Res*, 2010, 20(7): 981–988. [DOI]
- [25] Bontems F, Baerlocher L, Mehenni S, Bahechar I, Fari-nelli L, Dosch R. Efficient mutation identification in zebrafish by microarray capturing and next generation sequencing. *Biochem Biophys Res Commun*, 2011, 405(3): 373–376. [DOI]
- [26] Sun M, Mondal K, Patel V, Horner VL, Long AB, Cutler DJ, Casparly T, Zwick ME. Multiplex chromosomal exome sequencing accelerates identification of ENU-induced mutations in the mouse. *G3 (Bethesda)*, 2012, 2(1): 143–150. [DOI]
- [27] Bolon YT, Haun WJ, Xu WW, Grant D, Stacey MG, Nelson RT, Gerhardt DJ, Jeddeloh JA, Stacey G, Muehlbauer GJ, Orf JH, Naeve SL, Stupar RM, Vance CP. Phenotypic and genomic analyses of a fast neutron mutant population resource in soybean. *Plant Physiol*, 2011, 156(1): 240–253. [DOI]
- [28] Wang JS, Sui JM, Xie YD, Guo HJ, Qiao LX, Zhao LL, Yu SL, Liu LX. Generation of peanut mutants by fast neutron irradiation combined with *in vitro* culture. *J Radiat Res*, 2015, pii: rru121. [DOI]
- [29] Suzuki T, Eiguchi M, Kumamaru T, Satoh H, Matsusaka H, Moriguchi K, Nagato Y, Kurata N. MNU-induced mutant pools and high performance TILLING enable finding of any gene mutation in rice. *Mol Genet Genomics*, 2008, 279(3): 213–223. [DOI]
- [30] Fukai E, Malolepszy A, Sandal N, Hayashi M, Andersen SU. Forward and reverse genetics: the *LORE1* retrotransposon insertion mutants. In: Tabata S, Stougaard J, eds. *The Lotus japonicus Genome*. Berlin Heidelberg: Springer, 2014: 221–227. [DOI]
- [31] Schierenbeck L, Ries D, Rogge K, Grewe S, Weisshaar B, Kruse O. Fast forward genetics to identify mutations causing a high light tolerant phenotype in *Chlamydomonas reinhardtii* by whole-genome-sequencing. *BMC Genomics*, 2015, 16(1): 57. [DOI]
- [32] Katoh H, Miyata S, Inoue H, Iwanami T. Unique features of a Japanese ‘Candidatus Liberibacter asiaticus’ strain revealed by whole genome sequencing. *PLoS One*, 2014, 9(9): e106109. [DOI]
- [33] Carviel JL, Wilson DC, Isaacs M, Carella P, Catana V, Golding B, Weretilnyk EA, Cameron RK. Investigation of intercellular salicylic acid accumulation during compatible and incompatible *arabidopsis-pseudomonas syringae* interactions using a fast neutron-generated mutant allele of *EDS5* identified by genetic mapping and whole-genome sequencing. *PLoS One*, 2014, 9(3): e88608. [DOI]
- [34] Ashelford K, Eriksson ME, Allen CM, D’Amore R, Johansson M, Gould P, Kay S, Millar AJ, Hall N, Hall A. Full genome re-sequencing reveals a novel circadian clock

- mutation in *Arabidopsis*. *Genome Biol*, 2011, 12(3): R28. [DOI]
- [35] Belfield EJ, Gan XC, Mithani A, Brown C, Jiang CF, Franklin K, Alvey E, Wibowo A, Jung M, Bailey K, Kalwani S, Ragoussis J, Mott R, Harberd NP. Genome-wide analysis of mutations in mutant lineages selected following fast-neutron irradiation mutagenesis of *Arabidopsis thaliana*. *Genome Res*, 2012, 22(7): 1306–1315. [DOI]
- [36] Manavella PA, Hagmann J, Ott F, Laubinger S, Franz M, Macek B, Weigel D. Fast-forward genetics identifies plant CPL phosphatases as regulators of miRNA processing factor HYL1. *Cell*, 2012, 151(4): 859–870. [DOI]
- [37] Schierenbeck L, Ries D, Rogge K, Grewe S, Weisshaar B, Kruse O. Fast forward genetics to identify mutations causing a high light tolerant phenotype in *Chlamydomonas reinhardtii* by whole-genome-sequencing. *BMC Genomics*, 2015, 16(1): 57. [DOI]
- [38] Billoud B, Jouanno É, Nehr Z, Carton B, Rolland É, Chenivresse S, Charrier B. Localization of causal locus in the genome of the brown macroalga *Ectocarpus*: NGS-based mapping and positional cloning approaches. *Front Plant Sci*, 2015, 6: 68. [DOI]
- [39] Uchida N, Sakamoto T, Kurata T, Tasaka M. Identification of EMS-induced causal mutations in a non-reference *Arabidopsis thaliana* accession by whole genome sequencing. *Plant Cell Physiol*, 2011, 52(4): 716–722. [DOI]
- [40] Wenger JW, Schwartz K, Sherlock G. Bulk segregant analysis by high-throughput sequencing reveals a novel xylose utilization gene from *Saccharomyces cerevisiae*. *PLoS Genet*, 2010, 6(5): e1000942. [DOI]
- [41] Doitsidou M, Poole RJ, Sarin S, Bigelow H, Hobert O. *C. elegans* mutant identification with a one-step whole-genome-sequencing and SNP mapping strategy. *PLoS One*, 2010, 5(11): e15435. [DOI]
- [42] Pomraning KR, Smith KM, Freitag M. Bulk segregant analysis followed by high-throughput sequencing reveals the *Neurospora* cell cycle gene, *ndc-1*, to be allelic with the gene for ornithine decarboxylase, *spe-1*. *Eukaryot Cell*, 2011, 10(6): 724–733. [DOI]
- [43] Leshchiner I, Alexa K, Kelsey P, Adzhubei I, Austin-Tse CA, Cooney JD, Anderson H, King MJ, Stottmann RW, Garnaas MK, Ha S, Drummond IA, Paw BH, North TE, Beier DR, Goessling W, Sunyaev SR. Mutation mapping and identification by whole-genome sequencing. *Genome Res*, 2012, 22(8): 1541–1548. [DOI]
- [44] Voz ML, Coppieters W, Manfroid I, Baudhuin A, Von Berg V, Charlier C, Meyer D, Driever W, Martial JA, Peers B. Fast homozygosity mapping and identification of a zebrafish ENU-induced mutation by whole-genome sequencing. *PLoS One*, 2012, 7(4): e34671. [DOI]
- [45] Bowen ME, Henke K, Siegfried KR, Warman ML, Harris MP. Efficient mapping and cloning of mutations in zebrafish by low-coverage whole-genome sequencing. *Genetics*, 2012, 190(3): 1017–1024. [DOI]
- [46] Obholzer N, Swinburne IA, Schwab E, Nechiporuk AV, Nicolson T, Megason SG. Rapid positional cloning of zebrafish mutations by linkage and homozygosity mapping using whole-genome sequencing. *Development*, 2012, 139(22): 4280–4290. [DOI]
- [47] Greenberg MVC, Ausin I, Chan SWL, Cokus SJ, Cuperus JT, Feng SH, Law JA, Chu C, Pellegrini M, Carrington JC, Jacobsen SE. Identification of genes required for *de novo* DNA methylation in *Arabidopsis*. *Epigenetics*, 2011, 6(3): 344–354. [DOI]
- [48] Abe A, Kosugi S, Yoshida K, Natsume S, Takagi H, Kanzaki H, Matsumura H, Yoshida K, Mitsuoka C, Tamiru M, Innan H, Cano L, Kamoun S, Terauchi R. Genome sequencing reveals agronomically important loci in rice using MutMap. *Nat Biotechnol*, 2012, 30(2): 174–178. [DOI]
- [49] Hartwig B, James GV, Konrad K, Schneeberger K, Turck F. Fast isogenic mapping-by-sequencing of ethyl methanesulfonate-induced mutant bulks. *Plant Physiol*, 2012, 160(2): 591–600. [DOI]
- [50] Birkeland SR, Jin N, Ozdemir AC, Lyons RH, Weisman LS, Wilson TE. Discovery of mutations in *Saccharomyces cerevisiae* by pooled linkage analysis and whole-genome sequencing. *Genetics*, 2010, 186(4): 1127–1137. [DOI]
- [51] Nowrousian M, Teichert I, Masloff S, Kück U. Whole-genome sequencing of *Sordaria macrospora* mutants identifies developmental genes. *G3 (Bethesda)*, 2012, 2(2): 261–270. [DOI]
- [52] Zhu Y, Mang HG, Sun Q, Qian J, Hipps A, Hua J. Gene discovery using mutagen-induced polymorphisms and deep sequencing: application to plant disease resistance. *Genetics*, 2012, 192(1): 139–146. [DOI]
- [53] Nordström KJV, Albani MC, James GV, Gutjahr C, Hartwig B, Turck F, Paszkowski U, Coupland G, Schneeberger K. Mutation identification by direct comparison of whole-genome sequencing data from mutant and wild-type individuals using *k*-mers. *Nat Biotechnol*, 2013, 31(4): 325–330. [DOI]
- [54] Chen LY, Shi DQ, Zhang WJ, Tang ZS, Liu J, Yang WC. The *Arabidopsis* alkaline ceramidase TOD1 is a key turgor pressure regulator in plant cells. *Nat Commun*, 2015, 6:

- Article number 6030. [DOI]
- [55] Sakai H, Kanamori H, Arai-Kichise Y, Shibata-Hatta M, Ebana K, Oono Y, Kurita K, Fujisawa H, Katagiri S, Mukai Y, Hamada M, Itoh T, Matsumoto T, Katayose Y, Wakasa K, Yano M, Wu JZ. Construction of pseudomolecule sequences of the *aus* rice cultivar kasalath for comparative genomics of Asian cultivated rice. *DNA Res*, 2014, 21(4): 397–405. [DOI]
- [56] Lindner H, Raissig MT, Sailer C, Shimosato-Asano H, Bruggmann R, Grossniklaus U. SNP-Ratio Mapping (SRM): identifying lethal alleles and mutations in complex genetic backgrounds by next-generation sequencing. *Genetics*, 2012, 191(4): 1381–1386. [DOI]
- [57] Austin RS, Vidaurre D, Stamatiou G, Breit R, Provart NJ, Bonetta D, Zhang JF, Fung P, Gong YC, Wang PW, McCourt P, Guttman DS. Next-generation mapping of *Arabidopsis* genes. *Plant J*, 2011, 67(4): 715–725. [DOI]
- [58] Fukushima A, Kusano M, Mejia R F, Iwasa M, Kobayashi M, Hayashi N, Watanabe-Takahashi A, Narisawa T, Tohge T, Hur M, Wurtele ES, Nikolau BJ, Saito K. Metabolomic characterization of knockout mutants in *Arabidopsis*: development of a metabolite profiling database for knockout mutants in *Arabidopsis*. *Plant Physiol*, 2014, 165(3): 948–961. [DOI]
- [59] Roycewicz PS, Malamy JE. Cell wall properties play an important role in the emergence of lateral root primordia from the parent root. *J Exp Bot*, 2014, 65(8): 2057–2069. [DOI]
- [60] Laitinen RAE, Schneeberger K, Jelly NS, Ossowski S, Weigel D. Identification of a spontaneous frame shift mutation in a nonreference *Arabidopsis* accession using whole genome sequencing. *Plant Physiol*, 2010, 153(2): 652–654. [DOI]
- [61] Allen RS, Nakasugi K, Doran RL, Millar AA, Waterhouse PM. Facile mutant identification via a single tal backcross method and application of whole genome sequencing based mapping pipelines. *Front Plant Sci*, 2013, 4: 362. [DOI]
- [62] Earley EJ, Jones CD. Next-generation mapping of complex traits with phenotype-based selection and introgression. *Genetics*, 2011, 189(4): 1203–1209. [DOI]
- [63] Andolfatto P, Davison D, Erezylmaz D, Hu TT, Mast J, Sunayama-Morita T, Stern DL. Multiplexed shotgun genotyping for rapid and efficient genetic mapping. *Genome Res*, 2011, 21(4): 610–617. [DOI]
- [64] Zou GH, Zhai GW, Feng Q, Yan S, Wang A, Zhao Q, Shao JF, Zhang ZP, Zou JQ, Han B, Tao YZ. Identification of QTLs for eight agronomically important traits using an ultra-high-density map based on SNPs generated from high-throughput sequencing in sorghum under contrasting photoperiods. *J Exp Bot*, 2012, 63(15): 5451–5462. [DOI]
- [65] Zhou G, Chen Y, Yao W, Zhang CJ, Xie WB, Hua JP, Xing YZ, Xiao JH, Zhang QF. Genetic composition of yield heterosis in an elite rice hybrid. *Proc Natl Acad Sci USA*, 2012, 109(39): 15847–15852. [DOI]
- [66] Zhebentyayeva TN, Fan SH, Chandra A, Bielenberg DG, Reighard GL, Okie WR, Abbott AG. Dissection of chilling requirement and bloom date QTLs in peach using a whole genome sequencing of sibling trees from an F₂ mapping population. *Tree Genet Genomes*, 2014, 10(1): 35–51. [DOI]
- [67] Jaramillo-Lambert A, Fuchsman AS, Fabritius AS, Smith HE, Golden A. Rapid and efficient identification of *Caenorhabditis elegans* Legacy Mutations Using Hawaiian SNP-based mapping and whole genome sequencing. *G3(Bethesda)*, 2015, pii: g3.115.017038. [DOI]
- [68] Jiao WQ, Fu XT, Dou JZ, Li HD, Su HL, Mao JX, Yu Q, Zhang LL, Hu XL, Huang XT, Wang YF, Wang S, Bao ZM. High-resolution linkage and quantitative trait locus mapping aided by genome survey sequencing: building up an integrative genomic framework for a bivalve mollusc. *DNA Res*, 2014, 21(1): 85–101. [DOI]
- [69] O'Rourke SM, Yochem J, Connolly AA, Price MH, Carter L, Lowry JB, Turnbull DW, Kamps-Hughes N, Stiffler N, Miller MR, Johnson EA, Bowerman B. Rapid mapping and identification of mutations in *Caenorhabditis elegans* by RAD mapping and genomic interval pull-down sequencing. *Genetics*, 2011, 189(3): 767–778. [DOI]
- [70] Liu S, Yeh CT, Tang HM, Nettleton D, Schnable PS. Gene mapping via bulked segregant RNA-seq (BSR-seq). *PLoS One*, 2012, 7(5): e36406. [DOI]
- [71] Hill JT, Demarest BL, Bisgrove BW, Gorski B, Su YC, Yost HJ. MMAPPR: mutation mapping analysis pipeline for pooled RNA-seq. *Genome Res*, 2013, 23(4): 687–697. [DOI]
- [72] Miller AC, Obholzer ND, Shah AN, Megason SG, Moens CB. RNA-seq-based mapping and candidate identification of mutations from forward genetic screens. *Genome Res*, 2013, 23(4): 679–686. [DOI]
- [73] Fairfield H, Gilbert GJ, Barter M, Corrigan RR, Curtain M, Ding Y, D'Ascenzo M, Gerhardt DJ, He C, Huang W, Richmond T, Rowe L, Probst FJ, Bergstrom DE, Murray SA, Bult C, Richardson J, Kile BT, Gut I, Hager J, Sigurdsson S, Mauceli E, Di Palma F, Lindblad-Toh K, Cunningham ML, Cox TC, Justice MJ, Spector MS, Lowe

- SW, Albert T, Donahue LR, Jeddloh J, Shendure J, Reinholdt LG. Mutation discovery in mice by whole exome sequencing. *Genome Biol*, 2011, 12(9): R86. [DOI]
- [74] Andrews TD, Whittle B, Field MA, Balakishnan B, Zhang Y, Shao Y, Cho V, Kirk M, Singh M, Xia Y, Hager J, Winslade S, Sjollem G, Beutler B, Enders A, Goodnow CC. Massively parallel sequencing of the mouse exome to accurately identify rare, induced mutations: an immediate source for thousands of new mouse models. *Open Biol*, 2012, 2(5): 120061. [DOI]
- [75] Haraksingh RR, Jahanbani F, Rodriguez-Paris J, Gelernter J, Nadeau KC, Oghalai JS, Schrijver I, Snyder MP. Exome sequencing and genome-wide copy number variant mapping reveal novel associations with sensorineural hereditary hearing loss. *BMC Genomics*, 2014, 15(1): 1155. [DOI]
- [76] del Viso F, Bhattacharya D, Kong Y, Gilchrist MJ, Khokha MK. Exon capture and bulk segregant analysis: rapid discovery of causative mutations using high-throughput sequencing. *BMC Genomics*, 2012, 13: 649. [DOI]
- [77] Mascher M, Richmond TA, Gerhardt DJ, Himmelbach A, Clissold L, Sampath D, Ayling S, Steuernagel B, Pfeifer M, D'Ascenzo M, Akhunov ED, Hedley PE, Gonzales AM, Morrell PL, Kilian B, Blattner FR, Scholz U, Mayer KF, Flavell AJ, Muehlbauer GJ, Waugh R, Jeddloh JA, Stein N. Barley whole exome capture: a tool for genomic research in the genus *Hordeum* and beyond. *Plant J*, 2013, 76(3): 494–505. [DOI]
- [78] Mascher M, Jost M, Kuon JE, Himmelbach A, Aßfalg A, Beier S, Scholz U, Graner A, Stein N. Mapping-by-sequencing accelerates forward genetics in barley. *Genome Biol*, 2014, 15(6): R78. [DOI]
- [79] Pankin A, Campoli C, Dong X, Kilian B, Sharma R, Himmelbach A, Saini R, Davis SJ, Stein N, Schneeberger K, von Korff M. 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*, 2014, 198(1): 383–396. [DOI]
- [80] Ryan S, Willer J, Marjoram L, Bagwell J, Mankiewicz J, Leshchiner I, Goessling W, Bagnat M, Katsanis N. Rapid identification of kidney cyst mutations by whole exome sequencing in zebrafish. *Development*, 2013, 140(21): 4445–4451. [DOI]

(责任编辑: 胡松年)