遗传 Hereditas (Beijing) 2014 年 1 月, 36(1): 41－49
www.chinagene.cn

RAD-seq 技术在基因组研究中的现状及展望

王洋坤, 胡艳, 张天真

南京农业大学, 作物遗传与种质创新国家重点实验室/教育部杂交棉创制工程研究中心, 南京 210095

摘要: Restriction-site associated DNA sequencing(RAD-seq)技术是在二代测序基础上发展起来的一项基于全基因组酶切位点的简化基因组测序技术。该方法技术流程简单, 不受有无参考基因组的限制, 可大大简化基因组的复杂性, 减少实验费用, 通过一次测序就可以获得数以万计的多态性标记。目前, RAD-seq 技术已成功应用于超高密度遗传图谱的构建、重要性状的精细定位、辅助基因组序列组装、群体基因组学以及系统发生学等基因组研究热点领域。文章主要介绍了 RAD-seq 的技术原理、技术发展及其在基因组研究中的广泛应用。鉴于 RAD-seq 方法的独特性, 该技术必将在复杂基因组研究领域具有广泛的应用前景。

关键词: RAD-seq; 基因组; 遗传图谱; SNP; 双酶系统的 RAD(ddRAD)测序

Current status and perspective of RAD-seq in genomic research

Yangkun Wang, Yan Hu, Tianzhen Zhang

State Key Laboratory of Crop Genetics and Germplasm Enhancement/Cotton Hybrid R & D Engineering Center of the Ministre of Education, Nanjing Agricultural University, Nanjing 210095, *China*

Abstract: The restriction-site associated DNA sequencing (RAD-seq) is a high-throughput sequencing technique developed from the next-generation sequencing (NGS). This method can reduce the representation of the complex genome while mapping thousands of polymorphic markers with or without a reference genome. It has been extensively used for high-density genetic map construction, fine mapping of important genes, genome sequence assembly, population genomic research, as well as phylogenetic research and so on. Here, we introduce the technological principle and development of RAD-seq combined with the sequencing applications in various species. Due to its uniqueness, RAD-seq will have a wide application in genetic analysis of complex genomic research in the future.

Keywords: RAD-seq; genome; genetic map; SNP; double enzyme system of RAD (double digest RAD ddRAD) sequencing

2005 年, 美国 454 生命科学公司Margulies等[1]在国际顶级学术期刊*Nature*上报道了一种快速简单的测序方法：结合 DNA 扩增的乳胶系统(Emulsion system)和以皮升为单位的焦磷酸(Pyrophosphate)为

收稿日期: 2013–07–16; **修回日期:** 2013–08–15
基金项目: 国家重点基础研究发展规划(973 计划)项目(编号：2011CB109300)资助
作者简介: 王洋坤, 硕士研究生, 专业方向：基因组生物学。E-mail: cherrywyk@163.com
通讯作者: 张天真, 博士, 教授, 研究方向：作物遗传育种。E-mail: cotton@njau.edu.cn
DOI: 10.3724/SP.J.1005.2014.0041
网络出版时间: 2013-10-16 19:05:20
URL: http://www.cnki.net/kcms/detail/11.1913.R.20131016.1905.003.html

基础的测序方法——焦磷酸测序(Pyrosequencing)方法，二代测序(Next-generation sequencing, NGS)的时代由此开启。目前市场上主流的二代测序技术有Roche/454 焦磷酸测序(2005 年)、Illumina/Solexa 聚合酶合成测序(2006 年)和 ABI/SOLiD 连接酶测序(2007 年)。与传统的一代测序相比，新一代测序技术共有的突出特征是：单次运行(run)产出的序列数据量大，所以二代测序又被称为高通量测序技术。新一代测序技术的产生有助于人们以更低廉的价格，快捷、全面、深入地分析基因组、转录组及蛋白质之间交互作用的各项数据。

简化基因组测序(Reduced-representation sequencing)是在第二代测序基础上发展起来的一种利用酶切技术、序列捕获芯片技术或其他实验手段降低物种基因组复杂程度，针对基因组特定区域进行测序，进而反映部分基因组序列结构信息的测序技术。目前发展起来的简化基因组测序有：复杂度降低的多态序列(Complexity reduction of polymorphic sequences, CRoPS)测序[2]，限制性酶切位点相关的DNA (Restriction-site associated DNA, RAD)测序[3]，基因分型测序(Genotyping by sequencing, GBS)[4]，其中运用最为广泛的是限制性酶切位点相关DNA的测序技术，即RAD-seq。该技术利用限制性内切酶对基因组进行酶切，产生一定大小的片段，构建测序文库，对酶切后产生的RAD标记进行高通量测序。由于RAD标记是全基因组范围的呈现特异性酶切位点附近的小片段DNA标签，代表了整个基因组的序列特征，因此通过对RAD标记测序能够在大多数生物中获得成千上万的单核苷酸多态性(Single nucleotide polymorphism, SNP)标记[5,6]。该技术的优点在于：(1)通量高，通过一次测序开发 RAD 标记的数量是传统分子标记开发技术的 10 倍；(2)准确性高，数字化信号和高覆盖度使其较传统的分子标记准确性大大提升；(3)数据利用率高，性价比高，由于基因组的复杂度被大幅降低，从而降低了测序成本，因而特别适合在群体水平进行研究；(4)实验周期短，由于具有高通量的特点，经过一次测序能够产生数以万计的标记，大大缩短了传统标记的开发周期；(5)不受基因组序列的限制，对没有参考基因组的物种也可以进行大规模筛查 SNP 位点。RAD-seq已成功应用于SNP标记的开发、超高密度遗传图谱的构建、动植物重要经济性状的QTL定位、群体遗传结构、系统演化分析和辅助全基因组*de novo*测序等研究领域[7~10]。

1 RAD-seq 的主要技术流程

RAD-seq 的主要技术流程包括：基因组 DNA 的酶切，测序文库的构建，上机测序，数据分析等 4 个步骤。

(1)利用限制性内切酶对基因组DNA样品进行酶切。一般情况下，八碱基酶在基因组中出现的频率最低，其次是六碱基酶，出现频率最高的为四碱基酶。限制性内切酶的选择需要对目标物种的参考基因组(或已知BAC序列)进行系统分析，根据基因组的GC含量、重复序列情况等信息选择合适的酶。2008 年，Baird等[11]首先使用八碱基酶*Sbf* Ⅰ (CCTGCAGG) 对三刺鱼 (*Gasterosteus aculeatus*) 基因组DNA进行酶切，测序得到 14 万个RAD标记；而后，为了对三刺鱼的侧鳍性状进行精细定位，又使用三刺鱼基因组序列中出现频率更高的六碱基酶 *Eco*R Ⅰ (GAATTC)对亲本以及F_2 群体基因组DNA进行酶切，具有不完整侧鳍与完整侧鳍的两个亲本分别获得150 万和 250 万个RAD标记。很显然，与八碱基酶*Sbf* Ⅰ相比，通过*Eco*R Ⅰ的酶切能够产生更高密度的RAD标记。在选择限制性内切酶时要根据物种基因组序列信息以及实验目的来选择，保证产生的RAD标记能够在基因组上均匀分布，同时所获得的RAD标记数量能够达到实验所需的饱和度。

(2)测序文库的构建。首先，在酶切后的基因组片段两端加上 P1 接头。如图 1A 所示，P1 接头包含 4 个部分：与 PCR 扩增的前引物结合的互补序列；与 Illumina 测序引物结合的互补序列；用以对样品进行跟踪的 4~5 bp 的 Barcode(每个 Barcode 之间最好存在超过 2 个碱基的差异)；相应的限制性酶切位点。然后，将加好 P1 接头的序列进行打断(图 1B)。通过琼脂糖胶检测，选择符合大小的目的条带，一般选择目标条带在 400~500 bp。打断后的 DNA 片段连接上 P2 接头(图 1C)。这样 DNA 片段有的加上 P1、P2 接头，有的两端都加上 P1 接头，有的两端都加上 P2 接头。对这样的混合 DNA 进行 PCR 扩增。由于

P2 接头的“Y 型”特殊结构，使两端只有 P2 而没有 P1 的接头无法扩增，没有 P1 接头的 DNA 片段被过滤掉。通过 PCR 扩增富集得到既有 P1 接头、又有 P2 接头的 DNA 序列(图 1D)。

(3)上机测序。目前 RAD-seq 常用的测序平台为 Illumina GAII 或 Illumina HiSeq2000 平台。测序深度需要根据实验目的来选择，对于遗传连锁分析，一般要求亲本的平均测序深度为 10×以上，F_1、F_2 等临时性群体，推荐每个个体平均测序深度为 0.8-1×；RIL、DH 等永久性群体，推荐每个个体平均测序深度为 0.6×。对于群体遗传学分析，推荐每个个体平均测序深度为 1.5 ×。

(4)数据分析。目前，Stacks软件(http://creskolab.uoregon.edu/stacks/)被广泛用于RAD-seq的数据分析中[12]。该软件可以用于基于RAD-seq数据的遗传图谱的构建、群体基因组学研究、系统发生学研究。整个数据分析流程包含以下 3 个部分：

① 原始数据处理(Raw Reads)：该阶段为整个数据处理通路的起始准备阶段，要求输入的数据格式为 FASTA 或者 FASTAQ 格式，主要是利用 process_radtags 程序检测 barcode 与酶切位点是否完整并且按照不同的 barcode 将每个样本的 reads 分开，通过检测将 barcode 不完整，酶切位点处有一到两个碱基错配的序列进行修正。同时，对每条序列的质量进行评估，过滤掉那些被修正的可能性低于 90%的序列。通过原始数据处理，每条 clean reads 被分配到每一个样品下，保证了测序数据的质量可以用于以下核心程序的分析。

② 核心(Core)阶段：核心元件为 Ustacks 程序(图 2)。Ustacks 首先将从 process_radtags 程序获得的单个样本的 reads 进行聚类，得到 stacks(图 2A)。由于默认的测序深度必须大于 7 倍，因此每个 stack 最少不能低于 7 条 reads。能够聚类成为一个 stack 的 reads 为初级 reads，其他不能形成聚类的为次级 reads。而后，由 A 步骤获得的 stacks 被打乱重混，按照 k-mer 值重新聚类，将每个含有一个核苷酸差异的 stacks 以节点的形式连接起来(图 2B)。每一个圆形节点为一个 stack，两点间的距离为一个核苷酸差异，由节点和线段连接成一个基因座位(Loci)(图 2C)。需要注意的是，节点与节点之间的连接必须是单向的，如图 2C 中灰色圆点基因座位不符合此项规则，故舍去。接着，重新过滤一遍次级 reads，将与已

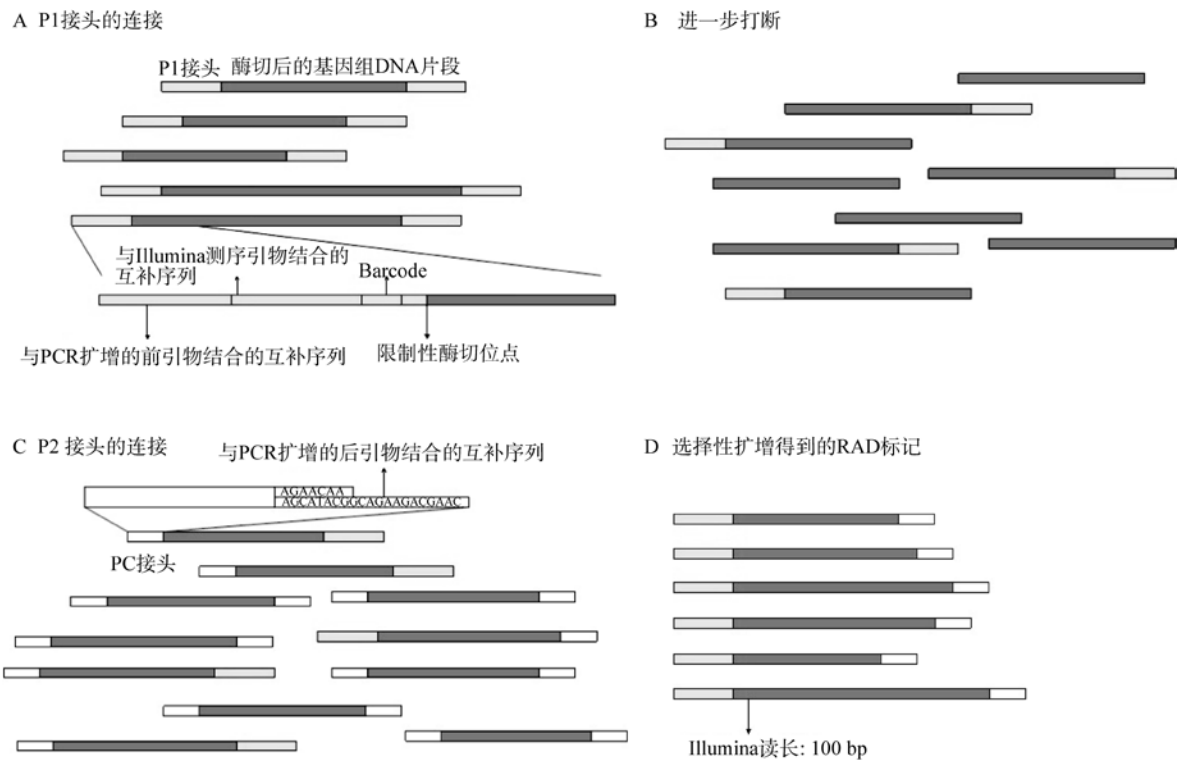

图 1　RAD-seq测序文库的构建流程[11]

存在 stacks 相差两个核苷酸以内的次级 reads 重新利用(图 2D), 这在一定程度上提高了数据的利用率, 增加了 stacks 的深度。图 2E 为 D 图中基因座位 1 的序列显示, 新加入的次级 reads 与初级 reads 除了有 C/A 差异位点之外, 还有其他两个核苷酸以内的差异, 但这种差异是可以忽略的, 并不影响将 C/A 差异位点视为一对等位基因(图 2F)。之后, 通过 Cstacks 程序将两个亲本中所出现的 stacks 综合编入, 形成一个含有双亲中所有基因座位的目录(图 2G)。最后, 由 Sstacks 程序将每个子代个体中出现的基因座位与双亲中出现的基因座位进行一对一搜索和概率计算, 定义出每一个基因座位上的等位基因(图 2H)。每一个步骤的结果都可以传入 MYSQL 数据库中。

③ 应用(Utilities)阶段: 该阶段为整个数据处理通路中最为灵活的阶段, 可以根据不同的实验目的选择不同的程序。Genotypes 程序具有自动纠正功能, 可以对每个位点的基因型进行检测, 例如对子代中纯合标记的检测, 确保其中没有出现 SNP, 保证了每个标记位点的准确性。通过该程序处理后的数据, 利用 Joinmap 或者 R/QTL 软件, 可直接进行遗传图谱的构建。Populations 程序在某种情况下可以代替 genotypes 软件的使用, 但其主要用于群体遗传分析, 该程序能够网页输出 VCF 格式的 SNP, 计算例如 *Pi*、F_{is} 和 F_{st} 等群体遗传学相关的统计数据。

经过以上 3 个阶段的分析, 基本上能够完成 RAD-seq 数据的分析。另外, Stacks 软件还有一些其他的应用程序。例如, 在原始数据处理阶段 process_shortreads 程序也可快速过滤掉一些低质量序列并

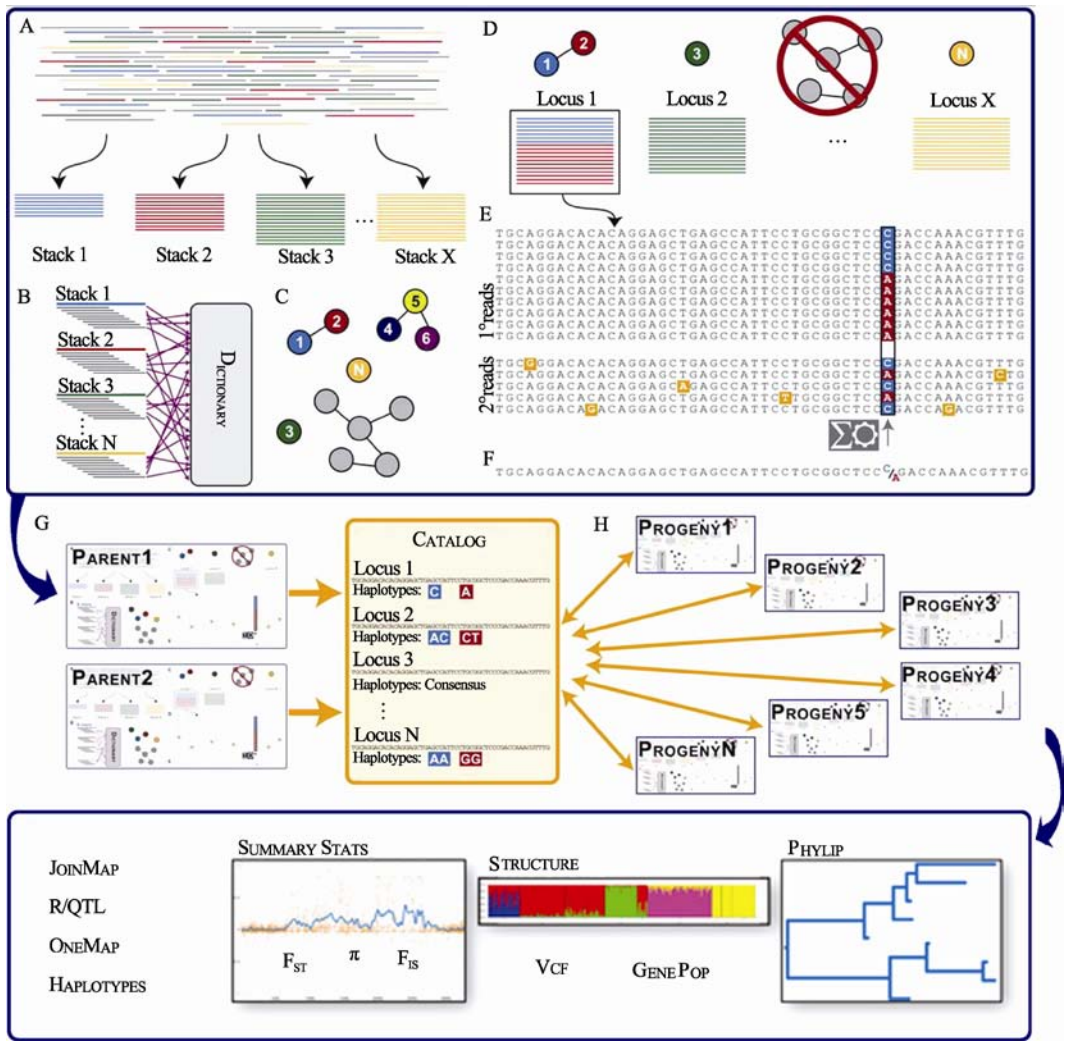

图 2 Stacks 软件的分析流程[12,13]

且将每个样本的 reads 分开，不同之处在于 process_shortreads 程序是修剪掉那些低质量的序列而非将其直接删除，因此并不适用于 RAD-seq 的数据处理。在核心阶段，当有参考基因组信息时，可使用 Pstacks 程序代替 Ustacks 程序，后续分析的程序 Cstacks、Sstacks 依然适用。在应用阶段也有更多的程序可以使用，在这里就不一一列举。

当然，Stacks软件并不是RAD-seq数据分析的唯一软件，聚类软件CLUSTER与比对软件MUSCLE[14]、BLAST[15,16]、SAMtools[17]等相结合也用于RAD-seq数据的分析。

2　RAD-seq 技术的发展

在已发表的一些针对没有参考基因组物种的RAD-seq文章中，有将近一半的原始数据因为测序错误被丢弃。同时，每个区域约有 30%~50%的基因座位由于含有 3 个以上的碱基多样性而被丢弃[8,9,11]。因此，为了提高数据的使用效率，增加可供分析的reads数量，提高每个基因座位的准确性，需要在方法上对传统的RAD-seq方法进行改善。目前，在单酶切 RAD-seq 技术上发展起来的有双酶切的 RAD (Double digest RAD, ddRAD)测序技术和IIB 型限制性内切酶的RAD(IIB digest RAD, 2b-RAD)技术。

双酶切的 RAD-seq 技术与单酶切 RAD-seq 技术的区别在于，基因组 DNA 通过一个稀有酶与一个常见酶相结合进行双酶切，这样处理免去打断的过程直接进行目的片段的筛选。在第二端的酶切位点后通过 PCR 扩增引入 Index，从而使更多的样品能够混在一起进行测序。该方法经过 Illumina Hiseq2000 的双端测序之后，能够获得相对于单酶切 RAD-seq 几倍的有效数据。

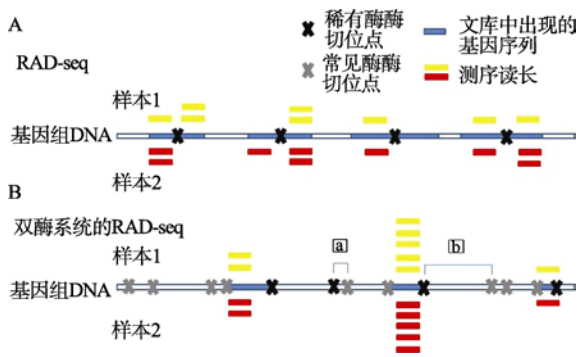

图 3　RAD-seq与double digest RAD-seq的比较[18]

ddRAD-seq 能够在改善测序效率的同时大大的减少实验成本。单酶切 RAD-seq(图 3A)，利用单一的限制性内切酶和随机打断对基因组进行切割，由于缺少方向性，酶切位点两边相邻的 100 bp 序列如蓝色区域所示都可能被测出，通过测序呈现出的序列分散度高，准确性就相对较低。如图 3B 所示，由双酶切系统对基因组进行切割并且辅以对酶切后产物片段大小的选择(一般为 500 bp 左右)，这样就把序列固定在了两端为不同酶切位点并且长度为 500 bp 左右的片段中，如图中的蓝色区域所示。a、b 两处虽然也在不同的酶切位点之间，但因大小并不符合规定的产物长度故不列入考虑范围。由此可见，双酶系统对 DNA 文库的筛选更为严格，通过测序得到的序列也就更为准确。在通量相同的情况下，利用双酶切系统的 RAD-seq 就能检测更多的样本，提高数据的利用率，减少成本。

2b-RAD技术采用的是利用IIB型限制性内切酶对基因组DNA进行酶切，这类酶(比如*Bsa*ⅩⅠ和*Alf* Ⅰ)能在基因组DNA上靶标位点上游和下游位点切断DNA，获得长度一致的DNA片段。该技术无需预知基因组信息，文库构建简单快捷，标签密度易于调节，成本低廉。Wang等[19]在拟南芥中对该方法进行了验证，结果表明 2b-RAD的准确性高，所需标记密度调整精细，这种方法特别适合于连锁图谱与自然群体中遗传变异图谱的构建。

3　RAD-seq 的应用

3.1　RAD-seq 在分子标记开发和基因分型上的应用

SNP 是基因组中最常见的变异类型，具有分布广、数量多的优点。传统的 SNP 标记开发方法通量低、开发成本高，极大地限制了 SNP 标记在高密度遗传图谱中的应用。RAD-seq 技术具有不依赖于基因组序列的优点，可进行高通量的 SNP 标记的开发。

2011 年，Barchi 等[20]将 RAD-seq 应用于茄子(*Solanum melongena*)的SNP标记开发。两个具有优良性状的育种亲本利用Illumina GAII平台进行 PE54 测序，共计获得约 45 000 条非冗余序列，70%为两个亲本共有序列，鉴定出约 10 000 个SNPs和约 1 000 个InDels，SNP和InDels频率分别为 0.8/kb和 0.07/kb。

通过RAD 序列预测到 2 000 个SSRs。研究表明，RAD-seq能够发掘大量的DNA分子标记，用于标记辅助选择和比较基因组学分析。

2012 年，Scaglione等[21]对 3 个洋蓟(*Cynara cardunculus*)群体及亲本进行RAD-seq测序，获得970 万条reads，大约 1 Gb数据。进行contigs组装后，利用不同样本的共有序列共开发出 34 000 个SNPs和大约 800 个InDels标记。杂合的SNP位点通过CAPS assays得到了较好的验证。此研究表明，RAD- seq技术也可用于高杂合物种的SNP标记的开发。

2012 年，Bus等[22]采用RAD-seq的方法，对 8 个油菜(*Brassica napus*)近交系种质材料进行了多态性检测和基因分型，共检测和鉴定到了 20 000 多个SNPs和 125 个InDels，约有 1/3 的RAD标记被聚类并比对到油菜参考序列。该研究表明，RAD-seq不仅仅是一个简单而经济有效的检测高密度多态性的方法，同时对于多倍体物种如油菜等的研究，也是一种进行SNP基因分型的有效方法。

3.2 RAD-seq 在图谱构建上的应用

将回交群体、F_2 群体和亲本同时进行测序，所得到的 RAD-seq 数据可以用于超高密度的多态性图谱的构建，进而用于关联性图谱和遗传图谱的构建。

2012 年，Poland等[10]利用限制性内切酶*Pst* Ⅰ(CTGCAG)与*Msp* Ⅰ(CCGG)的双酶系统对大麦和小麦的基因组分别进行RAD-seq，得到一张有 34 000 个SNPs和 240 000 个标记的俄勒冈州乌尔夫大麦(*Hordeum vulgare*)的高密度遗传图谱，以及一张有 20 000 个SNPs和 367 000 个标记的杂交小麦超高密度遗传图谱，证实了ddRAD-seq在大而复杂的多倍体基因组上的可用性。

2012年，Peterson等[18]对一新兴的啮齿目模式动物鹿鼠(*Genus Peromyscus*)利用 *Eco*R Ⅰ(GAATTC)和*Msp* Ⅰ(CCGG)双酶切系统的RAD-seq在两个姐妹物种*Maniculatus*和*Polionotus*的杂交群体中分离出了 1 000 多个有固定差异的SNPs位点，构建了一张含有1 158 个SNPs标记的遗传连锁图。之后，为了验证该方法针对野生物种也具有同样的适用性，Peterson又在自然种群*Leucopus*中捕捉到了 146 个野生种，分两次使用与前试验同样双酶系统的RAD- seq，第一次对 54 个个体进行测序，找到了 6 199 个多态性区域，15 962 个SNPs，第二次对 92 个个体进行测序，共找到 18 907 个SNPs。两次测序得到的SNPs 有大部分相同，因此ddRAD标记的可用性得到了进一步的验证。

更为重要的是，RAD-seq能够在不开发全新的标记情况下，添加新的来源于远缘物种或不同物种的构图个体。因为RAD-seq能够产生大量的遗传标记，有足够的标记可以对有一对单杂合的亲本杂交产生的F_1 家系利用测交法构建遗传图谱。2011 年，Amores等[23]就利用该方法绘制出了一张含有 8 406 个RAD标记的斑点雀鳝(*Lepisosteus oculatus*)高密度遗传图谱。

测交法的流程如图 4 所示，由一对杂合亲本产生的 F_1 群体可以用来生成以 RAD 为标记的遗传图谱。在一个亲本中是杂合，而在另一个亲本中是纯合的同一个标记(图 4A、B 位点)可以用于测交检测，这样的一对标记在每一个 F_1 群体中则会以一个纯合而另一个杂合的形式出现，通过结合两亲本图谱中均含有的杂合等位基因 C，即可合并成一个完整的图谱。

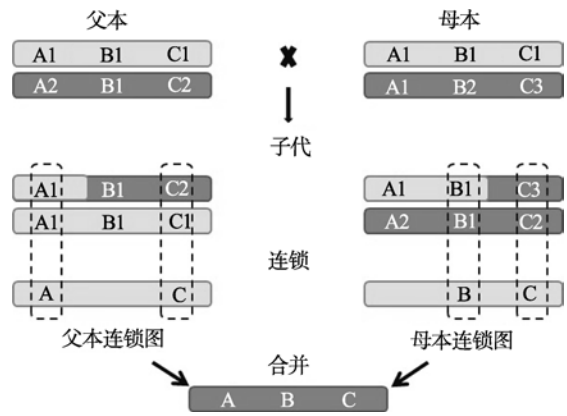

图 4 RAD 测交分析法[23]

3.3 RAD-seq 在动植物重要经济性状基因/QTL 定位上的应用

RAD-seq是一个功能强大的SNP开发平台，使用RAD-seq可以发掘大量的SNP标记，进行重要基因的定位。2008 年，Baird等[11]首次将RAD标记应用于第二代测序中，以三刺鱼为研究对象，论证了RAD-seq可以独立地识别控制侧鳍性状的基因，以

及在连锁群Ⅳ中一些与缺失侧鳍相关的其他位点。研究样本采用了来自两个在侧鳍性状存在明显差异的亲本及其F_2群体中的 96 个个体。利用限制性内切酶*Sbf* Ⅰ (CCTGCAGG)酶切子代和亲本的基因组DNA, 通过测序共识别了 41 622 个均匀分布于基因组中的RAD标记, 得到了 13 000 个SNPs。最终, 将控制三刺鱼侧鳍缺失的*Eda*位点定位在了连锁群Ⅳ上, 距离最近的RAD标记为 1.5 Mb。

2011 年, Pfender等[9]通过对黑麦草(*Lolium perenne*)锈病的抗感和易感亲本及由其杂交产生 188 个F_1 家系进行RAD-seq, 两个亲本各得到约 100 万条reads, 聚类分析后得到 1.7 万的初级RAD标记, 配合F_1 家系的测序数据按规定的准则筛选过滤之后母本获得 1 156 个RAD标记, 父本获得 1 216 个RAD标记, 构建了亲本的高密度遗传图谱。再通过F_1 代 193 个家系接种病原菌后的侵染表型, 结合SSR与STS标记, 定位到了 3 个锈病相关的QTL位点, 分别为位于黑麦草的 7 号连锁群上贡献值为 30~ 38 的主效 QTL(qLpPg1), 以及两个分别位于 1 号连锁群(qLpPg2)和 6 号连锁群(qLpPg3)的贡献值为 10 的QTL。

2012 年, Houston等[24]选取两种对于传染性胰脏坏死病病毒(Infectious Pancreatic Necrosis, IPN)感染抗性和敏感的三文鱼(*Salmo salar*)亲本及 14 个子代(7 个纯合抗性子代和 7 个纯合敏感子代个体), 利用RAD-seq技术进行基因分型, 构建遗传连锁图谱, 并结合相关表型数据进行了IPN抗性相关的QTL定位分析。鉴定到 6 712 个分离的SNPs, 其中 50 个SNPs与QTL连锁, 获得的这些QTL连锁SNPs可用于IPN感染后对三文鱼鱼苗的高通量分析和基因分型检测。

3.4　RAD-seq 在群体遗传及系统发生学上的应用

RAD-seq另一个非常强大的应用为利用RAD标记基因分型的结果, 进行高精度群体遗传、生态遗传和亲缘地理学以及系统发生学的研究。由于方法上的限制, 传统的群体遗传和亲缘地理研究只能利用得到的少量的基因座位进行分析, 无法满足对许多群体遗传相关参数的精确评估。RAD-seq能够得到数量巨大的多态性标记, 从而解决了传统方法基因座位少, 基因组信息代表性差的问题[25]。

2010 年, Hohenlohe等[26]用该方法研究了三刺鱼自然群体的多样性分化, 选择了两个来自阿拉斯加南部的海洋群体和 3 个淡水群体的 100 个个体, 利用限制性内切酶 *Sbf* Ⅰ (CCTGCAGG) 酶切基因组DNA, 通过测序得到 45 789 个SNPs。总体估计了三刺鱼群体的遗传多样性, 进而证明了大型随机交配的海洋群体会多频产生表型变异的淡水群体的生物地理假说。

2010 年, Emerson等[7]仅利用适量的RAD数据, 就揭示了来自 21 个不同群体的北美瓶草蚊(*Wyeomyia smithii*)的进化关系。通过对这 21 个样品的基因组DNA进行RAD-seq, 共获得 2 750 万条reads, 1 490 万条通过几项标准的过滤, 平均每个个体为 711 702± 85 779 条。利用Stacks软件分别分析每个北美瓶草蚊个体, 平均每个个体得到 20 868±1 681 个stacks, 覆盖了 13 627±1 177 个基因座。最后, 共获得 3 714 个SNP标记用于揭示这 21 个不同北美瓶草蚊群体的亲缘关系, 即*Appalachian*种群与多数的南部种群有较近的亲缘关系, 而来自北美五大湖和加拿大中部的大陆种群则与更偏东的横跨圣劳伦斯河走廊的种群产生了分离。

雷默瑞丽蜗牛(*Cepaea nemoralis*)是一个优秀的传统生态遗传学模型, 但是由于缺乏遗传标记, 对其进化研究和多样性保护被停滞。2013 年, Richards等[27]利用RAD-seq技术对雷默瑞丽蜗牛两个亲本和 22 个子代个体进行测序。在控制色彩和色带的基因座位上找到了 44 个标记, 另外又开发了 11 个能够在 22 个子代中独立遗传的最优标记, 并在其他 146 个子代中得到了进一步验证。最近的两个RAD标记被定位在了 0.6 cM 之内, 最终构建了一张 35.8 cM 的连锁图, 重新建立了雷默瑞丽蜗牛在生态遗传学上优秀的分子模型地位。

3.5　RAD-seq 在辅助全基因组测序上的应用

RAD-seq还能够通过辅助全基因组测序对非模式物种进行遗传基础的研究。2013 年, Jia等[28]在*Nature*上发表了一篇关于构建二倍体小麦粗山羊草(*Aegilops tauschii*)框架图的文章。在进行scaffolds锚定的过程中, 利用RAD-seq和全基因组测序相结合

的方法，对粗山羊草Y2280、AL8/78 以及由其杂交得到的 490 株F_2家系进行测序，从而得到了一张具有 151 083 个SNPs标记的高密度遗传图谱，该图谱总长 1 059.806 cM，包含 13 688 个scafflolds，序列总长 1.277 Gb，是迄今为止密度最高的一张粗山羊草遗传图谱，该图谱在辅助scaffolds的锚定上起到了至关重要的作用。

2013 年，Xu等[29]利用RAD-seq与全基因组测序相结合，完成了对孟加拉虎(*Panthera tigris*)出现白色条纹的遗传机理的研究。通过对 3 个亲本进行全基因组测序同时辅以对具有“白色”基因位点的同一个血统的 16 个圈养虎进行RAD-seq，发现了致病突变是一种氨基酸变化(A477V)SLC45A2 转运蛋白，经过蛋白质构象的三维结构同源性检测表明，这样的替代可能部分阻止运输通道，从而影响黑素原的生成，该结论在 130 只无关的老虎中得到了验证。

4 展 望

RAD-seq 技术操作简便，周期短，实验成本低，同时不受参考基因组的限制，一次实验即获得的大量 SNP 信息，可以用于任何物种的高密度图谱的构建、基因(QTLs)定位及群体遗传分析。由此可见，随着 RAD-seq 技术趋于成熟，将被广泛应用于不同的生物学研究领域。

在分子育种领域，可以利用 RAD-seq 与转录组分析相结合的方法寻找目的基因。利用该方法，将显著提高复杂性状相关基因定位的效率，为分子育种领域的研究开辟广阔的发展前景。

在动植物多样性保护方面，可以利用 RAD-seq 与全基因组测序相结合的方法对珍惜物种的遗传多样性进行研究，通过高密度遗传图谱和物理图谱的构建获得更多的遗传信息，再与其亲缘关系相近的物种进行比对，最终找出该物种在进化上的特征，为物种多样性保护奠定了基础。

在基础医学研究方面，由于人类疾病的复杂性，单一位点的序列多样性并不足以导致表现型的变化，大部分控制疾病的基因多为数量性状位点。通过将RAD-seq 与 QTL 定位相结合便能够找到控制疾病的基因，后期再设计药物抑制该基因的表达，便能够达到治疗疾病的效果，对于肿瘤、心脏病、动脉硬化等疾病的研究有着十分重要的意义。

总之，随着测序和实验成本的进一步降低，RAD-seq 技术必将在复杂基因组遗传分析研究领域具有广泛的应用前景。

参考文献

[1] Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 2005, 437(7057): 376–380.[DOI]

[2] Altshuler D, Pollara VJ, Cowles CR, van Etten WJ, Baldwin J, Linton L, Lander ES. An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature*, 2000, 407(6803): 513–516. [DOI]

[3] Davey JL, Blaxter MW. RADSeq: next-generation population genetics. *Brief Funct Genomic*, 2010, 9(5–6): 416–423. [DOI]

[4] Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS ONE*, 2011, 6(5): e19379. [DOI]

[5] Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Res*, 2007, 17(2): 240–248. [DOI]

[6] van Tassell CP, Smith TP, Matukumalli LK, Taylor JF, Schnabel RD, Lawley CT, Haudenschild CD, Moore SS, Warren WC, Sonstegard TS. SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nat Methods*, 2008, 5(3): 247–252. [DOI]

[7] Emerson KJ, Merz CR, Catchen JM, Hohenlohe PA, Cresko WA, Bradshaw WE, Holzapfel CM. Resolving postglacial phylogeography using high-throughput sequencing. *Proc Natl Acad Sci USA*, 2010, 107(37): 16196–16200. [DOI]

[8] Hohenlohe PA, Amish JS, Catchen MJ, Allendorf WF, Luikart G. Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Mol Ecol Resour*, 2011, 11(Suppl. 1): 117–122. [DOI]

[9] Pfender WF, Saha MC, Johnson EA, Slabaugh MB. Mapping with RAD (restriction-site associated DNA) markers to rapidly identify QTL for stem rust resistance in *Lolium perenne*. *Theor Appl Genet*, 2011, 122(8): 1467–1480. [DOI]

[10] Poland JA, Brown PJ, Sorrells ME, Jannink JL. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PloS ONE*, 2012, 7(2): e32253. [DOI]

[11] Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Johnson EA. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, 2008, 3(10): e3376. [DOI]

[12] Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. Stacks: building and genotyping Loci *de novo* from short-read sequences. *G3*, 2011, 1(3): 171–182. [DOI]

[13] Catchen JM, Hohenlohe P, Bassham S, Amores A, Cresko WA. Stacks: an analysis tool set for population genomics. *Mol Ecol*, 2013, 22(11): 3124–3140. [DOI]

[14] Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*, 2004, 32(5): 1792–1797. [DOI]

[15] Altschul SF, Madden TL, Schäffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, 1997, 25(17): 3389–3402. [DOI]

[16] Kent WJ. BLAT-the BLAST-like alignment tool. *Genome Res*, 2002, 12(4): 656–664. [DOI]

[17] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Durbin R. The sequence alignment/map format and SAMtools. *Bioinformatics*, 2009, 25(16): 2078–2079. [DOI]

[18] Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE*, 2012, 7(5): e37135. [DOI]

[19] Wang S, Meyer E, McKay JK, Matz MV. 2b-RAD: a simple and flexible method for genome-wide genotyping. *Nat Methods*, 2012, 9(8): 808–810. [DOI]

[20] Barchi L, Lanteri S, Portis E, Acquadro A, Valè G, Toppino L, Rotino GL. Identification of SNP and SSR markers in eggplant using RAD tag sequencing. *BMC Genomics*, 2011, 12(1): 304. [DOI]

[21] Scaglione D, Acquadro A, Portis E, Tirone M, Knapp SJ, Lanteri S. RAD tag sequencing as a source of SNP markers in *Cynara cardunculus* L. *BMC Genomics*, 2012, 13(1): 3. [DOI]

[22] Bus A, Hecht J, Huettel B, Reinhardt R, Stich B. High-throughput polymorphism detection and genotyping in *Brassica napus* using next-generation RAD sequencing. *BMC Genomics*, 2012, 13(1): 281. [DOI]

[23] Amores A, Catchen J, Ferrara A, Fontenot Q, Postlethwait JH. Genome evolution and meiotic maps by massively parallel DNA sequencing: spotted gar, an outgroup for the teleost genome duplication. *Genetics*, 2011, 188(4): 799–808. [DOI]

[24] Houston RD, Davey JW, Bishop SC, Lowe NR, Mota-Velasco JC, Hamilton A, Guy DR, Tinch AE, Thomson ML, Blaxter ML, Gharbi K, Bron JE, Taggart JB. Characterisation of QTL-linked and genome-wide restriction site-associated DNA (RAD) markers in farmed Atlantic salmon. *BMC Genomics*, 2012, 13(1): 244. [DOI]

[25] Hohenlohe PA, Catchen J, Cresko WA. Population genomic analysis of model and nonmodel organisms using sequenced RAD tags. In: Data Production and Analysis in Population Genomics. New York: Humana Press, 2012: 235–260. [DOI]

[26] Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genet*, 2010, 6(2): e1000862. [DOI]

[27] Richards PM, Liu MM, Lowe N, Davey JW, Blaxter ML, Davison A. RAD-Seq derived markers flank the shell colour and banding loci of the *Cepaea nemoralis* supergene. *Mol Ecol*, 2013, 22(11): 3077–3089. [DOI]

[28] Jia J, Zhao SC, Kong XY, Li YR, Zhao GY, He WM, Appels RD, Pfeifer M, Tao Y, Zhang XY, Jing RL, Zhang C, Ma YZ, Gao LF, Gao C, Spannagl M, Mayer KFX, Li D, Pan SK, Zheng FY, Hu Q, Xia XC, Li JW, Liang QS, Chen J, Wicker T, Gou CY, Kuang HH, He GY, Luo YD, Keller B, Xia QJ, Lu P, Wang JY, Zou HF, Zhang RZ, Xu JY, Gao JL, Middleton C, Quan ZW, Liu GM, Wang J, Yang HM, Liu X, He ZH, Mao L, Wang J. *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature*, 2013, 496: 91–95. [DOI]

[29] Xu X, Dong GX, Hu XS, Miao L, Zhang XL, Zhang DL, Yang HD, Zhang TY, Zou ZT, Zhang TT, Zhuang Y, Bhak J, Cho YS, Dai WT, Jiang TJ, Xie C, Li RQ, Luo SJ. The genetic basis of white tigers. *Curr Biol*, 2013, 23(11): 1389–1504. [DOI]