

## 农业基因组学项目最合适的平台选择

### 不牺牲数据质量和周转时间

新一代测序的快速发展帮助农业科学家建立起基因组的广泛资源，这将打造一个“基因组文库的生动世界”。科研学者、动物育种专家和商业种子子公司都开始涉足这个庞大的基因组文库资源，从而加强农业基因组学策略。通过应用基因组标记来鉴定和选择重要性状，他们的目标是提高生产力和商业可行性。

本技术指南整合了同行评议的杂志中介绍的基于序列的基因分型方法的经验，并比较了芯片在农业基因分型应用中的表现，以便协助您作出基因分型技术的决策。

#### 基于序列的基因分型概述

基因组选择和关联作图或连锁不平衡(LD)定位技术需要大量的标记，才能准确估计与基因型相关联的性状。这就要求获得基因型信息所使用的技术必须是经济高效，且高通量的。全基因组测序以及利用序列捕获的靶向基因分型比较昂贵，而产生基因型数据的低通量方法对于常规应用而言仍然是不实际的。在确定经济型测序的目标下，基于测序的基因分型方法，如基于酶切的简化DNA测序<sup>1</sup> (RADseq)和genotyping-by-sequencing (GBS)<sup>2</sup>已不断发展，它们在科研和日常应用中的潜力被不断引用。

基于测序的方法依赖于对多个样品添加条形码，并降低基因分型的成本。这种技术利用限制性内切酶来消化目标限制位点和低拷贝基因组区域，以降低基因组复杂度。这样就能够避免带有重复序列的区域，它们容易产生模糊或假的SNP，且增加测序成本。利用测序而获得的基因型数据在质量和数量上大有不同，这高度依赖于生物体的基因组大小和结构以及评估的群体。基因组结构的复杂度，如倍性水平、GC含量和重复序列、待研究群体的遗传多样性，以及群体内的交配系统，都对测序技术准确轻松收集基因型数据的成本、准确性和效率有着直接的影响。对于那些标记探索落后或不完善的物种而言，基于测序的基因分型技术很有用。GBS和RADseq都能用于至少96个样品，而不需要访问参考数据库或之前发现的标记。这种技术还特别适合筛查数千个多态性，以了解遗传变异的后果，之前人们依赖极其少量的标记，如微卫星和扩增片段长度多态性(AFLP)。基于测序的基因分型技术已被用于标记的发现。杂志上已经发表了在多个物种上开展的各种实验的结果，如大麦、玉米、小麦、牛和鳟鱼等。在常规的基因分型中使用基于测序的基因分型技术依然遥遥无期，这有几方面的原因，本文中也列出了其中一些。《Molecular Ecology》杂志关于genotyping-by-sequencing技术的特刊<sup>3</sup>也总结指出，新的genotyping-by-sequencing技术仍然是不完善的，无法在不同的植物和动物中充分地扩展。

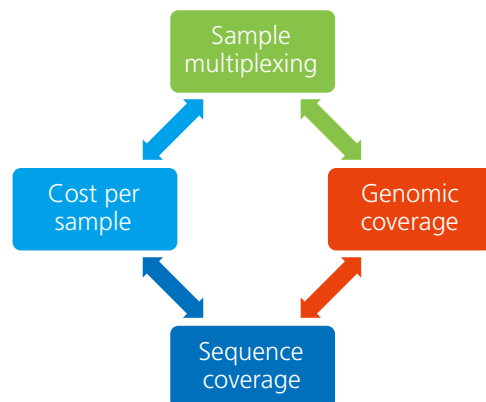
#### 基于序列的基因分型中的关键实验因素

所有的新一代测序平台都有序列碱基数量的限制，它们由每个测序运行产出。这个有限的产出能力意味着基于测序的基因分型运行必须平衡四个关键参数：样品多重分析的水平、基因组覆盖度、序列覆盖度以及每个样品的成本。

样品的多重分析很关键，因为测序仪的有限产出能力必须由运行中包括的所有样品来共享。更多样品意味着每个样品的测序碱基更少。

基因组覆盖度也很重要，因为它决定了被分析基因组的百分比，因此，也决定了基因组中可获取的标记数量。更高的基因组覆盖度是以牺牲其他参数中的一个为代价而实现的，因为它需要更多的测序仪产出能力。

图1: 新一代测序平台上的所有应用必须平衡四个关键参数。



序列覆盖度(或序列“深度”)决定了数据集中每个序列的平均读取数量。实际上,一些序列频繁被读取,而一些较少被读取,或根本没有。序列覆盖度影响数据中缺口的百分比以及基因型准确性。准确的基因型检出通常需要每个SNP上30倍或更高的覆盖度。增加序列覆盖度也迫使在其他地方妥协,以平衡测序仪能力的使用。

当然,样品多重分析、基因组覆盖度和序列覆盖度都能通过在测序仪上投入更多运行来改善,但这会使成本迅速增加。

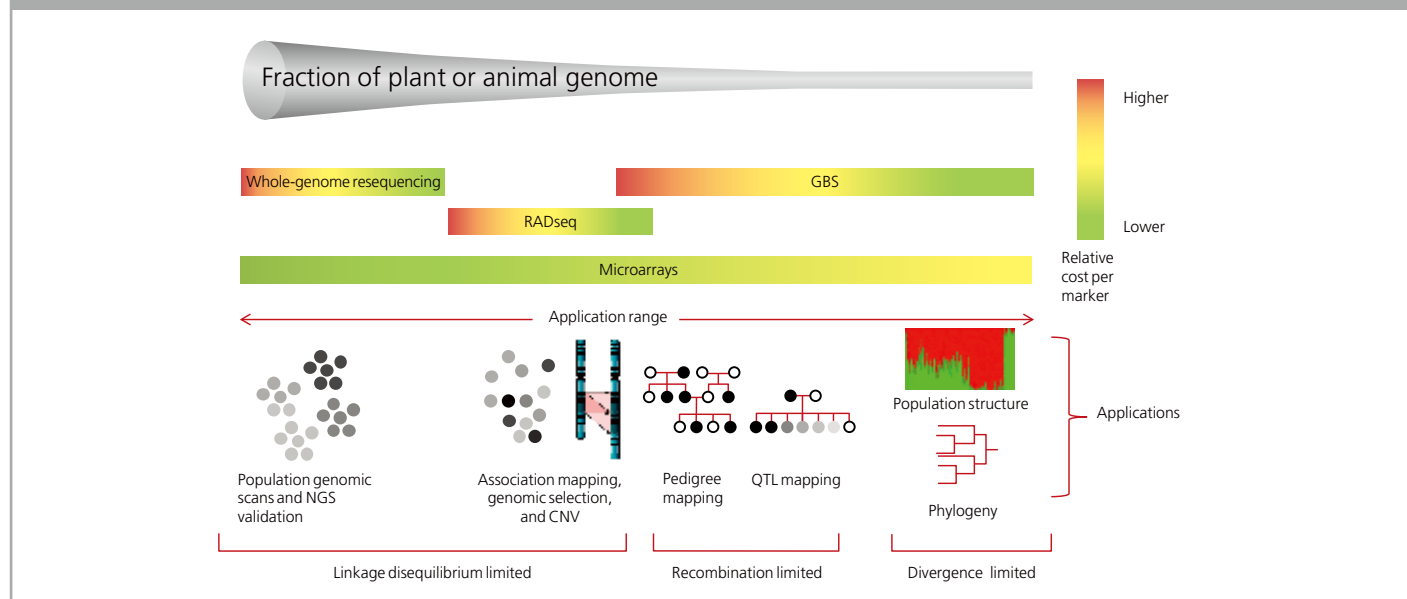
本技术指南讨论了每种新型测序技术中实验方法的影响、基因组复杂度对标记数量的影响,以及应用范围。实验方法的变化可明显增加任何基因分型项目的成本,对于30,000个标记的项目,可能增加五倍。

## 应用

基因组覆盖度高度依赖于基因分型技术和方法的选择,而这个选择最终取决于感兴趣的应用。每种方法提供了不同水平的植物或动物基因组覆盖度。这影响了可获取的标记数量,也决定了哪种方法最适合目标应用。这些应用的范围从群体基因组扫描到确定系统发育。图2显示了不同的基于测序和芯片的基因分型方法如何定位到各种应用,以及与覆盖基因组相关的相对成本。

每种基于测序的方法所覆盖的标记数量取决于实验参数,如限制性内切酶的类型、DNA的质量和数量以及分析技术。每个应用的标记数量如图2所示,是被测序的基因组部分的函数。准确基因分型所需的标记数量是基因组水平的连锁不平衡的数量、系谱中捕获的重组事件、各组之间分歧的函数。<sup>4</sup> 通过改变分析中的限制性内切酶,以增加标签数量,可提高基因组覆盖度。然而,正如上文指出的,提高基因组覆盖度是以更低的样品多重分析、更低的序列覆盖度或每个样品的更高成本为代价的。

图2: 各个农业基因分型应用的最佳基因组覆盖度以及成本相对于基因分型方法的扩展性的表示。颜色渐变表示利用每种方法获得所需数量的标记时每个标记的成本。测序成本随着待测序基因组部分的增加而增加。因此,在仅仅对数百个系统发育或群体结构的标记进行测序时,每个标记的成本较低,或与芯片成本不相上下。芯片作为一种经济高效的技术,可验证重测序或基于测序的基因分型技术所发现的标记,这些技术提供了筛查不同群体的能力。



- **群体基因组扫描和测序验证:** 芯片和基于测序的基因分型技术已经用于开展群体扫描和验证那些利用新一代测序发现的标记。基于测序的基因分型策略容易检出假的SNP,因为测序技术存在固有误差,拷贝数变异无法定位到参考基因组,或来自旁系同源或同源基因。通过更深度的序列覆盖,假的SNP可排除,但这会增加每个样品的成本,通过使用双单倍体或高质量的参考序列,也可避免这一问题,但这些会导致更复杂的信息学分析,而严格的过滤条件会丢弃大部分的测序数据。通过运行群体内的大量样品来验证标记,可鉴定出信息量大且重复的标记。高密度的Axiom®芯片已成功应用于验证测序发现和排除假的SNP,这些SNP是许多物种测序错误的结果,包括鸡<sup>5</sup>和三文鱼<sup>6</sup>。芯片带来了一种简单的方法,可以评估不同群体中的数百万个标记,并验证那些通过不同测序技术(如RADseq、RNAseq和重测序)发现的标记。

- **关联作图、基因组选择和拷贝数应用：**关联作图(AM)技术使用大量的多态性标记来克服QTL定位中的挑战和限制。关联作图依赖连锁不平衡和现有基因库中存在的重组，来开展随机交配群体、各系或种质间的表型-基因型关联。<sup>8</sup> 在关联作图研究中，更多的标记增加了找到或定位致病变异的可能性，<sup>9</sup> 因此标记越多越好。尽管关联作图可通过基于测序的基因分型方法来完成，但芯片通常能更经济高效地对高密度的标记进行基因分型，且有着更好的数据质量和完整性。

密集的标记也可用在基因组选择中，其中在基因分型和表型检测中同时估计标记的影响，或训练群体，然后用来预测选择候选物的价值。基因组选择的准确性随标记密度的增加而增加。据估计，50,000个标记已足以准确预测这些关系。<sup>10</sup> 拷贝数变异检测实现了复杂性状中可遗传变异的研究和鉴定。

- **系谱和数量性状位点(QTL)作图：**与关联作图不同，QTL作图查看多个基因对数量性状的影响，如三文鱼控制对海虱的抗性或鱼卵大小的QTL。QTL鉴定是基于双亲杂交，需要通过精细作图鉴定染色体区域的单个基因，因此需要大量的杂交来产生足够数量的减数分裂事件。系谱基因分型利用QTL检测中的育种材料，它们覆盖多代，通过多次杂交与系谱中的共同祖先相关联。这实现了育种项目中存在的大部分等位基因的鉴定和使用。
- **系统发育和群体定位：**在各个研究机构维护的种质和育种者维护的种质中，群体结构可能有所不同。不同的群体结构需要不同的全基因组关联研究(GWAS)方法。通过构建遗传或连锁图谱来调查群体结构和开展系统发育分析，可提供基因组重组率的信息。了解群体结构也有助于选择适当的标记和密度。通过研究很小一部分基因组中的标记，可完成群体分析。

**工作流程：**基于芯片的技术和基于测序的基因分型技术的工作流程比较如图3所示。基于测序的基因分型技术依赖条形码技术对样品进行多重分析(例如，96个样品在单个通道中测序将需要96个样品条形码)。文库制备需要选择适合该物种和所需标记数量的限制性内切酶。此过程需要优化，以避免引物二聚体等问题，这些可能增加测序的费用。在文库制备之后，真正的测序约需11个小时至11天不等，这取决于仪器的能力和测序基因组的百分比。更高的样品多重分析也并非不可能，但正如之前提到的，必须平衡基因组大小，测序基因组的百分比以及单个通道中的序列覆盖度。测序之后，数据经过过滤，条形码被解复用，以提取每个样品的标记。

使用任一基因分型技术的重要考虑因素是计算设备和分析流程。分析流程需要根据感兴趣的物种、实验方法、待研究的群体和技术本身来定制。

James Hutton研究所近期发表的一篇文章<sup>11</sup>得出结论，利用GBS来研究大麦的一个重要结果是，与目前实验室中使用的多重SNP分析技术相比，GBS数据在处理和随后的分析上更具挑战性。采用基于测序的基因分型技术存在诸多挑战，包括计算设备，维护定制分析流程的生物信息学专家，开展比对和分析的软件，以及提取有用的基因分型数据所需的时间。测序技术的数据分析通常在“云端”开展，以尽量减少本地数据存储和计算要求。变异检出往往通过定制的软件来开展，这些软件检出各种基因型。每个存储技术都具有与数据转移、存储和检索相关的成本，这会影响基因分型项目的成本。

相比之下，基于芯片的基因分型技术能轻松地利用台式工作站，对每个样品的数百万个数据点进行基因分型，在降低设备费用的同时提高操作效率。这种芯片法基因分型的简约、易用让芯片可在各种场景和环境下使用，对常规育种应用而言尤其有吸引力，因为大量样品处理和周转时间都很重要。

**文库制备中的DNA质量和数量：**基于测序的基因分型技术对DNA浓度和DNA质量的要求仍然是实际应用中的严重挑战之一。DNA测序需要几微克(μg)已纯化的高分子量基因组DNA，且无污染和共生体。细菌污染可能影响测序，因为DNA材料的随机扩增意味着细菌DNA会与待分型的生物样品一起被测序。

**标记的数量和类型：**芯片技术与测序技术的最大差异在于芯片靶定特定的基因组区域或特定SNP的能力，如图4所示。基于芯片的技术能够靶定特定染色体区域内任何数量的标记，其设计策略采用在基因组中均匀间隔的标记，如有必要，采用基因组特定区域内更高间隔的标记。这种灵活性让芯片可应用于GWAS、<sup>12</sup> QTL作图、关联作图和基因组选择，并具有一定的确定偏差。通过在多个品种上开展SNP研究，可降低确定偏差。

图3: 基于芯片的基因分型与任一种基于测序的基因分型的工作流程比较。

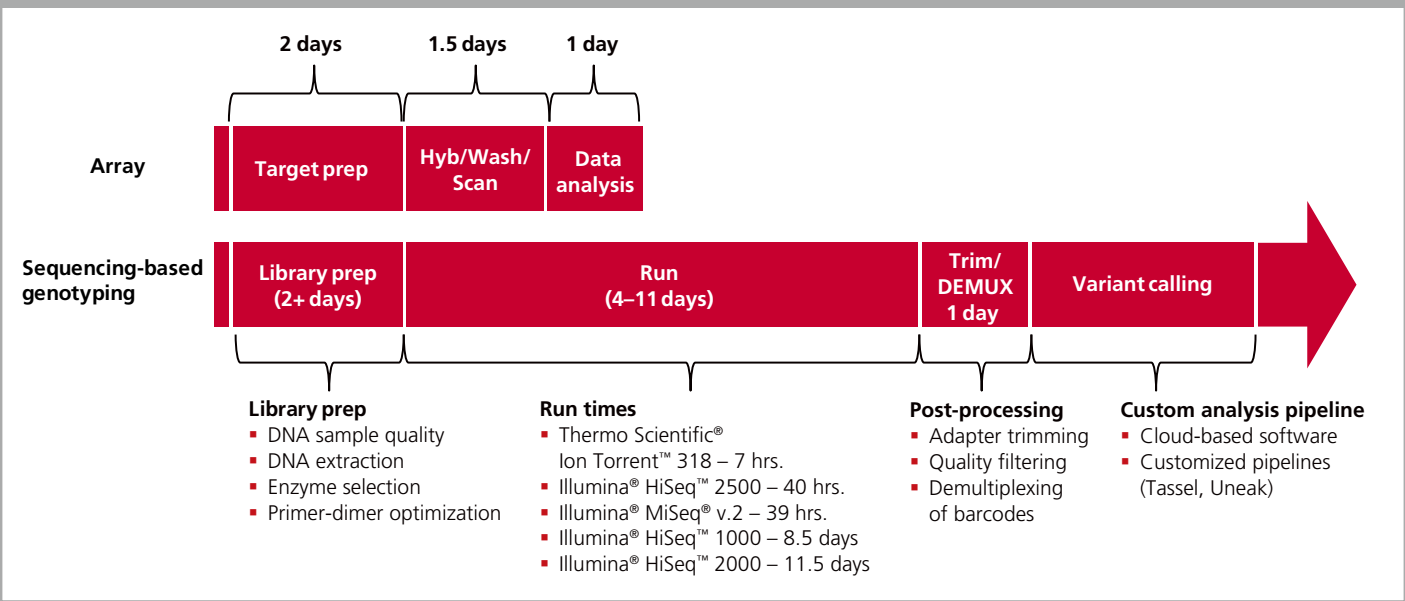
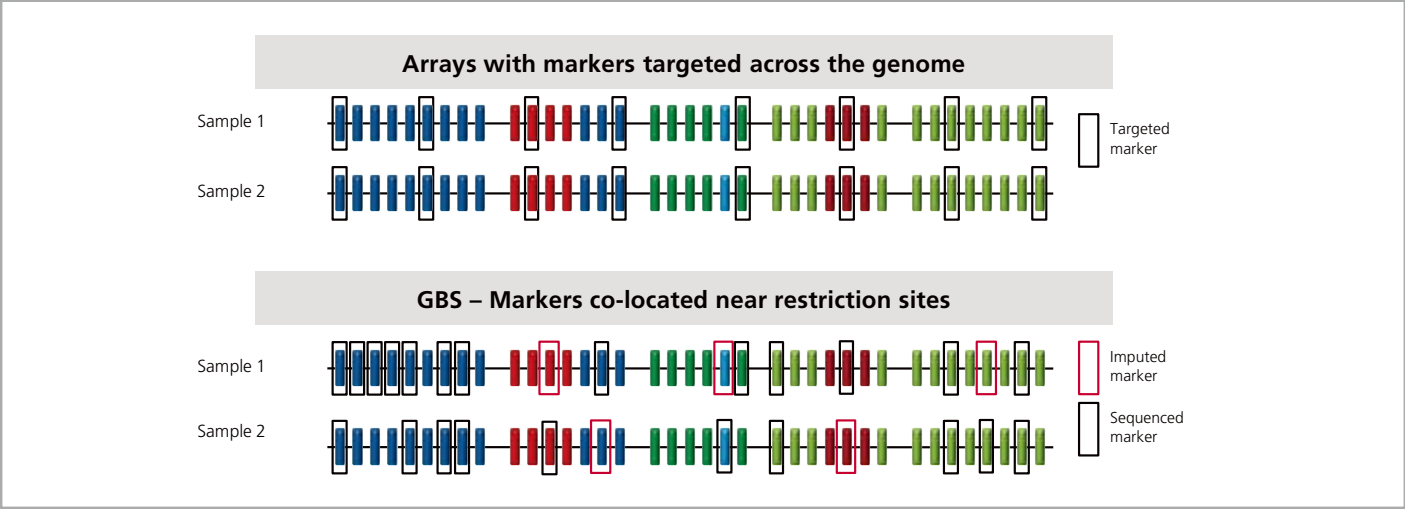


表1: 利用芯片和基于测序的基因分型技术对生物样品进行基因分型所需的DNA量。基于测序的基因分型技术所需的浓度是芯片的2倍至30倍。

Amount of DNA recommended	Microarrays	Sequencing-based genotyping methods
Diploid plants and animals	100 ng	200–3,000 ng
Polyploid plants and animals	150 ng	

图4: 图像显示了芯片和基于测序的基因分型技术的标记选择技巧。芯片能够利用靶定那些特定区域的标记, 靶定单倍型块。相比之下, 基于测序的基因分型技术则通过靠近限制性酶切位点的标记, 靶定整个基因组内的区域。



基于测序的基因分型技术依赖DNA库的随机抽样, 而标记的数量与待测序区域的数量和大小成正比。当待研究群体的限制性位点保守时, 基因组区域预计没有偏差。因此, 样品间的标记不保守, 并且没有两个样品能提供相同的一组标记。这导致数据丢失, 并需要复杂的信息学通过推算来恢复丢失的数据。样品间不保守的标记必须通过参考基因组来推算, 或利用覆盖度非常高的测序(18倍或更高)通过相关品系的单倍型来推算。



## 选择测序方法的考虑因素

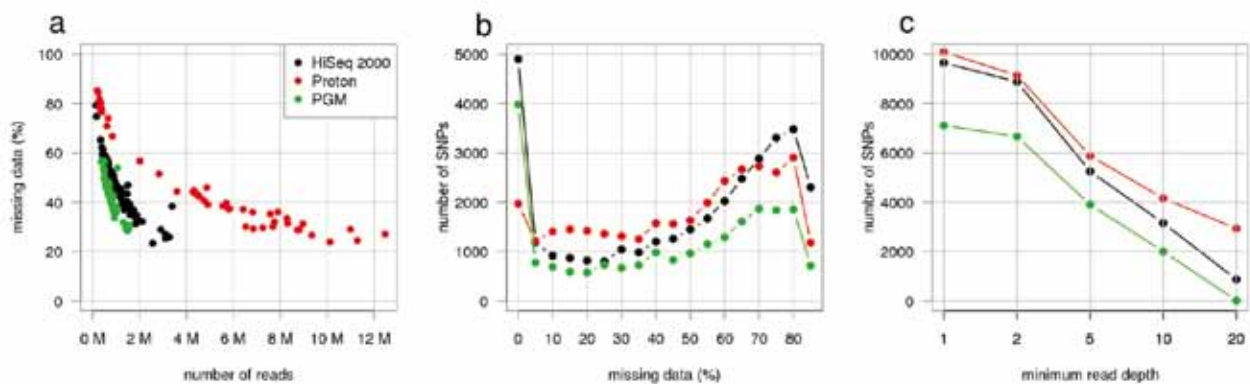
对于上面提到的任何应用，在决定采用哪种方法之前，必须考虑到影响基因型的各种因素。

- **杂合子检出错误**：基于测序的基因分型技术，尤其是GBS，依赖低覆盖度来降低成本，并获得大量标记，这些标记可用于关联作图。这种实验方法的缺点在于杂合子的检出明显过低，这影响了基因型准确性。GBS检出不足50%的杂合子。一项关于DNA测序所需覆盖度的研究<sup>14</sup>预测，对于每个杂合二倍体，检测99.75%位点上的两个等位基因至少一次，需要13.5倍的深度。而检测每个等位基因至少两次，将需要18倍的深度。增加测序覆盖度导致每个样品的成本更高，并使得测序比芯片更为昂贵。关于葡萄的研究表明，以5.7倍的平均深度基因分型时，30-50%的杂合子未检出。<sup>15</sup>而芯片上杂合子的检出准确性是由芯片设计决定的，这是高度可预测的，使得基因型检出准确性接近100%。芯片所使用的进一步设计方法能够对GC含量高于60%的基因组区域进行基因分型。
- **基因组覆盖度**：任何技术所带来的标记数量有望实现基因组的均一覆盖。基于测序的基因分型技术表现出数据丢失，这导致基因组的不均一覆盖。丢失的数据是实验条件和基因组结构的函数所造成，源于文库复杂度(即独特序列标签的数量)和文库的序列覆盖度的组合。丢失数据的量与文库制备的多重水平以及RE消化所使用的酶直接相关。测序技术中限制性内切酶的选择影响了等位基因信号丢失，从而影响群体遗传学的统计数据。稀有标记需要切割不频繁的酶，随后产生较少的标记。若使用频繁切割的酶，会产生较多的标记，但覆盖度明显降低，导致大量的数据丢失。

所有代表性降低的测序技术依赖基因组复杂度降低，从而降低成本并增加通量。复杂度降低的缺点在于所获得的基因型数据有着明显的丢失数据。<sup>16</sup>基因型数据可能丢失，因为基因组结构中的内在差异，如存在一缺失差异、多态性限制位点的变异，以及差异甲基化，这影响代表性降低的测序技术中所使用的甲基化敏感的酶。丢失的数据对QTL作图很重要，其中亲本系的基因型数据质量对作图群体的基因型检出至关重要。亲本系需要以非常高的覆盖度测序。

图5说明了序列覆盖度与丢失数据之间的关系，这项结果是由近期一篇论文发表的，它比较了不同平台上基于测序的基因分型。<sup>17</sup>此研究表明，在10倍覆盖度下，可获得1,000个标记，且50%的数据丢失，而低覆盖度下的标记数量增加至30,000个时，90%的数据丢失。

图5：近期一篇论文的数据说明了序列reads的数量、测序深度和read检出之间的关系，这些数据来自于GBS在大麦上的应用。(a)将每个测序样品中每个SNP的丢失数据的平均百分比相对于该样品的序列reads数量作图。(b)每个SNP的丢失数据的直方图。(c)将SNP检出的数量相对于为使基因分型成功特定样品中变异位置的最低深度作图。所有SNP检出都是通过SAMtools管道实现的。次要等位基因频率被设为30%，而丢失数据的最大值被设为50%。本研究所使用的测序平台包括Illumina® HiSeq™ 2000 (黑色)、Ion Torrent™ PGM (绿色)和Ion Torrent™ Proton (红色)。所有分析的颜色代码如(a)图中的图例所示。doi:10.1371/journal.pone.0076925.g002.



基因型数据的预期量和实际量可能差异巨大。最近一项使用GBS的玉米研究<sup>18</sup>表明，基因组位置分布的倾斜覆盖和不成比例的区域不代表最初预计的信息。这限制了测序技术的范围和应用，被认为无法用于关联研究的精细作图。大多数位点的基因型数据只能通过大幅增加read深度来实现，而这会影响测序成本。

丢失的数据可利用数据推算技术来恢复，也就是将数据与参考基因组比对，这需要大量的投资、先进的分析，以及复杂的流程，能过滤、排序并比对序列数据。缺乏简单易用且统一的信息学流程仍然是在常规应用中采用基于测序的基因分型技术的第二大障碍。推算特别适合亲缘关系相近的个体，但对于高度多样化的样品，丢失的数据可替换为近邻的等位基因。<sup>19</sup> 当丢失数据的比例高时，基于测序的基因分型技术也会丢失低频率的等位基因。替代方案是追求更高深度的测序，这会导致每个样品的成本更高。

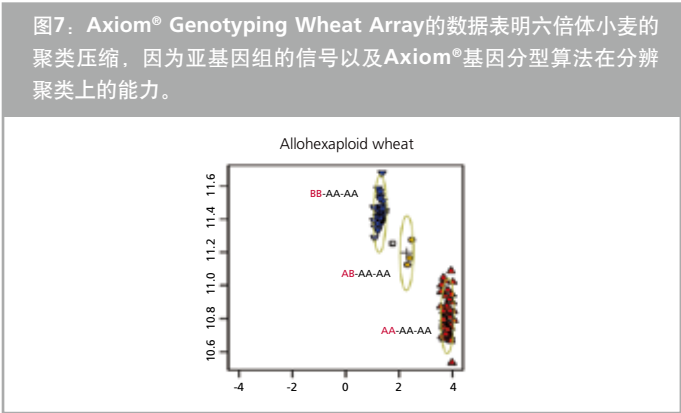
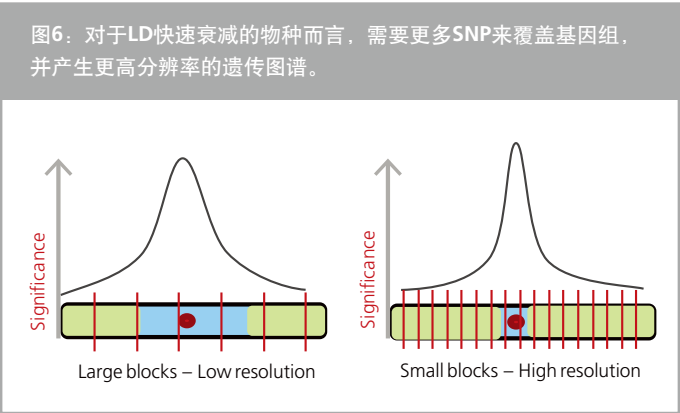
- **LD和多态性频率：**对于收集基因型数据的群体而言，它的基因组多样性和交配系统对测序成本有很大的影响。从一个较窄的遗传基础衍生而来的群体表现出较少的多态性，需要更多的测序，并增加总成本。四倍体棉花等物种便是如此，其每1,000-1,500个碱基表现出一个多态性。物种内的LD衰减也决定了多个群体的关联作图所需的标记数量。图6显示了LD衰减对标记分辨率的影响。对LD衰减高的物种而言，标记密度的低分辨率将导致基因组的覆盖度不足。水产养殖物种(如鳙鱼)和植物(包括玉米、葡萄和甜菜)表现出低的LD，在关联分析时需要大量的片段。近期一项关于鳙鱼的全基因组关联研究<sup>20</sup>使用了基于测序的基因分型技术，并得出结论，LD的快速衰减需要更高水平的标记密度，才能高效地开展全基因组关联研究。
- **拷贝数应用：**基于芯片和测序的基因分型技术可被用来开展拷贝数研究，以鉴定复杂性状的遗传变异。这两种技术都能检测拷贝数获得。但基于测序的基因分型技术在低覆盖度下难以鉴定拷贝数丢失，因为片段丢失显示为低覆盖度的标记。<sup>21</sup> 更高覆盖度将实现CNV丢失的检测，但成本有望增加40-50%。
- **基因组复杂度：**多倍性是植物和某些动物的更复杂属性之一。60-70%的被子植物是多倍体，其倍性水平从葡萄籽的四倍体到草莓的八倍体，而甘蔗更为复杂，其倍性水平从12-16倍不等。多倍体物种表现出基因组复制。多倍体的挑战如下：(i) 多倍体物种需要更高的序列覆盖度，才能高效覆盖更大的基因组，而这增加了测序成本。(ii) 基因组组装和作图算法很复杂，容易出错，特别是在组装旁系同源/直系同源区域时。对于多倍体且杂合的物种，每个指定位点的数据推算都需要复杂的分析流程，而这不能用于常规的育种应用。<sup>22</sup> 此外，更深度的测序增加了总成本。测定基因组中每个位点的等位基因剂量信息对基因组选择模式很重要。在使用芯片对多倍体物种进行基因分型时，来自亚基因组的信号导致聚类压缩。多倍体物种也表现出不同水平的倍性，因为干扰突变导致复杂度降低。常规的育种应用必须有一个分析流程，能自动聚类并分配基因型，以满足严格的育种时限要求。Axiom<sup>®</sup> GT1算法用贝叶斯统计来准确分配基因型并让多倍体基因组的数据聚类。图7显示了一个例子。自动流程让人们能够轻松准确地对数千个样品的数千个标记进行基因分型。

**为您的基因分型项目选择适当方案的指南：**鉴于芯片技术和测序技术的进步，科学家们需要认识到使用测序技术的挑战，以及使用测序和芯片技术的偏向。下列问题可帮助您选择适当的技术，应用在科研或基因组育种项目所考虑的物种上：

- 获取基因型数据所使用的分析是否与所考虑物种的基因组结构兼容，它是否能带来足够量的可靠标记？
- 是否有需要靶定特定染色体区域，需要采用何种标记策略以覆盖整个基因组？
- 所考虑物种的潜在LD结构如何？
- 此物种是不是多倍体，倍性水平如何？
- 将此技术引入科研或育种项目需要哪种信息学流程和专业知识？
- 需要多少个小时才能检出基因型并聚类数据？
- 所考虑的群体是近交群体，还是多样化的无关个体，预计杂合水平如何？
- 需要对多少个样品进行基因分型，周转时间或出结果的时间是否有限制？
- 需要何种深度的序列覆盖，才能准确检出基因型？
- 数据缺口有什么影响，您将如何恢复丢失的基因型？
- 考虑到丢失数据和生物信息学流程和分析所需的资源，分析的成本如何？
- 技术的通量、周转时间、分析的可靠性以及技术所使用的仪器如何？
- 育种项目可接受的偏差量如何，是否有办法绕过偏差？
- 需要将多少种不同的技术或分析整合到科研或育种项目中，进行有效验证、标记性状或常规使用？

芯片技术不断发展，形成Axiom® 384HT的格式。这个能以非常经济的价格点同时处理384个样品的创新让此技术从科研走向主流的商业化农业基因组学。在优先考虑周转时间、易用性和数据质量的应用中，芯片仍然是首选技术。

基于芯片的技术在单一平台上合并了多个基因分型应用，提供了灵活性和经济性。分析和信息学分析流程的创新让所有感兴趣的相关标记能够不受限制地基因分型，其结果可通过简单的流程在几小时内得到。Axiom®基因分型解决方案，来自Affymetrix的芯片技术演化，为全基因组分析到常规筛查的应用提供了完整的解决方案，具有最高的准确性和重复性、简化的流程和最低的成本。



基于测序的基因分型技术的挑战总结在表2。

表2：表2比较了基于测序的基因分型技术(如RADseq和GBS)与Axiom® Genotyping Arrays的特点。新技术较低成本忽视了实际实验条件和基因组复杂度，但却往往被宣传为替代芯片技术的理由。

属性	RADseq		GBS	Affymetrix Axiom® Genotyping Arrays
DNA质量	<ul style="list-style-type: none"><li>对质量和数量的严格要求(200 ng – 3,000 ng)</li><li>RADseq文库生产需要几微克纯化好的高分子量基因组DNA</li></ul>			100–200 ng DNA
标记选择	<ul style="list-style-type: none"><li>位于或靠近限制性位点的标记</li><li>无确定偏差</li></ul>			<ul style="list-style-type: none"><li>能够靶定特定的染色体</li><li>通过多个品系的重测序，可减少确定偏差</li></ul>
应用范围	<ul style="list-style-type: none"><li>不能靶定特定的染色体</li><li>不能控制与消化、连接和扩增相关的可变性</li><li>不能控制样品的消化位点(基因组中的突变、变异)</li><li>不能控制不同区域的read深度(RE敏感的甲基化)</li><li>数据丢失不能避免</li></ul> 对于参考基因组不可用的物种，能够对96个样品进行基因分型			<p>基于芯片上SNP的广泛应用</p> <ul style="list-style-type: none"><li>NGS验证</li><li>GWAS/标记性状关联</li><li>基因组选择</li><li>常规应用</li></ul> 不能用于SNP发现
基因组复杂度	<ul style="list-style-type: none"><li>增加遗传差异可能会影响酶切位点，并增加丢失数据的频率</li><li>大型基因组需要更多的测序</li><li>重复让组装很混乱</li></ul>			<ul style="list-style-type: none"><li>没有酶切位点的影响</li><li>能够拷问大型基因组或高杂合性基因组的标记</li></ul>
多倍体能力	<ul style="list-style-type: none"><li>多倍性、高杂合性和CNV需要深度测序</li><li>纯合6倍，杂合13.5倍(P ≥ 0.9975)</li><li>组装可能是嵌合的</li><li>不能检测CNV丢失</li></ul>			<ul style="list-style-type: none"><li>能够在1小时内分析异源多倍体数据(从四倍体到八倍体)</li></ul>
SNP的数量	1,000 – 150,000。 限于酶切位点附近的位置。		1,000 – 50,000。 限于酶切位点附近的位置。	多层SNP, 1,500 – 250万个SNP
周转时间	2-20周(服务供应商宣传的是20-24周)			4-5天，包括数据分析
分析	<ul style="list-style-type: none"><li>丢失的数据使推算复杂化</li><li>复杂的信息学</li></ul>			五步分析，通过基于Windows的Affymetrix Genotyping Console™ Software或基于命令行的Affymetrix® Power Tools Software

Affymetrix的农业基因组学基因分型方案为育种人员和研究人员提供了一种强大而经济的工具，可鉴定、验证和筛查植物或动物中复杂的遗传性状，实现更快速、更精确的育种。Axiom®基因分型一开始是从SNP文库资源中选择标记内容，接着设计SNP芯片，最后用芯片来鉴定样品的基因型。这为育种人员和研究人员提供了一种功能性的基因分型工具，让其应用在标记-性状关联、全基因组关联研究(GWAS)、数量性状位点(QTL)分析和基因组选择项目中。

鉴于genotyping-by-sequencing技术在数据管理、计算需求上的挑战，且定制信息学流程需要根据每个物种和样品群体来定制，芯片在数据质量、完整性、分析和常规育种的应用上仍是到目前为止最简单的技术。

总的来说，适用于动物和植物基因分型的Axiom®基因分型解决方案让人们能够为具有商业价值的物种定制芯片上的基因分型内容。Axiom基因分型解决方案包括物种特异和定制的芯片，其经过验证的基因组内容来自Axiom®基因组数据库，以及完整的试剂盒、数据分析工具，和一个利用GeneTitan®多通道(MC)仪器的全自动流程。

## References and publications

- 1 Baird N. A., et al. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* **3**:e3376 (2008). doi:10.1371/journal.pone.0003376
- 2 Elshire R. J., et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* **6**:e19379 (2011). doi:10.1371/journal.pone.0019379
- 3 Narum S. R., et al. Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology* **22**: 2841–2847 (2013). doi:10.1111/mec.12350
- 4 Peterson B. K., et al. Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS ONE* **7**(5):e37135 (2012). doi:10.1371/journal.pone.0037135
- 5 Kranis A., et al. Development of a high density 600K SNP genotyping array for chicken. *BMC Genomics* **14**:59 (2013). doi:10.1186/1471-2164-14-59
- 6 Houston R. D., et al. Development and validation of a high density SNP genotyping array for Atlantic salmon (*Salmo salar*). *BMC Genomics* **15**:90 (2014). doi:10.1186/1471-2164-15-90
- 7 Affymetrix application note *Mitigating sequencing errors, monomorphs, and poor performing markers during de novo SNP selection for genotyping applications* (2013) P/N DNA02261 Rev. 1
- 8 Ersoz E. S., Yu J., Buckler E. S. Applications of linkage disequilibrium and association mapping in crop plants, in *Genomics-Assisted Crop Improvement: vol 1: Genomics Approaches and Platforms*, eds. Varshney R. K., Tuberosa R. Springer, pp. 97-119 (2007). doi:10.1007/978-1-4020-6295-7\_5
- 9 Poland J. A., et al. Genotyping-by-sequencing for plant breeding and genetics. *The Plant Genome* **5**(3):92–102 (2012). doi:10.3835/plantgenome2012.05.0005
- 10 Meuwissen T., et al. Accelerating improvement of livestock with genomic selection. *Annual Review of Animal Biosciences* **1**:221-237 (2013). doi:10.1146/annurev-animal-031412-103705
- 11 Hui L., et al. An evaluation of genotyping by sequencing (GBS) to map the *Breviaristatum-e (ari-e)* locus in cultivated barley. *BMC Genomics* **15**:104 (2014). doi:10.1186/1471-2164-15-104
- 12 Liu S., et al. Development of the catfish 250K SNP array for genome-wide association studies. *BMC Research Notes* **7**:135 (2014). doi:10.1186/1756-0500-7-135
- 13 Cavangh C. R., et al. Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. **110**(20):8057–8062. doi:10.1073/pnas.1217133110
- 14 Wendl M. C., et al. Aspects of coverage in medical DNA sequencing. *BMC Bioinformatics* **9**:239 (2008). doi: 10.1186/1471-2105-9-239
- 15 Hyma K. E., GBS Usage Cases: Non-model Organisms. (2013). [http://cbsu.tc.cornell.edu/lab/doc/GBS\\_nonmodel\\_Sept\\_2013.pdf](http://cbsu.tc.cornell.edu/lab/doc/GBS_nonmodel_Sept_2013.pdf)
- 16 Davey J. W., et al. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics* **12**(7):499–510 (2011). doi:10.1038/nrg3012
- 17 Mascher M., et al. Application of genotyping-by-sequencing on semiconductor sequencing platforms: a comparison of genetic and reference-based marker ordering in barley. *PLoS ONE* **8**(10):e76925 (2013). doi:10.1371/journal.pone.0076925
- 18 Beissinger T. M., et al. Marker density and read depth for genotyping populations using genotyping-by-sequencing. *Genetics* **193**(4):1073–1081 (2013). doi:10.1534/genetics.112.147710
- 19 Huang X., et al. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nature Genetics* **42**(11):961–967 (2010). doi:10.1038/ng.695
- 20 Rexroad C. E., et al. Estimates of linkage disequilibrium and effective population size in rainbow trout. *BMC Genetics* **10**:83 (2009). doi:10.1186/1471-2156-10-83
- 21 Donato MD et al. Genotyping-by-sequencing (GBS): a novel, efficient and cost-effective genotyping method for cattle using next-generation sequencing. *PLoS ONE* **8**(5): e62137. doi:10.1371/journal.pone.0062137.
- 22 Brummer, et al. Applied genetics and genomics in alfalfa breeding. *Agronomy* **2**:40-61 (2012). doi:10.3390/agronomy2010040



Applied Biosystems  
技术支持服务中心  
800-820-8982  
400-820-8982

© 2017 Thermo Fisher Scientific Inc.

**ThermoFisher**  
SCIENTIFIC