

精准升级:CLR 与 HiFi 取长补短, 任你选择

PacBio是三代测序的主流平台之一, 共有CLR和HiFi两种测序模式。CLR测序模式称为超长测序模式(插入片段30 Kb以上), 产生的数据就是基于单循环测序的结果, 一个插入片段只测序一次, 准确率和PacBio 常规测序保持一致。

PacBio HiFi Reads是准确度超过Q20 (>99%), 并且还能兼顾超过10kb, 甚至可达20 kb的长读长测序模式。在HiFi测序模式下, 酶读长与CLR测序模式相当甚至更长(超过100 Kb), 但插入片段只有10-20 Kb, 因此测序时酶会绕着DNA模板(插入片段)进行滚环测序, 即插入片段会被多次测序。这样单次测序中造成的随机测序错误, 可以通过算法进行自我纠错校正, 最终得到高准确度的HiFi Reads。

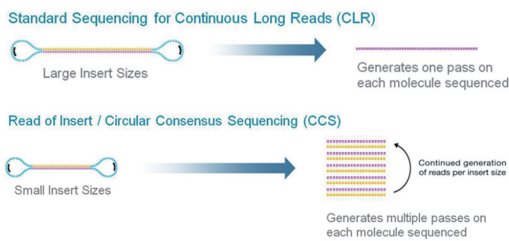


图1 CLR与CCS测序模式的对比

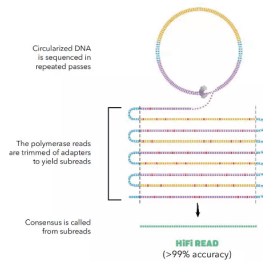


图2 HiFi测序原理

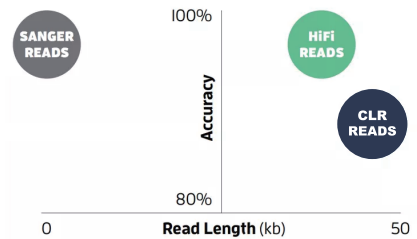


图3 CLR与HiFi读长与准确性的比较

表1 CLR与HiFi的对比

测序模式	文库构建	数据产出/cell	准确率	Subreads N50	在变异检测中的应用
CLR	30-40kb	100-200G	85%	25-30kb	不可检测SNV、InDel; 可检测SV大概位置, 断点不精确; 需要高深度来构建基因组
HiFi	10-20kb	15-30G	99%	10-20kb	可检测SNV、InDel、以及精确SV断点; 15X可用于构建基因组单体型(以人基因组为例)

HiFi Reads进行变异检测的技术要点

1. 应用方向

- (1) **开发分子标记**: 建立遗传多态性数据库, 为后续揭示进化关系、挖掘功能基因等提供数据基础。
- (2) **群体基因型分析**: 通过不同个体、群体之间比较分析SNV、InDel、PAV等变异, 来构建和完善物种核心种质资源, 进而可辅助遗传育种。
- (3) **癌症和疾病突变检测**: 通过在全基因组范围内检测各种变异类型精确定位SNV、InDel、SV, 为挖掘致病变异和致病基因提供参考依据。

2. 送样要求

表2 HiFi变异检测送样要求

测序模式	送样要求	建库总量 (DNA)	纯度	项目周期
HiFi	动物5g 以上 植物10g 以上	≥15μg	Nanodrop检测OD260处主峰明显且无其它杂峰, Nanodrop浓度/Qubit浓度 ≤2; 电泳检测胶孔处无中度蛋白污染;	45个自然日

3. 分析内容

数据质控, 与参考基因组比对, 根据比对结果, 进行SNP、InDel和SV检测及注释。分析流程图如下图所示。

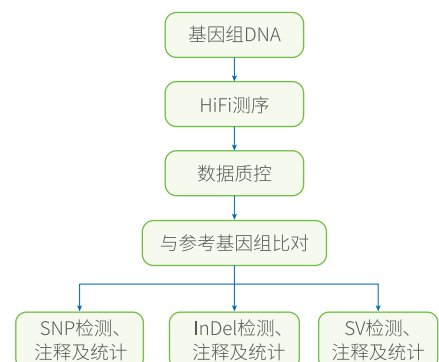


图4 HiFi变异检测分析流程图

运用HiFi Reads进行变异检测的测序深度推荐

1. 文献数据

15×的HiFi Reads可检测到99.5%的SNVs和90%的InDel (与NGS相同), 15X深度时检测SV也可达到90%的检出率和准确率。

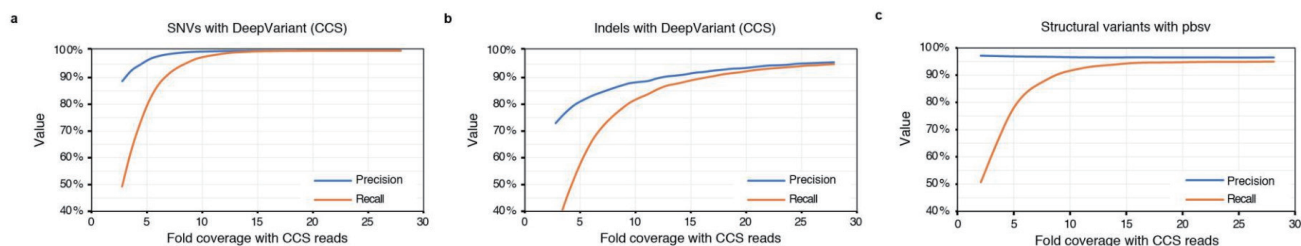


图5 不同测序深度对SNV、InDel和SV的检测结果

2. 菲沙实测数据

(1) 不同测序模式检测SV的IGV图比较

与CLR相比, HiFi Reads检测到的SV边界更明显, SV连续性更好。

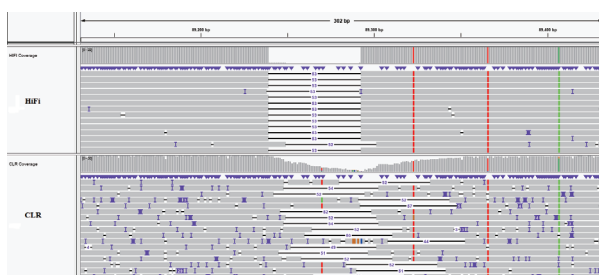


图6 不同测序模式检测SV的IGV图比较

(2) 不同HiFi Reads测序深度的Coverage率比较

实测数据表明, 5×以上的测序深度Coverage率超过98%, 当测序低于5×时, Coverage率急剧下降。

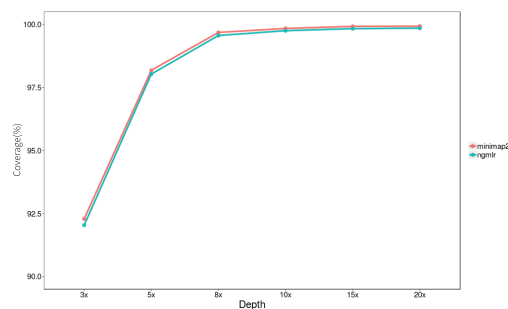


图7 不同测序深度Coverage率的比较

(3) 不同HiFi Reads测序深度的SV检测结果比较

利用HiFi Reads对某植物进行变异检测, 结果表明当测序深度在3-15×之间时, 随着测序深度的增加SV检测数目也增加; 而当测序深度在15-25×之间时, 随着测序深度的增加, SV的检测数目逐渐趋于饱和状态。

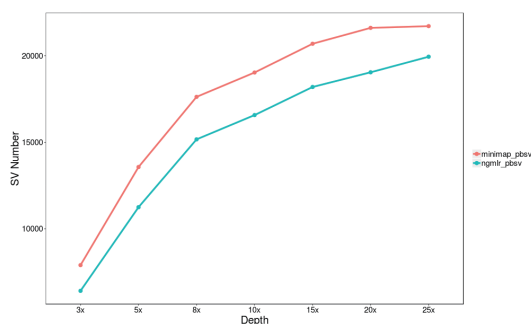


图8 不同测序深度SV检测结果的比较

参考文献:

Wenger A M, Peluso P, Rowell W J, et al. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome[J]. *Nature Biotechnology*, 2019, 37(10): 1155-1162.

