基因组测序

1. 第一代 DNA 测序【sanger 测序技术】:
   1.1. 原理：以待测 DNA 为模板，使用带有标记的碱基类似物体外合成新链，可在任意一个碱基位置终止 =》凝胶电泳时形成彼此只差一个碱基的梯形条带 =》得到序列。
   1.2. 原理：链终止法（完成了人类基因组计划）、化学降解法。
   1.3. 凝胶电泳对于信号捕捉是存在缺陷的，且不高效 =》毛细管电泳、高效毛细管电泳
2. 第二代 DNA 测序:【高通量测序平台】
   2.1. 焦磷酸测序【光点测序】
   2.2. DNA 芯片测序
   2.3. 将二代测序技术按测序原理分类：
      2.3.1. 边合成边测序 SBS：454、illumina、HiSeq/MiSeq/NextSeq、Ion torrent/proton
      2.3.2. 边连接边测序 SBL：solid【缺点：读长短】
3. 第三代测序技术:【高通量测序平台】
   3.1. 原理：单分子测序 SMS
   3.2. 特点：
      3.2.1. 测序读长：平均测序读长达到 10 ~ 18 kb ，最长可超过 60 kb ；
      3.2.2. 准确度高：测序深度达到 30× 时，准确度达到 99.999% （ Q50 ）；
      3.2.3. 敏感性强：可以检测频率在 0.1% 的 Minor Variants ；
      3.2.4. 无 PCR 扩增偏好性：样本不需要进行 PCR 扩增，避免了覆盖度不均一以及 PCR Artifacts 的产生；
      3.2.5. 最小的 GC 偏好性（ GC bias ）：在极端高 GC 和极端低 GC 区域，可以轻松测 定，从而保证序列的均匀覆盖度；
      3.2.6. 可直接检测碱基修饰：利用测序过程聚合酶反应的动力学变化，首次实现在 测序的同时对碱基修饰进行直接检测
   3.3. 技术路线：
      单分子实时测序技术 SMRT：读长超过 10kb，插入缺失错误率 1%【Sparc】
      纳米孔单分子技术【Sparc】:纳米孔单分子 DNA 电流阻遏、纳米孔单分子 DNA 碱基序列的电子阅读
4. 全基因组测序注意事项：
   4.1. 基因组测序的覆盖面 P0 = e-m【P0 ：丢失概率、m：为覆盖面（单倍体基因组数）】
      m= 1，P0 = 37% ，覆盖率 63%
      m=5， P0 = 0.67% ，覆盖率 99.33%
      m=10， P0 = 0.0045% ，覆盖率 99.9955%
   4.2. 物理间隙（Physical gap） 和 序列间隙（Sequence gap）
      4.2.1. 物理间隙：构建基因组文库时被丢失的 DNA 序列
      4.2.2. 序列间隙：测序时遗漏的序列，这个序列仍保留在尚未挑选到的克隆中
   4.3. 插入片段的两端测序：
      同一个载体的两段有两个引物，每个克隆读序只有 400bp，每个克隆内部不能进行连续的测序，因为缺少引物，所以>800bp 的片段中间就存在 gap
5. 序列读取模拟软件：ART、pIRS、PBSIM、Wessim、IgSimulato

6. ART_454:

6.1. 单末端测序 SINGLE-END SIMULATION art_454

6.2. 双末端测序 PAIRED-END SIMULATION art_454

6.3. 扩增子测序 AMPLICON SEQUENCING SIMULATION art_454

PS:SAM 是一种序列比对格式标准， 由 sanger 制定，是以 TAB 为分割符的文本格式。

主要应用于测序序列 mapping 到基因组上的结果表示、表示任意的多重比对结果

SAM 分为两部分：注释信息部分（header section）、比对结果部分（alignment section）

GenomeABC: http://crdd.osdd.net/raghava/genomeabc/