



武汉大学

Wuhan University

基因组与基因组学的研究方法和技术

武汉大学生命科学学院

丁毅

2016-10-25





武汉大学

Wuhan University

主要内容

一、遗传学定律与基因组的基本概念

二、遗传图的构建方法

三、物理图的构建方法

四、基因组与基因组学的研究方法和技术



一、遗传学定律与基因组的基本概念

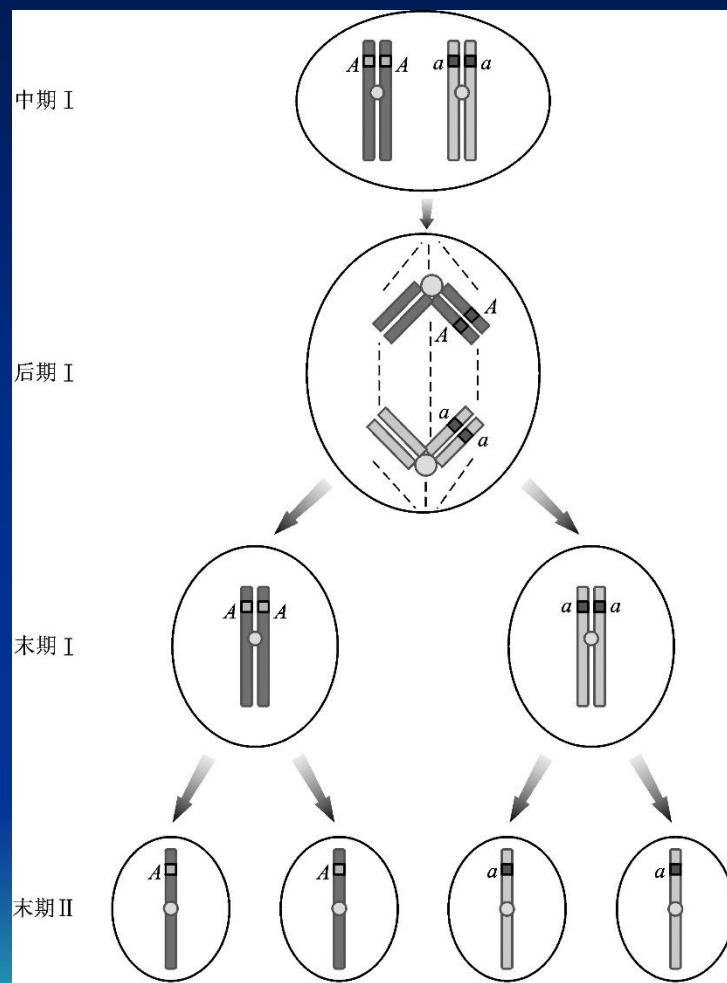
1、遗传学三定律

(1) 分离定律

(principle of segregation)

Mendel's First law The two members of a gene pair (alleles) segregate (separate) from each other in the formation of gametes; half the gametes carry one allele, and the other half carry the other allele.

对孟德尔的分离定律可以这样理解：**在第一次减数分裂时，由于同源染色体的分离，使位于同源染色体上的等位基因分离，从而导致性状的分离。**



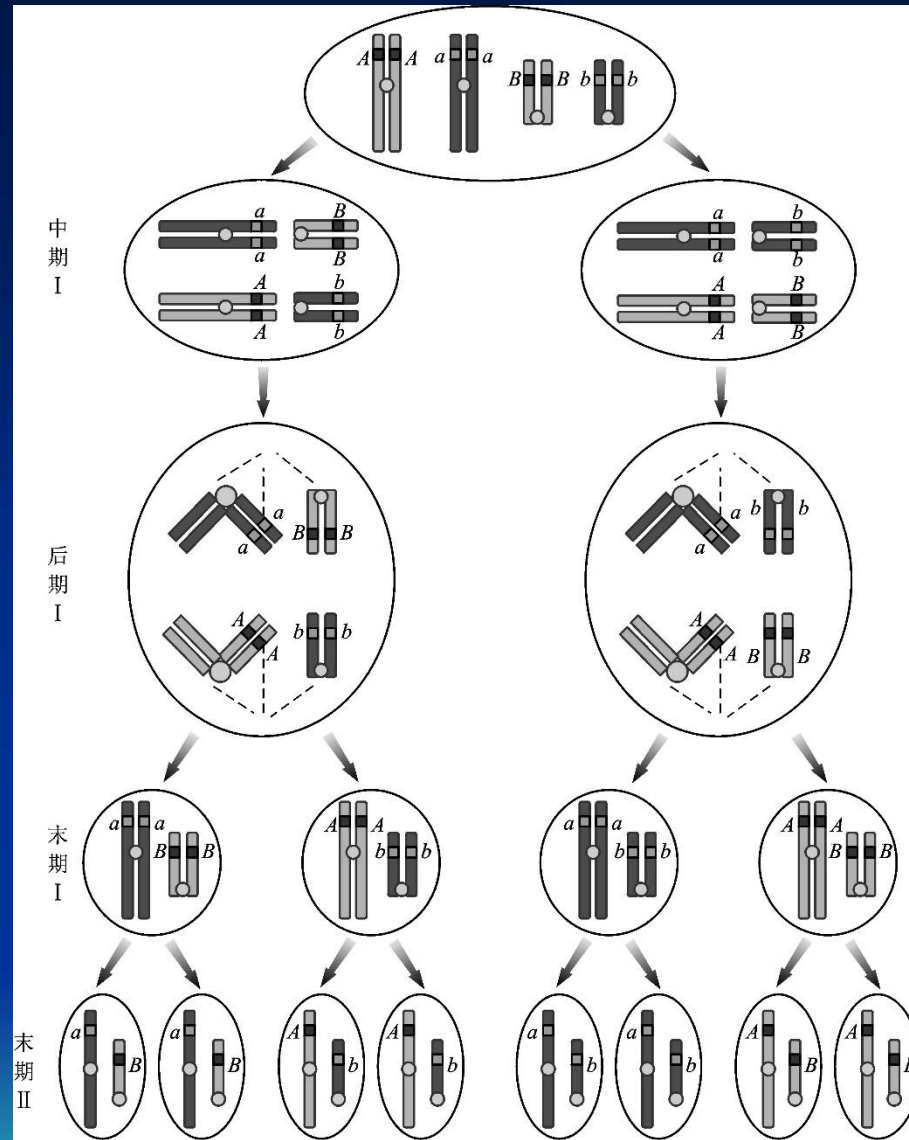
孟德尔分离定律的染色体基础
(以一对同源染色体为例的模式图)

(2) 自由组合定律 / 独立分配定律 (law of independent assortment)

Mendel's second law, the principle of independent assortment, states that genes for different traits assort independently of one another in the production of gametes.

不同对的遗传因子在形成配子中自由组合

由于决定不同性状的两对非等位基因分别位于两对非同源染色体上，形成配子时同源染色体上的等位基因分离，非同源染色体上的非等位基因以同等的机会在配子内自由组合，从而实现性状的自由组合。



孟德尔独立分配定律的染色体基础
(以两对非同源染色体为例的模式图)



(3) 基因的连锁和交换定律

在进行减数分裂形成配子时，位于同一条染色体上的不同基因，常常连在一起进入配子；在减数分裂形成四分体时，位于同源染色体上的等位基因有时会随着非姐妹染色单体的交换而发生交换，因而产生了基因的重组。

或者：

在生殖细胞形成过程中，位于同一染色体上的基因是连锁在一起，作为一个单位进行传递的；在生殖细胞形成时，一对同源染色体上的不同对等位基因之间可以发生交换。

或者：

是指位于同一染色体上的基因联合在一起伴同遗传的频率大于重新组合的频率，重组体或重组子（recombinant）的产生是由于在配子形成过程中同源染色体的非姊妹染色单体间发生了局部交换。

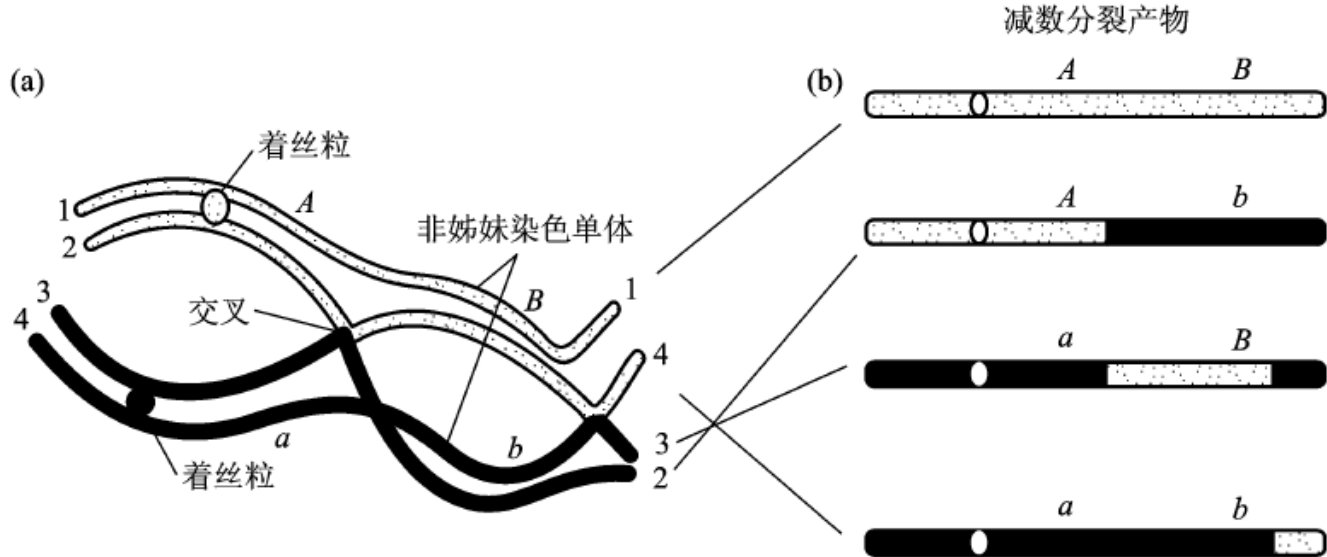
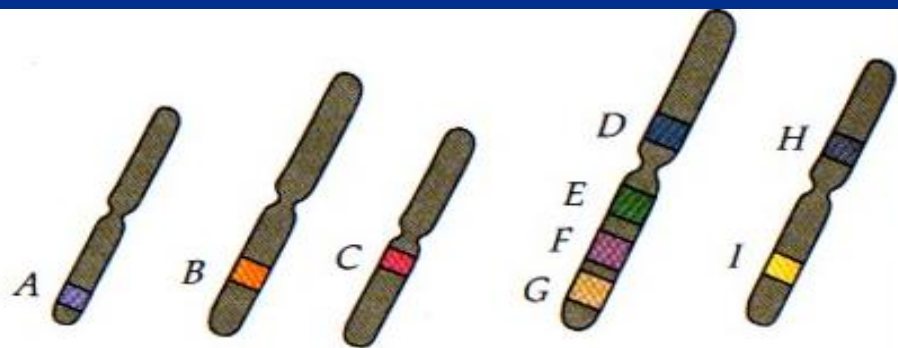


图 5-14 交叉及其产物的模式图

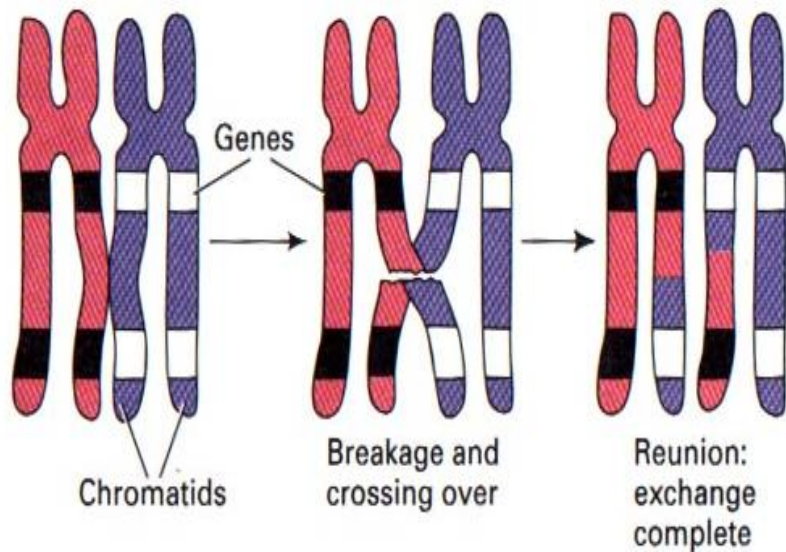
(a) 染色体交叉的模式图 (b) 减数分裂的产物



A, B, C: Unlinked genes
D-E-F-G: Genes on the same linkage group
H-I: Genes on the same linkage group

[*A*], [*B*], [*C*], [*D-E-F-G*], [*H-I*]: Five linkage groups

Homologous chromosomes





2、基因组

基因组 (genome) (1922年出现在遗传学的文献中):

一个细胞或病毒所包含的全部基因

通常在真核生物中指**一个物种的单倍体染色体组所含有的一整套基因**, 所以, genome被译作染色体组, **指的是单倍体细胞中所含的整套染色体**, 但现在基因组这个名词逐渐替代了染色体组。

原核生物一般只有一个环状的DNA分子, 其上所含有的基因为一个基因组。

真核生物细胞中的细胞器如叶绿体、线粒体中的DNA一般也为环状, 构成叶绿体基因组和线粒体基因组



现代基因组的概念:

Genome -- The complete set of sequences in the genetic material of an organism. It includes the sequence of each chromosome plus any DNA in organelles. (genes IX/X/Lewin's Genes XI)

- 基因组DNA测序的结果表明基因组中不仅包含着整套基因的编码序列，同时还包含着大量非编码序列，即基因之间的序列。这些序列同样包含着遗传指令 (genetic instruction)。因此，**基因组（应该）是整套染色体所包含的DNA分子以及DNA分子所携带的全部遗传指令。**



3、基因组学（genomics）：

研究基因组的组成、结构和功能的学科。

结构基因组学（structural genomics）：

是着重研究基因组的结构并构建高分辨的遗传图、物理图、序列图和转录图以及研究蛋白质组成与结构的学科；

功能基因组学（functional genomics）：

主要是利用结构基因组学研究所得到的各种信息在基因组水平上研究编码序列及非编码序列生物学功能的学科。



二、遗传图的构建

1、**遗传图**：A map of the position of loci or other genetic markers on a chromosome obtained by measuring recombination frequencies between markers

又称**连锁图**（linkage map），是指确定基因或DNA标记在染色体上的相对位置与遗传距离。它是通过连锁分析，计算遗传标记（或基因）间的交换频率，将某一染色体上的基因呈直线排列，确定基因之间的相对位置，一般用厘摩（cM）表示。

2、**经典的遗传学图谱**：

主要用来确定生物体的基因在染色体上的排列，只能标明基因之间的相对位置，无法指明基因在染色体上的具体位置，因此无法按图直接克隆分离基因。在现代基因组计划中利用价值不大。

3、**现代遗传学图谱**：

其概念是David Botstein等于1980年提出来的，当时由于DNA限制性内切酶和连接酶的应用，RFLP成为一种崭新的DNA多态性标记。他们提出来利用RFLP作为标记去构建多态性基因与这些标记连锁关系，进而确定多态性基因的位置。其精髓在于将单纯的表型研究深入到DNA分子的本质上去。



(一)、植物基因组遗传图谱构建的基本步骤

- 1、选择DNA分子标记
- 2、亲本的选配
- 3、分离群体类型的选择与构建作图群体
- 4、群体大小的确定
- 5、连锁图谱制作的统计学原理
- 6、DNA标记分离数据的处理与构建标记的遗传图谱





1、选择DNA分子标记

DNA分子标记：也称DNA多态性标记，是 DNA水平上遗传多态性的直接反映，有多种DNA分子标记：

- ✓ RFLP (Restriction fragment length polymorphism)
- RAPD (Random amplified polymorphic DNA)
- SCAR (Sequence characterized amplified region)
- STS (Sequence tagged site)
- SSCP (Single strand confirmational polymorphism)
- ✓ SSR (Simple sequence repeat)
- ISSR (Inter simple sequence repeat)
- AFLP (Amplified fragment length polymorphism)
- CAPS (Cleaved amplified polymorphic sequence)
- ✓ SNP (Single nucleotide polymorphism)
- ✓ InDel (insertion-deletion)



共显性的DNA分子标记

通常我们选择具有共显性的DNA分子标记用于遗传图谱的制作。

用做双亲的、它们间的DNA序列的多态性是发展分子标记的基础，基于PCR的共显性标记是常用的标记类型

通常共显性标记有四种类型：

RFLP (Restriction fragment length polymorphism)

SSR (Simple sequence repeat)

SNP (Single nucleotide polymorphism)

InDel (insertion-deletion)





(1) RFLP标记 (restriction fragment length polymorphism):

同一物种的亚种、品系或个体间基因组DNA 受到同一种限制性内切酶作用而形成不同的酶切图谱的现象，称为

限制性片段长度多态性 (RFLP)

这是由于基因组DNA某一位点上的变异有可能引起该位点特异性的限制性内切酶识别位点的改变，包括原有位点的消失或出现新的酶切位点，致使酶切片段长度随之发生变化而产生。



对RFLP的检测主要是用Southern杂交的方法进行

基本流程：

组织或细胞→基因组DNA→限制性内切酶消化
→琼脂糖凝胶电泳→印迹转移至滤膜→预杂交
→加入探针→杂交→洗膜→放射自显影。



RFLP标记的主要特点是：

- (1) 遍布于整个基因组，低拷贝序列；
- (2) 无表型效应，不受发育阶段及器官特异性限制；
- (3) 共显性，可区分纯合子和杂合子；
- (4) 结果稳定、可靠；
- (5) DNA需要量大，检测技术繁杂，难以用于大规模的育种实践中。

用途：RFLP主要用于群体水平和系统发育研究上

进行：个体识别；绘制遗传图谱；目的基因定位；
检测群体内或群体间序列差异程度；辅助育种等



(2) SSR (simple sequence repeats) 标记:

简单重复序列多态性，又称微卫星DNA多态性，即由二核苷酸，三核苷酸或四核苷酸串联重复的拷贝数目不等而出现的多态现象。

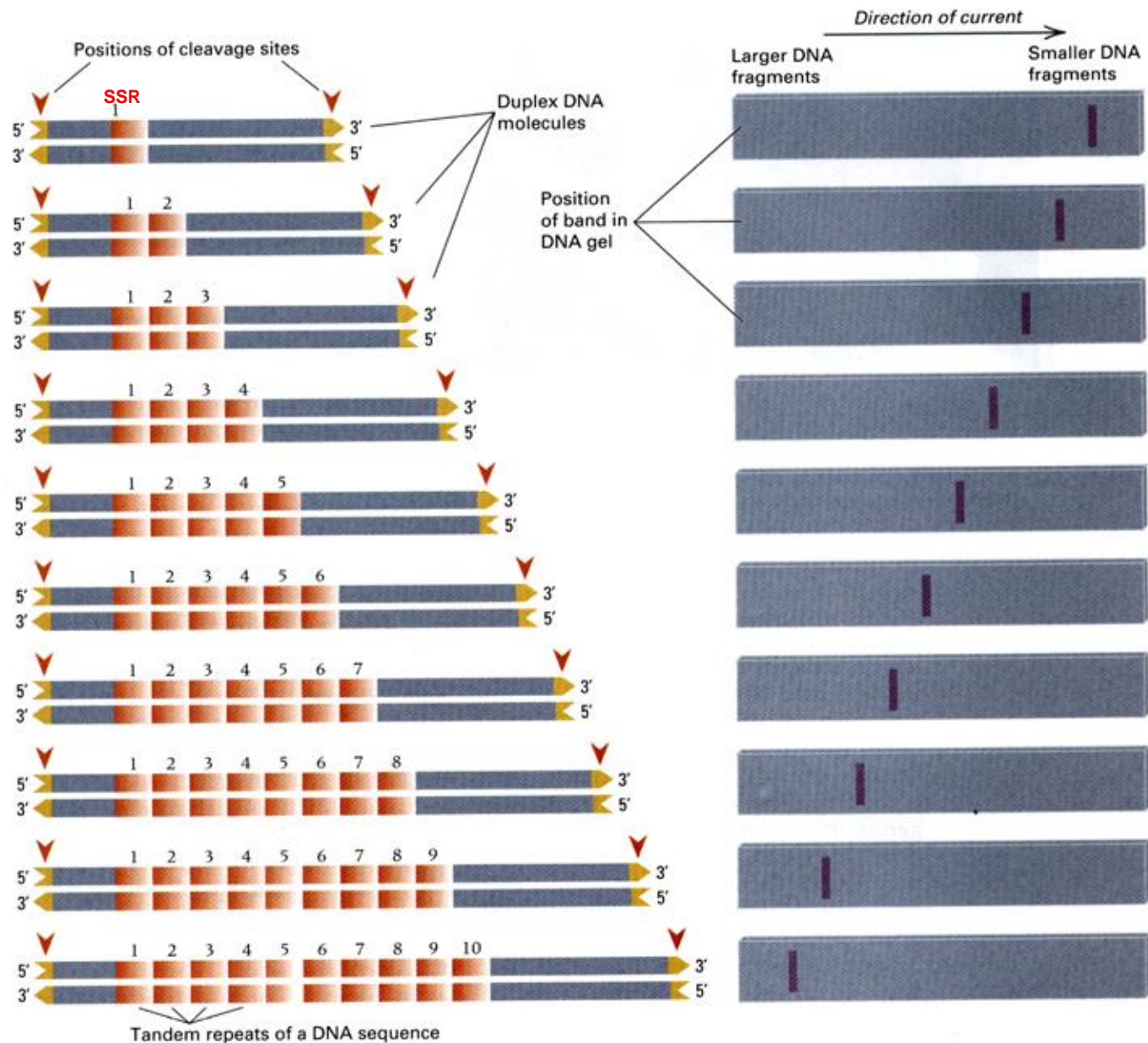


Figure 2.28 In a simple tandem repeat polymorphism (STRP), the alleles in a population differ in the number of copies of a short sequence (typically 2–60 bp) that is repeated in tandem along the DNA molecule. This example shows alleles in which the repeat number varies from 1 to 10. Cleavage at the sites shown, flanking the STRP, yields a unique fragment length for each allele. The alleles can also



典型的孟德尔式遗传；共显性（两个亲本的性状在一个个体中同时出现，没有显隐关系）。

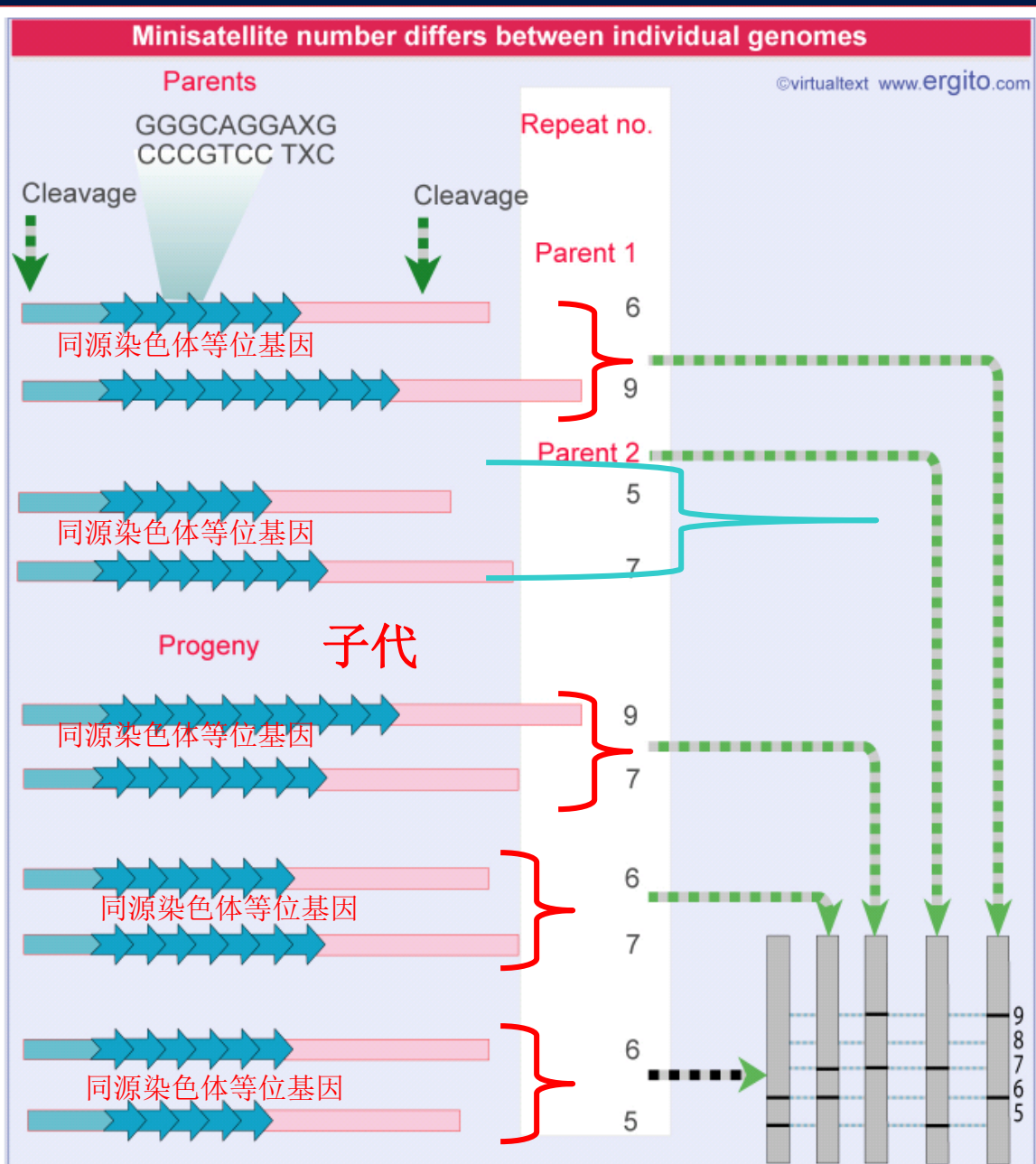


Figure 4.27 Alleles may differ in the number of repeats at a minisatellite locus, so that cleavage on either side generates restriction fragments that differ in length. By using a minisatellite with alleles that differ between parents, the pattern of inheritance can be followed.



SSR技术的优点是：

- (1) 在基因组中随机分布，检测的多态性频率高；
- (2) PCR特异引物，重复性好；
- (3) 共显性，操作相对简单。

问题是在没有进行基因组DNA测序的生物中：

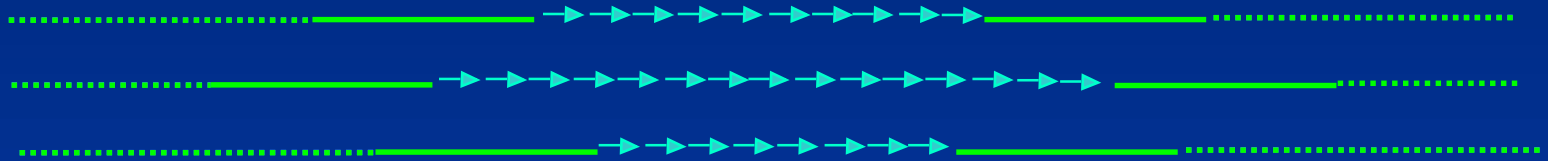
- (1) SSR需要测序和设计引物，因而需要大量的人力、物力和时间；
- (2) 另外其种属特异性强，开发所需的费用高昂，因此一些实验室进行了合作，共同开发微卫星引物

Simple Sequence Repeats (SSR)

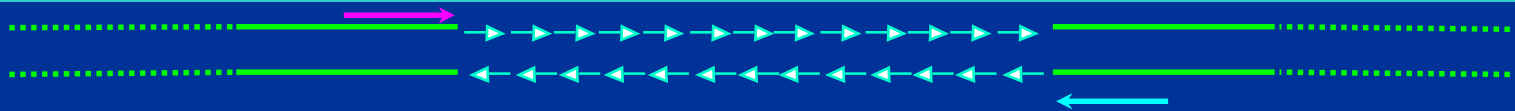
➤ 结构



➤ 不同个体间单元重复数高度可变



➤ 设计与侧翼区域配对的引物 (⇔)



➤ 扩增重复区域



SSR标记的基因变异来源于基因组DNA复制时的滑动



显然，利用SSR标记的关键是引物的设计。

对于已经进行基因组全序列测定的生物或遗传学研究比较经典的生物，可以直接从其基因组进行查找和利用有关程序进行引物的设计或利用别人已经发表的引物来进行实验；而对于没有基因组序列信息的生物，就需要进行有关引物的设计工作。



武汉大学

Wuhan University

SSR标记应用：

微卫星DNA 存在于绝大多数真核生物基因组中，已广泛应用于：

遗传图谱构建

品种指纹图谱绘制

品种纯度检测及目标性状分子标记筛选

遗传多样性分析

重要性状基因的定位等



(3) SNP (single nucleotide polymorphism)

标记:单核苷酸多态性。

同一物种不同个体基因组DNA的等位序列上单个核苷酸存在差别的现象。或者说不同个体基因组DNA序列同一位置上的单个核苷酸的差别。其比较的不是DNA的片段长度，而是相同序列长度里的单个碱基的差别。 (SNP)



SNP是指基因组内特定位置上存在两种不同的碱基，是二等位多态性，其中最少一种在群体中的频率不小于1%；如果出现频率低于1%，则视作点突变。

SNP只涉及单个碱基的变异，这种变异可以由单个碱基的转换（包括C与T互换，在其互补链上则为G与A互换），或颠换（包括C与A、G与T、C与G、A与T互换）引起，也可以由碱基的插入或缺失所致。但是，通常所说的SNP并不包括后两种情况。

根据SNP在基因中的位置，SNP可分为：

基因编码区SNP（coding SNP, cSNP）

基因周边SNP（peripheral SNP, pSNP）

基因间SNP（intronic SNP, iSNP）



尽管单一的SNP所提供的信息量远小于现在常用的分子标记，但SNP的数量极其丰富，并且可以自动化检验，因此其具有广泛的应用前景。

例如：如果假定1/1000的碱基是多态的话，那么人类30亿碱基当中有约300万SNP位点。由此可见，SNP数量比微卫星标记数要高出几个数量级。



例如：3-4个SNP这种相邻的界标构成的单体型 (haplotype)

就可以有8-16种：

3个SNP
可以形
成的单
倍型8种

1	+	+	+
2	+	⊗	+
3	+	+	⊗
4	⊗	+	+
5	⊗	+	⊗
6	⊗	⊗	+
7	+	⊗	⊗
8	⊗	⊗	⊗

Haplotype: 又称“单倍型”，“单元型”。一条同源染色体上的等位基因或遗传标记所构成的组合。

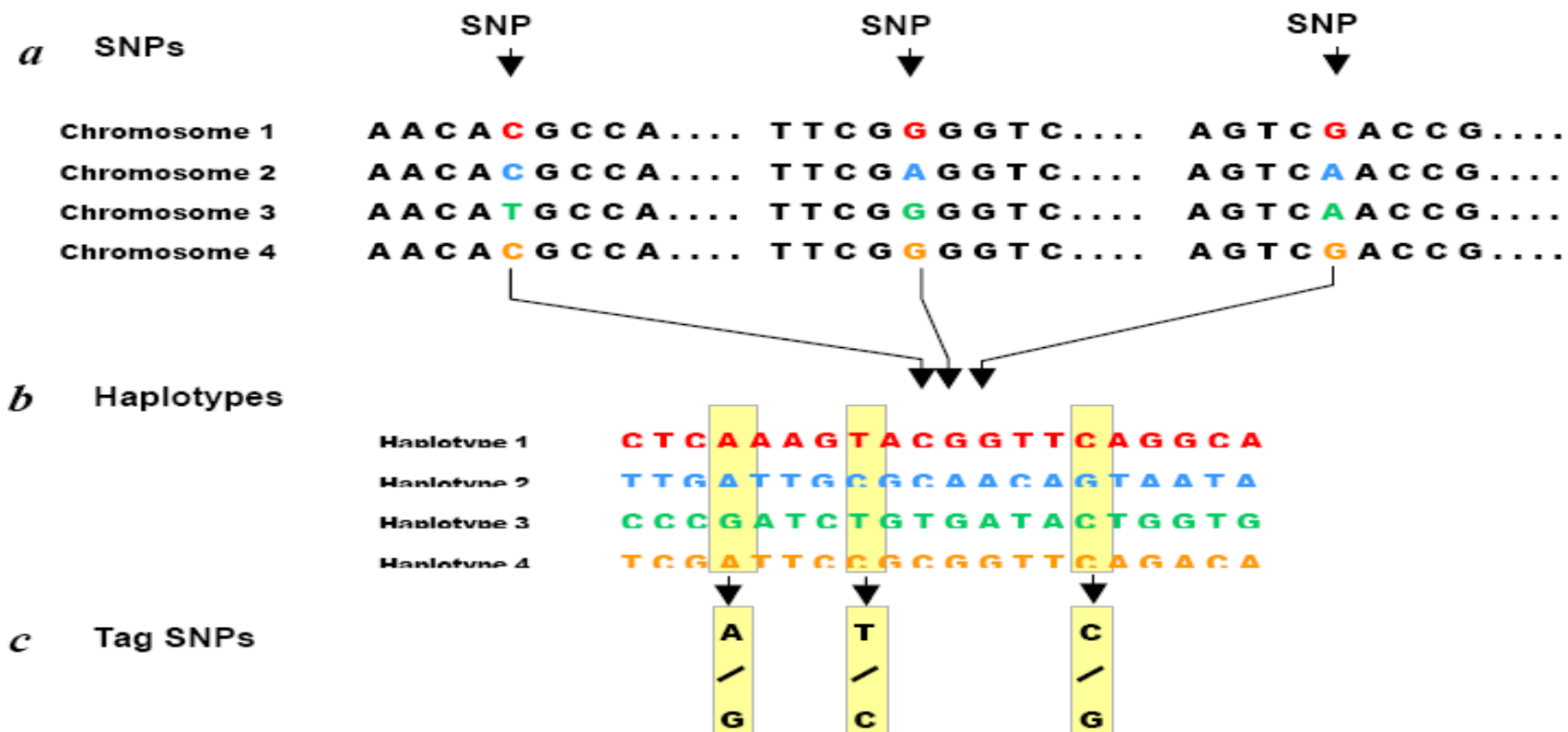
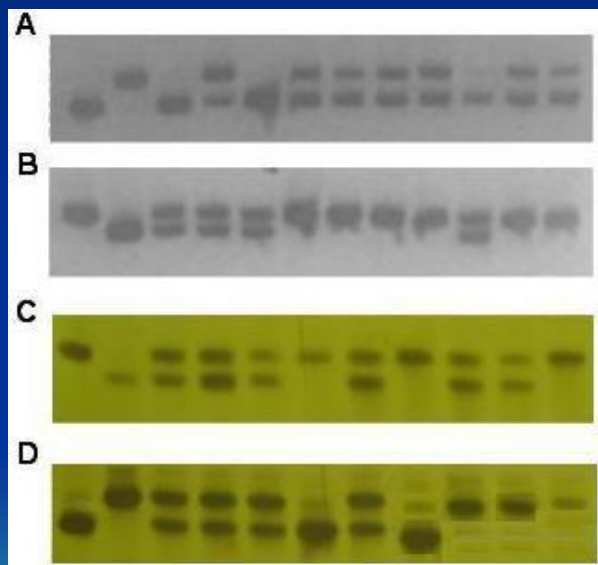


图 1 单体型图示意。a 一段 DNA 上的 SNP 位点。来自不同个体的四条染色体的这一段 DNA 的大部分序列完全相同但有 3 个位点表现出差异。每个 SNP 有两个可能的等位位点 (alleles)。b 单体型。一个 Haplotypes 由临近的一系列 SNP 等位位点组成，图中显示从一个 6000 bp DNA 片断上检定出的 20 个 SNPs 的排布类型，包括 a 中的 3 个位点 (第二排箭头)。c 标签 SNPs。只要检定分型 20 个 SNPs 中的 3 个，就足以确定这一段序列的单体型。例如，如果测出一个染色体的标签 SNPs 是 A - T - C，则这一序列是单体型 1。(转引自 The International HapMap Consortium. The International HapMap Project, Figure 1. *Nature*, 2003, in press)



(4) InDel (insertion-deletion)标记

插入缺失标记，指的是两种亲本中在全基因组中的差异，相对另一个亲本而言，其中一个亲本的基因组中有一定数量的核苷酸插入或缺失。根据基因组中插入缺失位点，设计一些扩增这些插入缺失位点的PCR 引物，这就是InDel 标记



刘耀光等用InDel标记引物

- L14-121 5'-CTTGACAATTGCAAGGCCAA3'
5'-GAGGTAGGCATGAGACATGC-3'
- J04-91 5'-GTGCGGATCGATCGGCAGCA-3'
5'-GCTCAGATCCGGCCAGATTTC-3'
- P24-83 5'-CAGAACAGCATTTCAGGCATC-3'
5'-ATCCTGATAATGTGTGGCGC-3'
- P24-93 5'-CATTGGTTAAAGGATGGTAC-3'
5'-TAGTACTGCTAGGACCATCC-3'

用InDel分子标记的水稻基因型检测





2、亲本的选配

亲本的选择直接影响到构建连锁图谱的难易程度及所建图谱的适用范围。对亲本的选择要考虑：

亲本间的DNA多态性：亲本间要有差异、DNA多态性非常丰富

亲本材料的纯度：影响试验结果，纯度不高的要自交进一步纯化

杂交后代的可育性等问题。

最后，选配亲本时还应对亲本及其 F_1 杂种进行细胞学鉴定。防止亲本间有染色体易位、缺失及 F_1 染色体异常等现象对作图的不利影响出现



3、分离群体类型的选择与构建作图群体

分离群体分为两类：暂时性分离群体 永久性分离群体

暂时性分离群体：分离单位是个体，经自交或近交后其遗传组成会发生变化，无法永久使用，这些群体包括 F_2 、 F_3 、 F_4 、BC和三交群体等

永久性分离群体：分离单位是株系，株系间基因组有差异，而株系内个体间的基因型是相同的是纯合的，是杂交不分离的，这类群体有

RILs (Recombinant inbred lines, **重组近交系**)

DH (Double haploid, 双单倍体) 群体等

它们可杂交或近交繁殖后代而不改变群体的遗传组成，可以永久使用。因此，构建DNA连锁图谱应根据具体情况选择不同类型的分离群体。

目前构建遗传连锁图谱主要应用 F_2 代群体、RILs和DH群体。



F₂群体是常用的作图群体，迄今大多数植物的DNA标记连锁图谱都是用F₂群体构建的。其主要优点是容易建立F₂群体，不足是对于显性纯合与显性杂合基因型无法识别，造成基因型信息简并现象的存在。

RILs群体是杂种后代经过多代自交而产生的一种作图群体，通常从F₂开始，采用单粒传的方法来建立。自交的作用是使纯合的基因型增加，杂合的基因型减少，因此，**RILs群体中每个株系都是纯合的**，因而RILs群体是一种可长期使用的永久性分离群体。又可以进行重复试验。它除了可用于构建分子标记连锁图外，还特别适用于数量性状基因座（QTL）的等位研究



DH群体是由植物的单倍体经过染色体加倍形成的二倍体即加倍单倍体或双单倍体（DH）而来。DH群体产生的途径很多，最常见的方法是通过花药培养，诱导产生单倍体植株，然后对染色体加倍产生DH植株。DH植株是纯合的，自交后产生纯系，这种纯系可以稳定繁殖，长期使用，是一种永久群体。DH群体的遗传结构直接反映了 F_1 配子中基因的分离和重组，因此**DH群体作图效率是最高的**。同时，DH群体也适合于QTL定位的研究。



4、群体大小的确定

遗传图谱的分辨率和精度与群体的大小有密切的关系。

群体越大，作图精度越高。但群体太大会增加实验工作量和费用。一般，构建DNA标记连锁图谱所需的群体远比构建形态性状特别是数量性状的遗传图谱要小，大部分已经发表的分子标记连锁图所用的分离群体一般都不足100个单株和家系。在实际工作中，构建分子标记骨架连锁图可基于大群体中的一个随机小群体（如150个单株或家系），当需要精细地研究某个连锁区域时，再针对性地在骨架连锁图的基础上扩大群体。



5、连锁图谱制作的统计学原理

(1) **两点测验** 对两个基因座位之间的连锁关系进行检测，称为两点检验。在进行连锁测验之前，必须了解各基因座位的等位基因分离是否符合孟德尔分离比例，这是连锁检验的前提。

在共显性条件下， F_2 群体中一个座位上的基因型分离比例为1: 2: 1，而 BC_1 和DH群体中分离比例均为1: 1；在显性条件下， F_2 群体分离比例为3: 1，而 BC_1 和DH群体中分离比例仍为1: 1。检验DNA标记的分离是否偏离孟德尔比例，一般采用 χ^2 检验。而对基因座位之间的连锁关系，则采用或然比检验的方法，即LOD值的大小来进行重组率的估计从而推断连锁是否存在。



(2) **多点测验** 在构建分子标记连锁图谱中，每条染色体都涉及到许多标记座位。要确定这些标记座位在染色体上的正确排列顺序及彼此间的遗传距离必须同时对多个座位进行联合分析，**利用多个座位间的共分离信息来确定它们的排列顺序，进行多点测验。多点测验通常也采用或然比检验法。**先对各种可能的基因排列顺序进行最大或然比估计，然后通过或然比检验确定出可能性最大的顺序。在一条染色体上，经过多次多点测验，就能确定出最佳的基因排列顺序，并估计出相邻基因间的遗传图距，从而构建出相应的连锁图。



遗传图谱构建中连锁与否的判断

人类遗传连锁图的构建是以系谱分析为基础，通过重组率的计算排列出不同多态位标的遗传距离而构建的。

在人类遗传学研究中，由于通常不知道父母的基因型和父母中标记基因的连锁相是相斥还是相引，因而无法简单地通过计算重组体出现的频率来进行连锁分析，而必须通过适当的统计模型来估算重组率，并采用**或然比**（似然比 likelihood ratio）**检验的方法来推断连锁是否存在**，



即比较假设两个座位间存在连锁 ($r < 0.5$) 的概率与假设没有连锁 ($r = 0.5$) 的概率。这两种概率之比可以用似然比统计量来表示:

即 $L(r) / L(0.5)$, 其中 $L()$ 为似然函数, 为方便起见, 常将 $L(r) / L(0.5)$ 取以10为底的对数, 称为LOD值(Lod score)或对数优势比 (logarithum of the odds)。为了确定两对基因之间是否存在连锁, 一般要求似然比大于1000: 1, 即 $LOD > 3$;而要否定连锁存在, 则要求似然比小于100: 1, 即 $LOD < 2$.

[返回](#)

[返回 \(93\)](#)



武汉大学

Wuhan University

在其他生物遗传图谱的构建中，似然比的概念也用来反映重组值的可靠程度或作为连锁是否真实存在的一种判断尺度。





1. 重组率与非重组率的计算

例：以下4个家系显示多态DNA标记与由一个显性等位基因D决定的性状的分离情况。

M和m分别代表分子标记的等位基因。

假设： 在每个家系中双杂合子亲本为MD/md

每个子代分为重组和非重组类型

在4个家系中子代只有家系（B）mD/md为重组

其余9人都是非重组

重组频率为： $r = 1/(1+9) = 0.1$

问题是数值太小以至于chi-square test 统计意义都不适合



我们如何确定这些家系中MD是否连锁？

第一步：是使用二项式展开式

计算出每个家系中子代的概率（重组）

仅仅考虑是重组子还是非重组子，

不考虑性别及其他的问题。

在家系（B）中，出现1个重组子和3个非重组子，

所以其出现的概率是：

$$(4! / 1!3!) r^1 (1-r)^3 = 4r(1-r)^3$$

(91)

每个家系中子代出现的概率都表示在方框图中第1行

第二步：对每个家系计算子代中重组子和非重组子分布

的概率，假设 $r=0.5$ （独立分配），基因位于

不同的染色体上或在相同的染色体上被分离

（非重组）。这些概率表示在方框图中第2行





2. 或然率的计算

$$\begin{aligned} \text{(A) Likelihood ratio} &= L(r) / L(0.5) \quad \underline{(91)} \\ &= (1-r) / 1/2 = 2(1-r) \end{aligned}$$

$$\text{Lod} = \log_{10}(2) + \log_{10}(1-r)$$

$$\begin{aligned} \text{(B) Likelihood ratio} &= L(r) / L(0.5) \\ &= 4r(1-r)^3 / 4(1/2)^4 = 16r(1-r)^3 \end{aligned}$$

$$\text{Lod} = \log_{10}(16) + \log_{10}(r) + 3\log_{10}(1-r)$$

.....





通常判定连锁关系是以LOD值大小为依据：

确定两对基因之间是存在**连锁**的，一般要求或然比大于1000：1，即**LOD>3**；而要**否定连锁**存在，则要求或然比小于100：1，即**LOD<2**。在其他生物遗传图谱的构建中，或然比的概念也用来反映重组值的可靠程度或作为连锁是否真实存在的一种判断尺度。

LOD值为0，意味着连锁假设与不连锁假设的可能性是相等的；LOD值为正值，有利于连锁；LOD值为负值，表示有一定重组率的连锁。显著性的域值是 +3和-2 。**LOD值=+3**时，**连锁的概率为95%**。如果两个基因座都在X染色体上，则LOD = 2.5，就足以说明两者是连锁的。



6、DNA标记分离数据的处理与 构建标记的遗传图谱

从分离群体中收集分子标记的分离数据，获得不同个体的DNA多态性信息，是进行连锁分析的第一步。

DNA标记基因型的表现——电泳带型——电泳带型数字化（是DNA标记分离数据进行数学处理的关键）。

必须区别所有可能的类型和情况，并赋予相应的数字或符号。例如，我们用RFLP标记构建遗传图谱，假设全部试验共100个 F_2 单株，检验了90个RFLP标记，经数字化处理，可得到一个由90（行） \times 100（列）的、由简单数字组成的RFLP数据矩阵。获得有关数据矩阵后，选择适合的构建DNA标记图谱的计算机软件对其进行分析和处理。



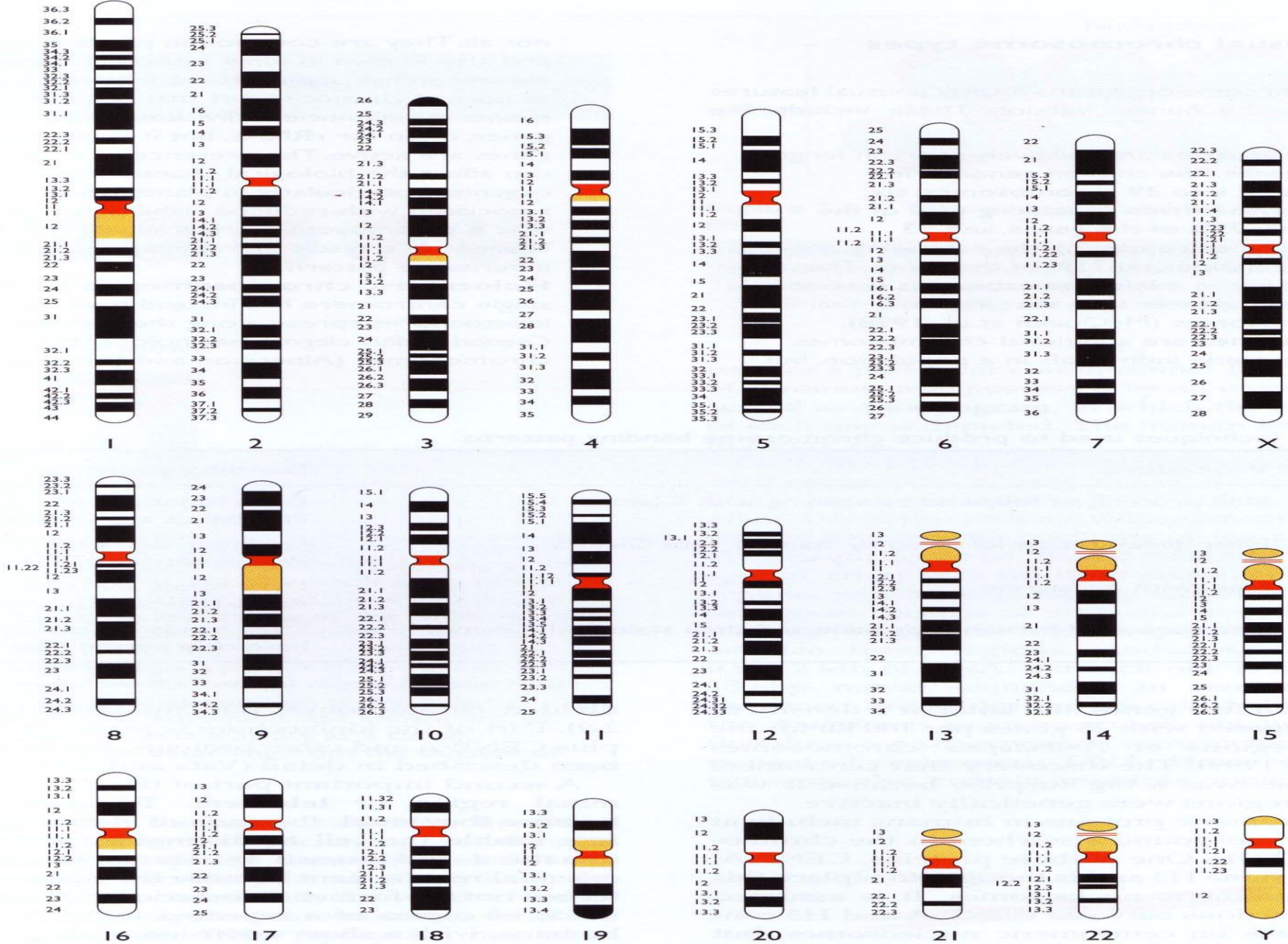
举例：中国莲SSR图谱构建

对以上SSR条带矩阵数据库用JoinMap program version 3.0软件进行莲基因组分子遗传图谱的构建。在 $LOD \geq 4.0$ 的情况下进行连锁群的估计，以 $LOD=2.0$ 确定连锁群内标记的顺序，以计算结果提供的最佳顺序排列。SNP标记作为备选标记用于作图，以增加图谱的标记密度和精确性。采用Kosambi函数将重组值转换成遗传图距(cM)



三、物理图的构建方法

1、**细胞遗传学图谱** (cytogenetic map)：是将基因或DNA片段直观定位于染色体上的物理图谱，因此也称为染色体图谱(chromosome map)。它是把基因或其他被分离出的DNA片段定位在它所在的染色体区域，并且粗略地测出它们之间相距的碱基长度。其图谱的制作主要采用原位杂交技术，将目标基因或特定的DNA片段定位到特定的染色体区带上



KEY

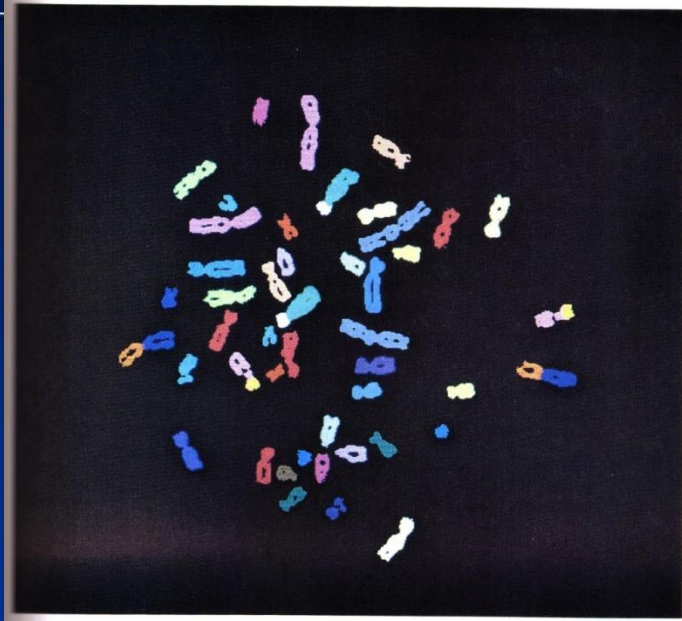
Centromere
 rDNA
 Constitutive heterochromatin



目前用得最多的是**荧光原位杂交 (Fluorescence in situ hybridization, FISH)** 技术。

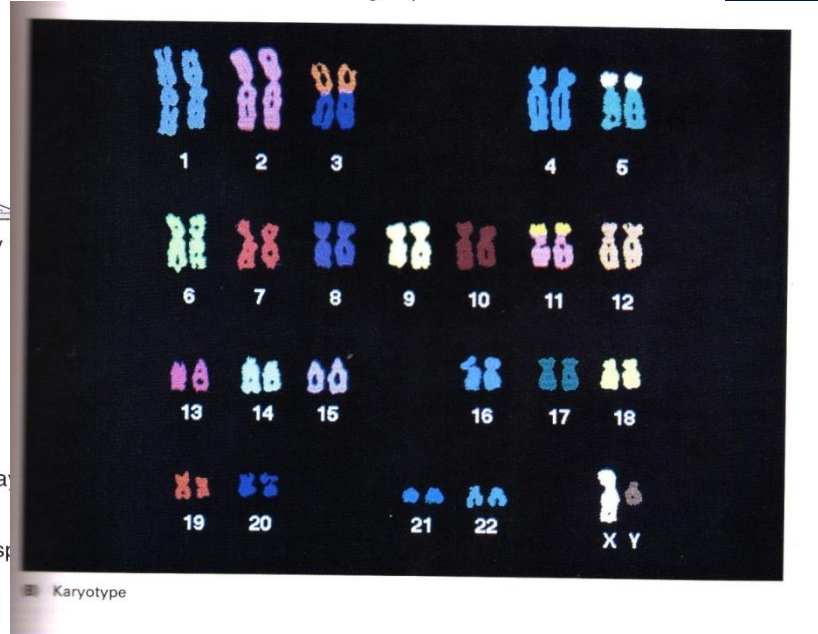
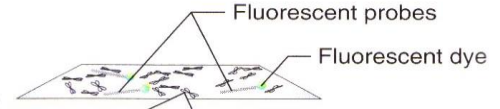
近年发展起来的**染色体DNA Fiber—FISH** 技术，可以将几组探针带上不同的荧光标记，同时与染色体杂交，利用不同的抗体进行检测，一次可以进行多个位点的杂交，并且有较好的分辨率。

(a)



Metaphase spread

Figure 7.13
 (A) Metaphase spread and (B) karyotype in which human chromosomes have been "painted" in 27 different colors according to their hybridization with fluorescent probes specific to individual chromosomes or chromosome arms. [Courtesy of M. R. Speicher and D. C. Ward. See M. R. Speicher, S. G. Ballard, and D. C. Ward. 1996. *Nature Genetics* 12: 368.]

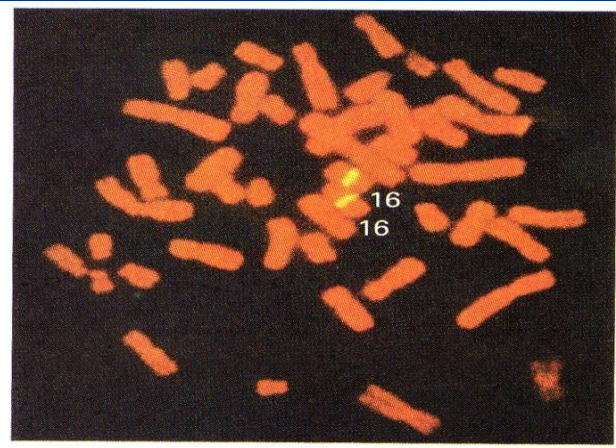
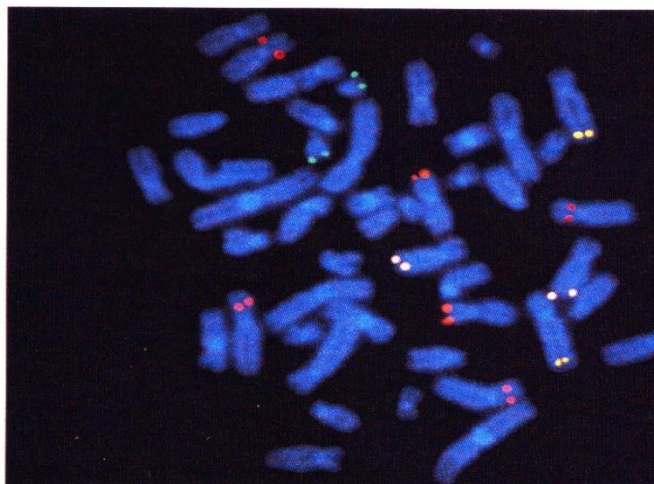
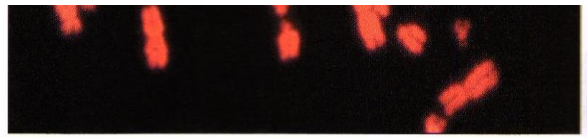


Karyotype

An
iza
we

dangerous short UV rays, allows needed long UV rays to pass through)

4. Expose to ultraviolet (UV) light.
 Take picture of fluorescent chromosomes.



▲ FIGURE 9-6 Use of simple-sequence DNA as chromosomal marker. Human metaphase chromosomes stained with a



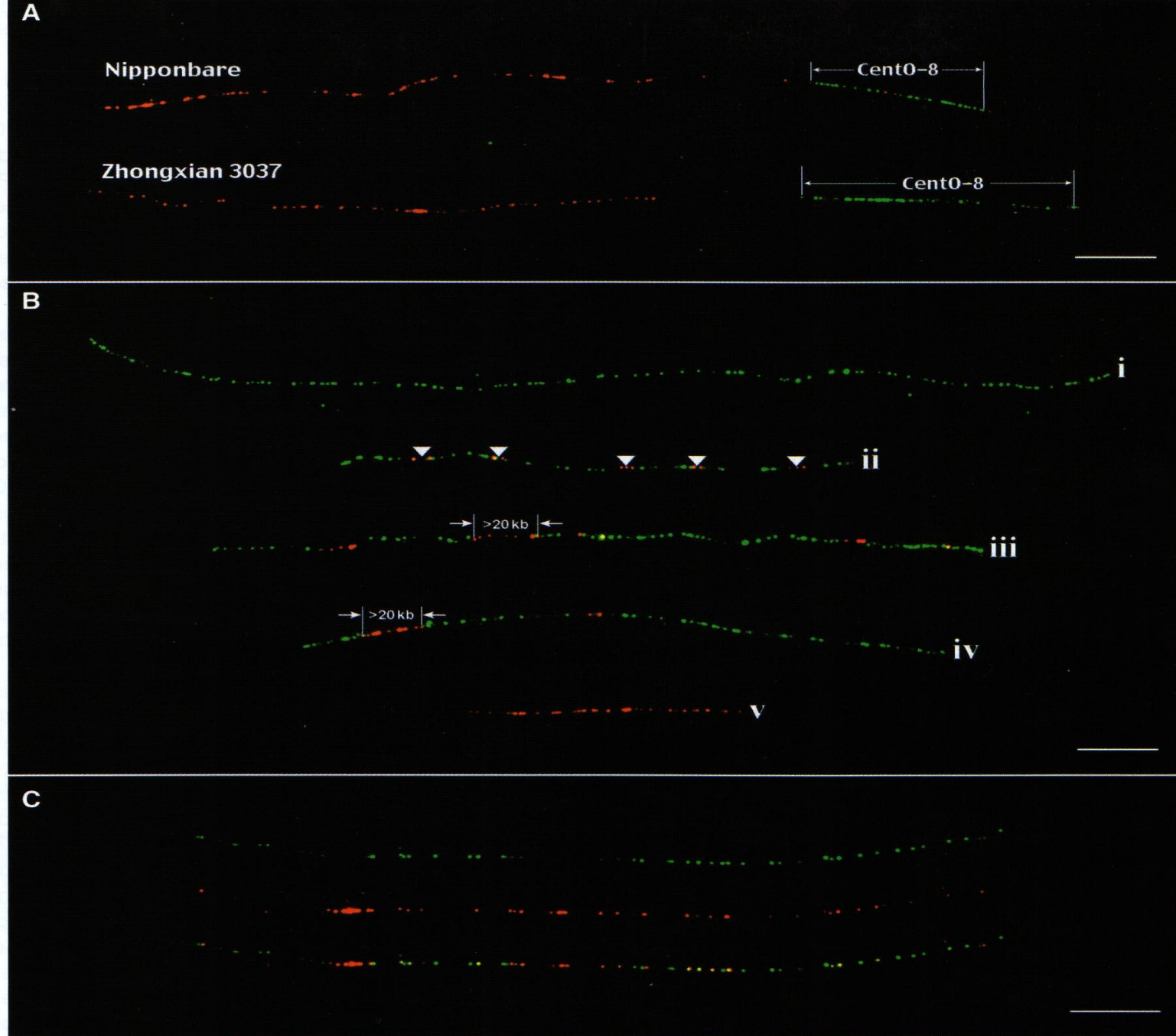
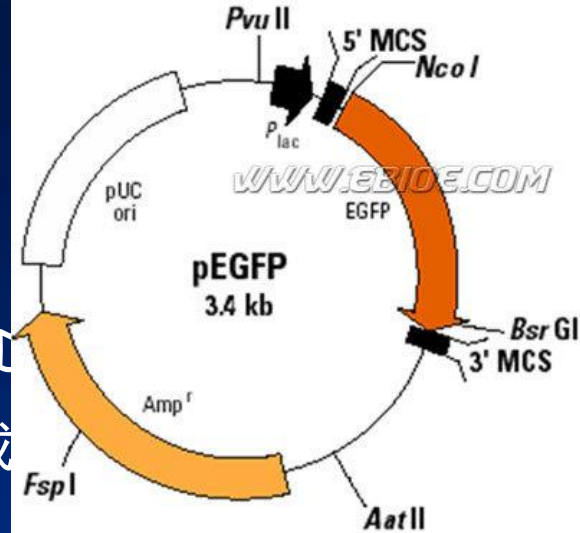


Figure 4. Fiber-FISH Analysis of Rice Centromeric DNA.

(A) FISH using PCC1 (red) and PCC2 (green) probes for the CentO1 and CentO2 (PCC1 and PCC2) FISH signals, respectively, in Nipponbare and Zhongxian 3037. Scale bar, 50 kb. (B) Fiber-FISH analysis of rice centromeric DNA. The tracks (i-v) show the same DNA molecule as in (A). The scale bar indicates a distance of >20 kb. (C) Fiber-FISH analysis of rice centromeric DNA. The tracks show the same DNA molecule as in (A). The scale bar indicates a distance of >20 kb.

(2) 限制酶切图谱

DNA限制性内切酶酶切图谱是一种重要的DNA分析工具。它是由一系列位置确定的多种限制性内切酶酶切位点组成，用于分析DNA片段的长度和位置。



其主要的步骤是：制备克隆DNA；克隆DNA指纹分析，包括酶切、电泳分离、杂交标记或图象处理；计算机对克隆的排列和DNA片段的排序；填空隙，包括分离新克隆，PCR验证等。

构建DNA限制性内切酶图谱有许多方法。通常结合使用多种限制性内切酶，通过综合分析多种酶单切及不同组合的多种酶同时酶切所得到的限制性片段大小来确定各种酶的酶切位点及其相对位置。

(3) 叠连群图谱

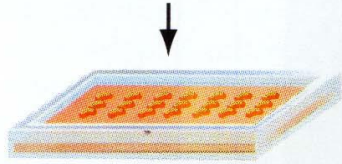
一组相互两两头尾拼接的可装配成长片段的DNA序列克隆群称为叠连群 (Contig)。

其具体步骤是把染色体切割成小片段后，克隆并排序，得到由排好序的插入DNA片段克隆组成的叠连群。克隆叠连群有多方面的重要应用，是人类基因组计划物理图谱的核心部分。使用这些叠连群图谱，可以重新确定克隆片段或其他DNA探针在基因组中的位置，一旦构建好叠连群图谱，则整个基因组就可以以克隆形式获得，从而可在碱基序列水平上用这些克隆来分析基因组的任何区域。

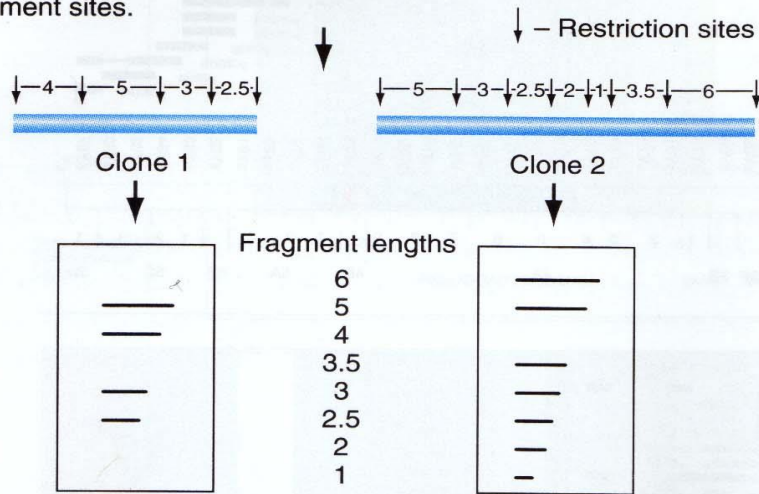
Bottom-up approach



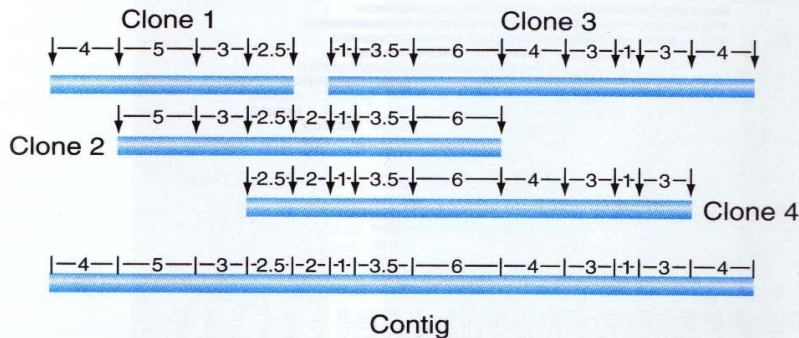
- (a) Digest chromosomes with restriction enzyme. Clone fragments into a genomic cosmid library.



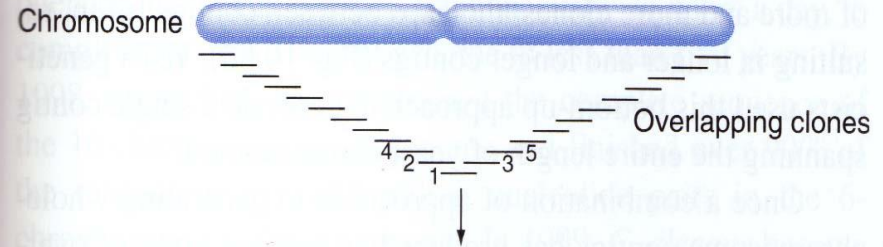
- (b) Fingerprint each cosmid for the pattern of restriction fragment sites.



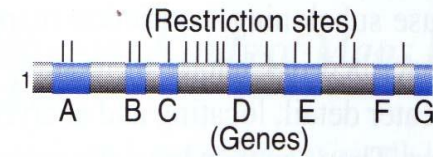
- (c) Feed fingerprint data into computer programmed to look for overlaps.
(d) Arrange overlapping cosmid inserts into 24 chromosomal contigs.



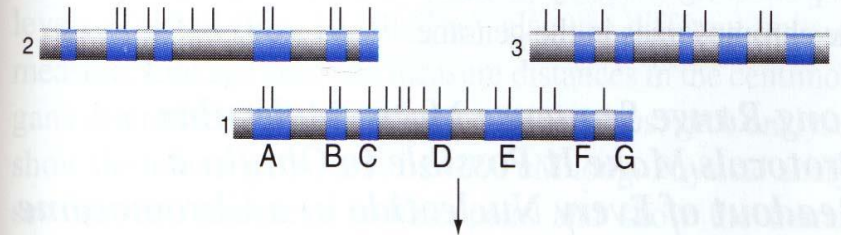
- (a) Identify an ordered series of overlapping genomic clones.



- (b) Analyze each clone for restriction sites and gene locations.



- (c) Create maps of overlapping genomic clones.



- (d) Combine information into a single continuous physical map that spans the length of the chromosome.



Figure 10.5 Building a whole-chromosome physical map.

(a) To produce a whole-chromosome physical map, you first order a



武汉大学

Wuhan University

DNA 测序技术及其进展

2016-6-6





一、关于DNA测序的回顾

1950年 已知 DNA是一个杂聚物 (heteropolymer)

1953年 Watson and Crick的DNA结构模型

暗示遗传信息的贮存机制——核苷酸的顺序

1959年 Arthur Kornberg DNA polymerase

1977年 Sanger(F)-Maxam(G), 酶法-化学降解法, DNA序列测

2000年 基因组测序, 第一份人类基因组草图

2003年 完成第一份人类全基因组序列图谱

2005年 刚刚开始进行新一代测序技术开发

2007年 首位“个人版”基因组图谱 —— Watson的“生命天书”

2008年 DNA sequencing -- 方法年



DNA 测序技术在现代生命科学各研究领域具有重要的作用，是促进基因组学、分子生物学等生命科学迅猛发展的关键技术之一。

以前用于测序的主要技术有：

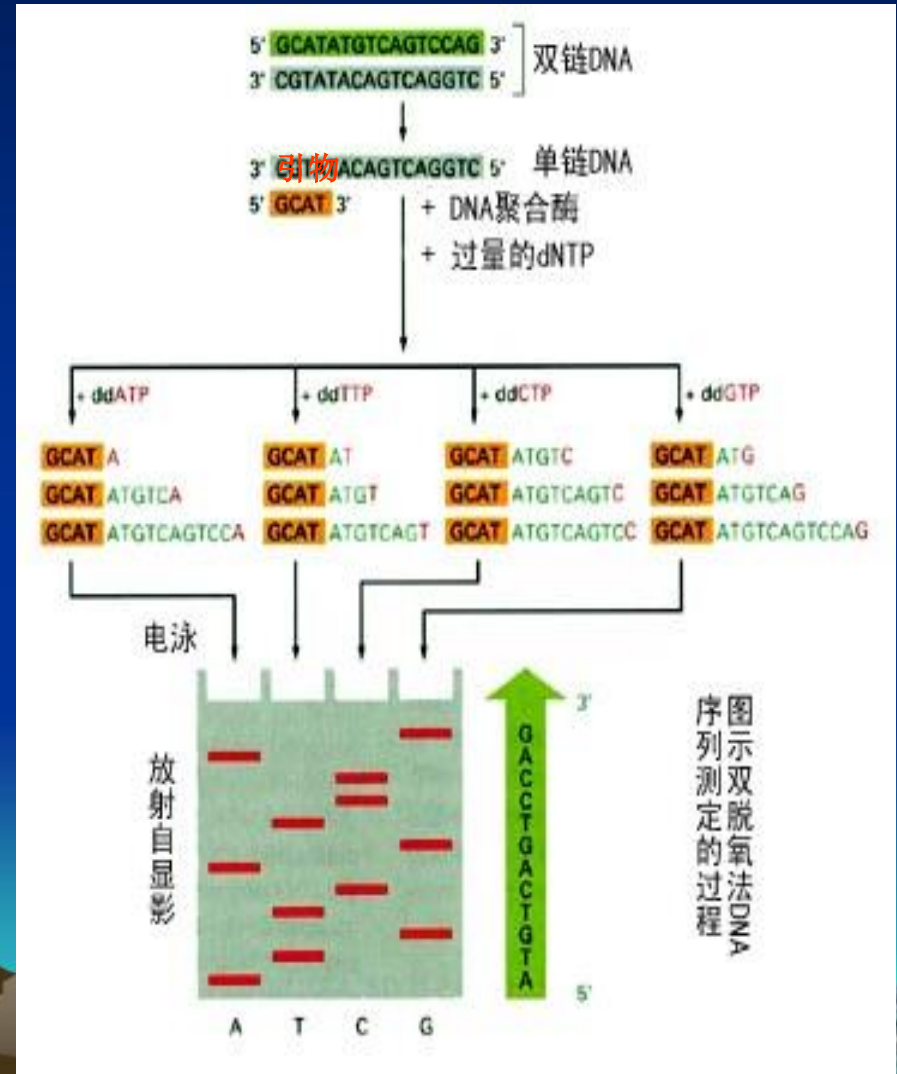
Sanger等（1977）发明的双脱氧链末端终止法

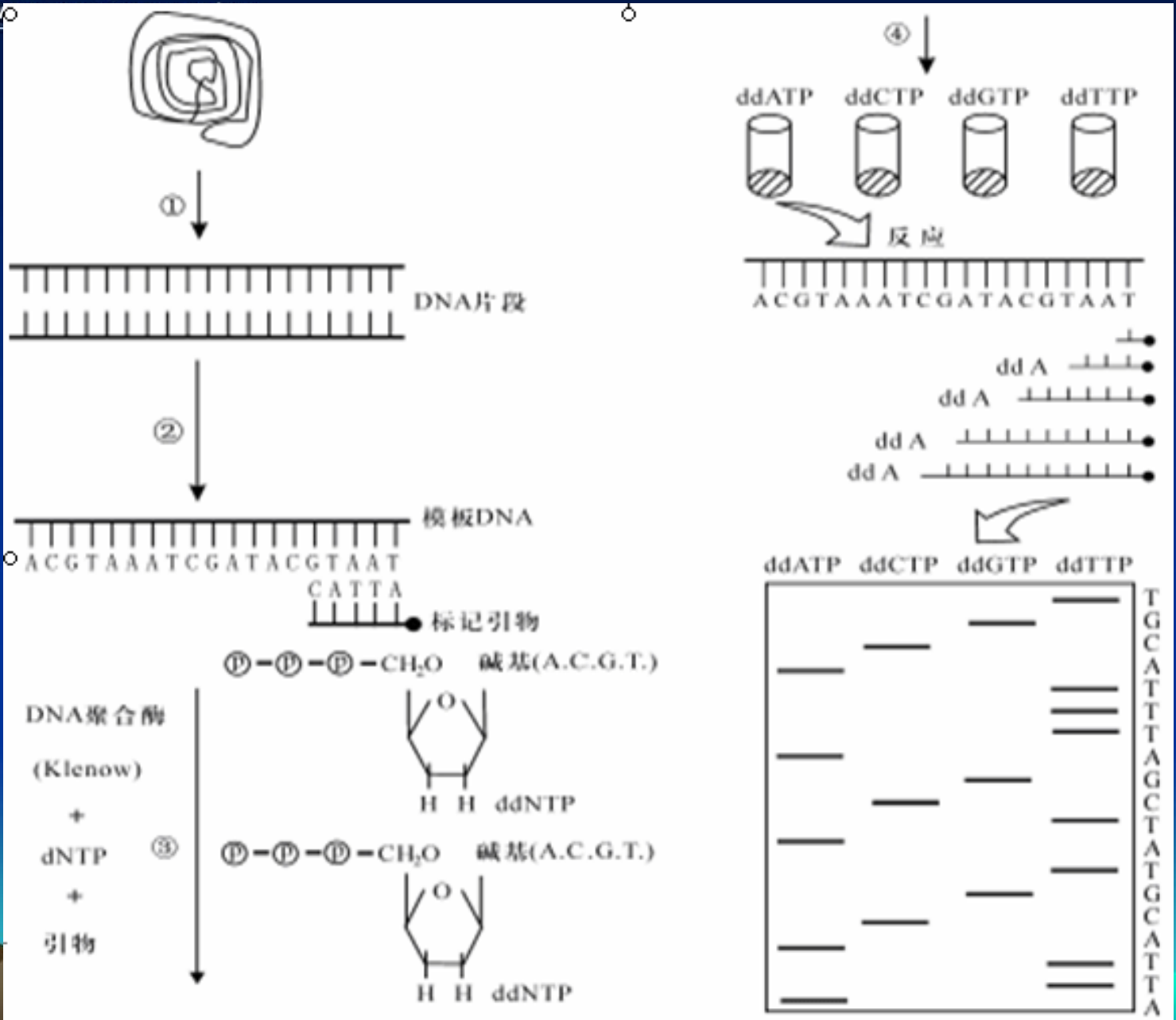
Maxam和 Gilbert（1977）发明的化学降解法

两者原理上差异很大，但都是根据核苷酸在某一固定的点开始，随机在某一个特定的碱基处终止，产生A、T、C、G四组不同长度的一系列核苷酸，然后在尿素变性的PAGE胶上电泳进行检测，从而获得DNA序列。

Sanger 等的双脱氧链末端终止法进行测序

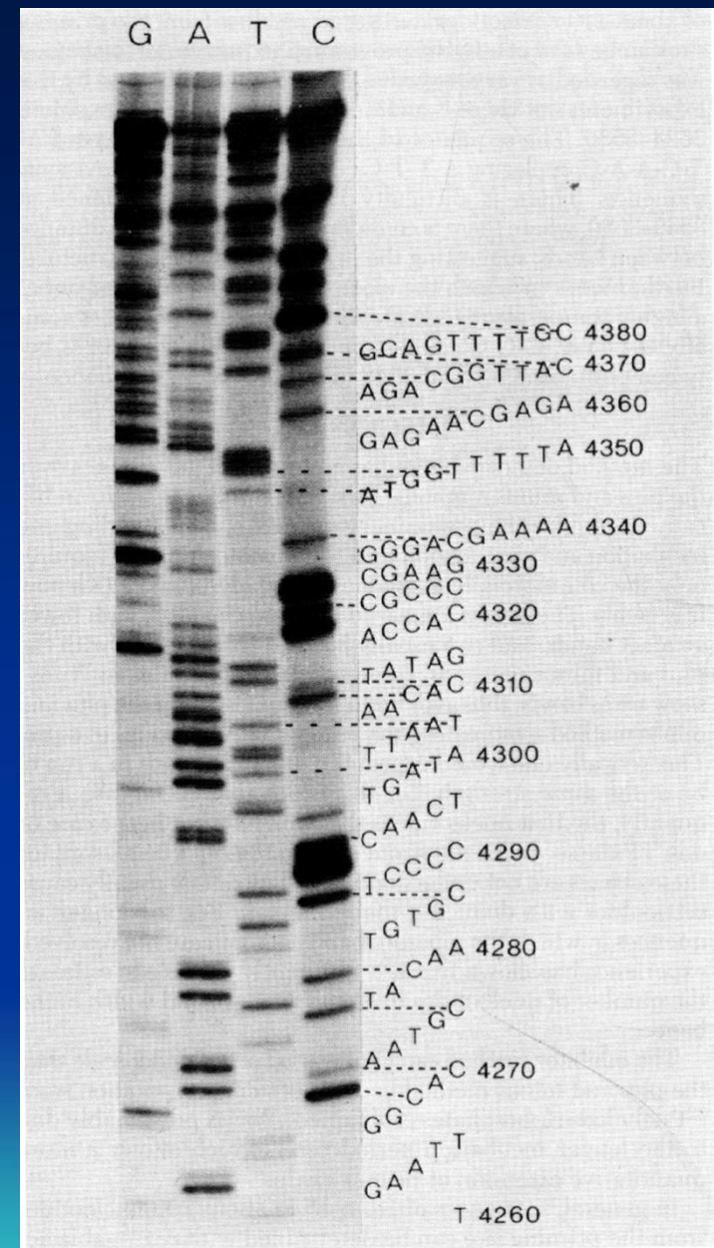
该技术引入了聚合酶和测序胶。
DNA聚合酶不能起始DNA链的合成，而是在退火于“模板”DNA的引物的3'端上进行延伸。双脱氧测序法利用了DNA聚合酶能以双脱氧核糖核苷酸为底物的特性。当ddNTP被掺入到引物延伸的引物的3'端时，由于链上3'羟基的缺如，链的延伸就终止于G、A、T或C。在4个测序反应中，每个反应只需各加入4种可能的ddNTP中的其中一种，就将产生如图所示的4个序列阶梯。



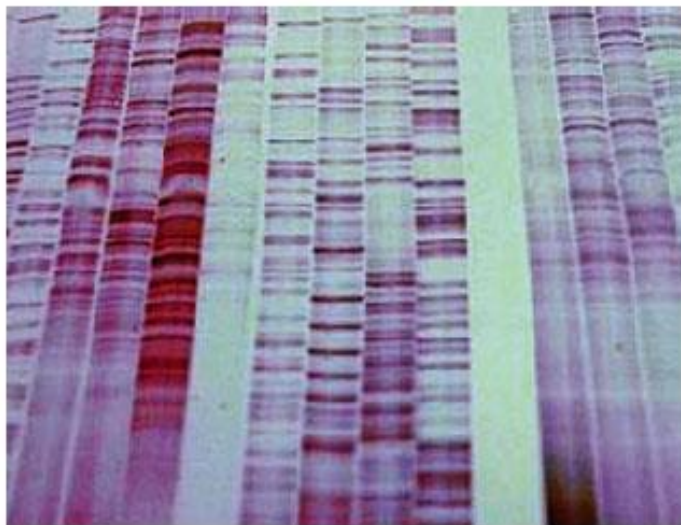
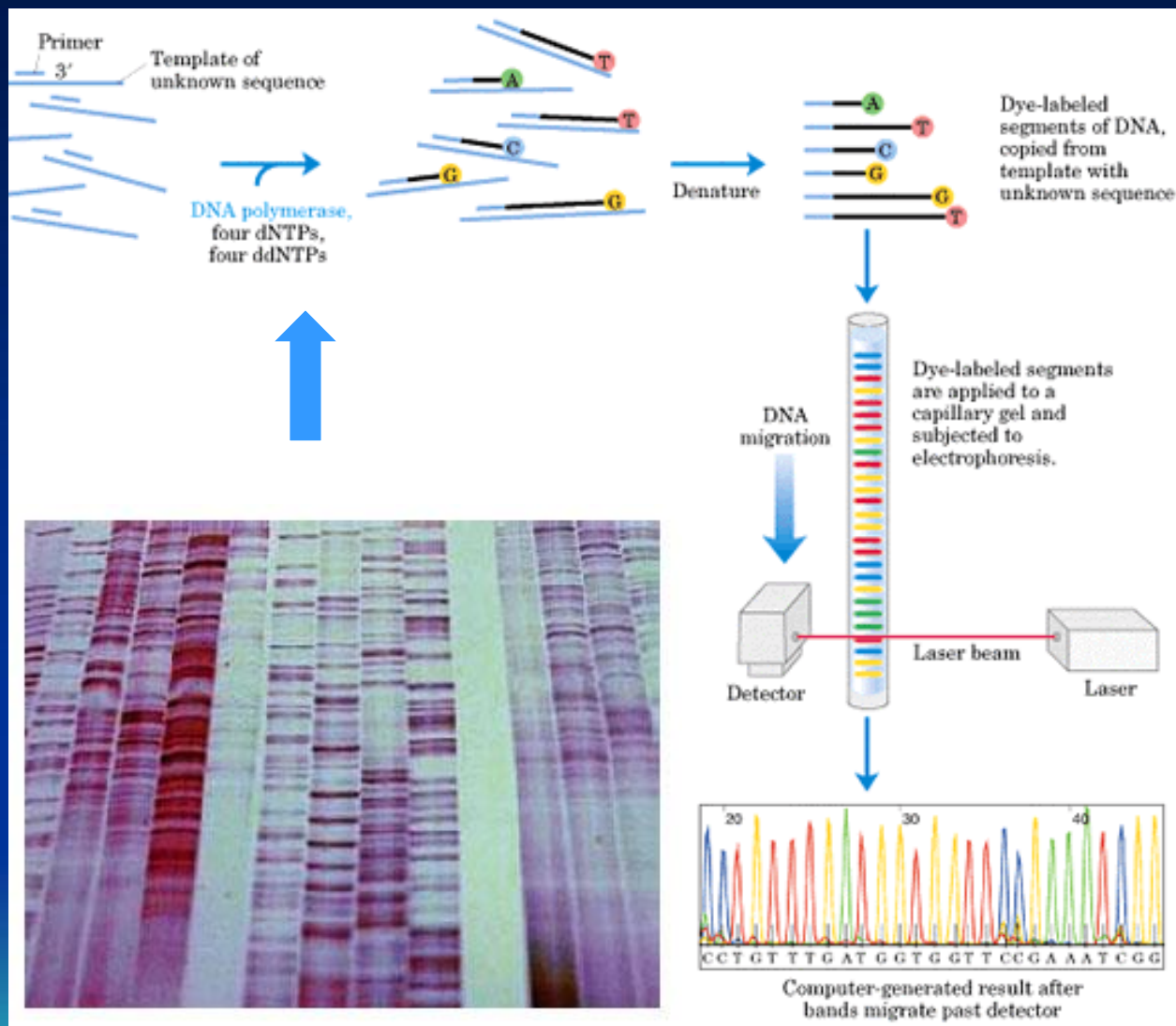


First generation DNA sequencing

Frederick Sanger's enzymatic dideoxy DNA sequencing technique based on the chain-terminating dideoxynucleotide analogues (Sanger et al. 1977)



聚合酶测序法
在经过改良后，成
为迄今为止应用最
为广泛的测序技术—
毛细管电泳测序仪





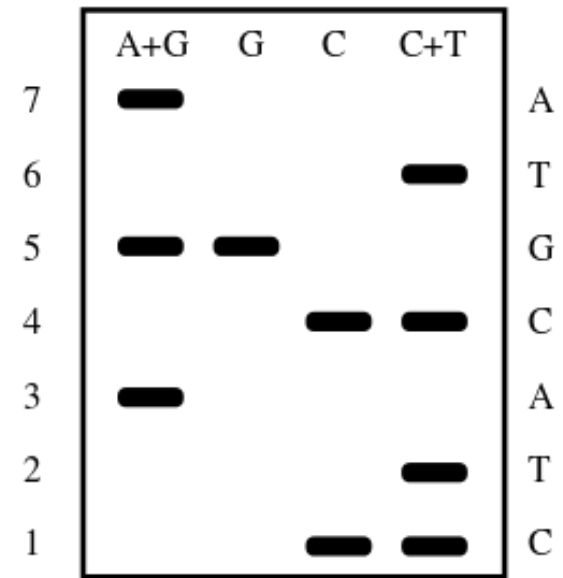
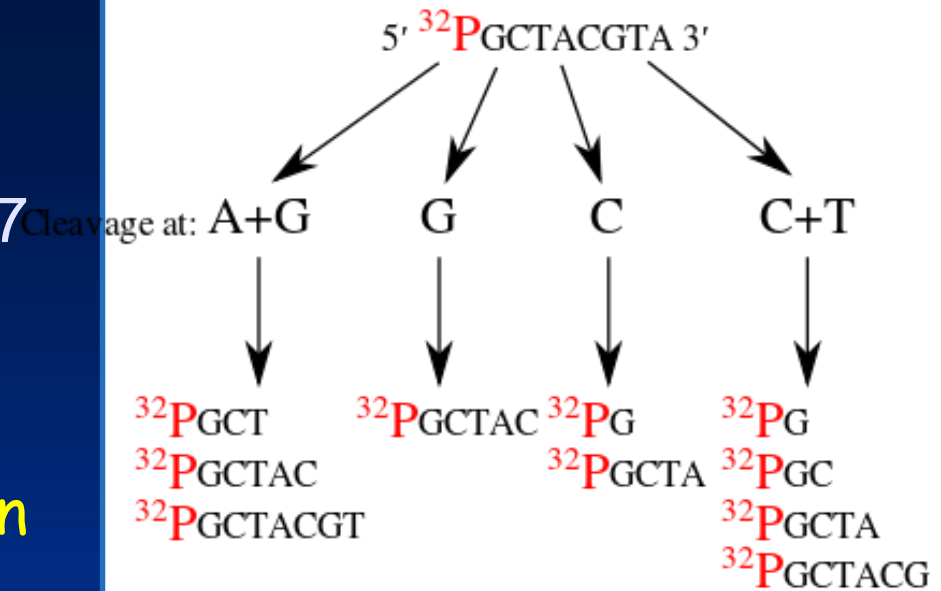
Maxam-Gilbert测序法—化学降解法

化学降解法需要双链的DNA片段。**不需要引物**，因为这种方法不是合成DNA片段，而是降解已有的DNA分子。化学降解法操作是要小心，因为降解DNA分子的化学药剂是剧毒的。

首先要标记双链DNA的5'端，然后加入二甲基亚砷并加热到90°C，使双链DNA分子变成单链。通过凝胶电泳分离两条链，其中一条链含有更多的嘌呤成为重链，另一条链为轻链。纯化两条链，分成四份，分别利用不同的降解试剂进行降解（哌啶，氮杂环己烷），产生了一系列的降解片段。可以通过电泳分离显示出来。

Maxam–Gilbert sequencing 1977

Allan Maxam and Walter Gilbert's chemical degradation DNA sequencing technique in which terminally labeled DNA fragments were chemically cleaved at specific bases and separated by gel electrophoresis

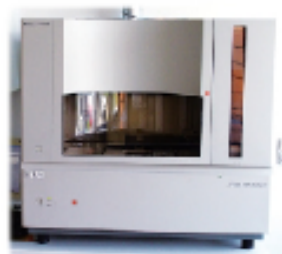


Sequencing Gel

“old” way of genome sequencing

Cloning and clone handling are very labor intensive

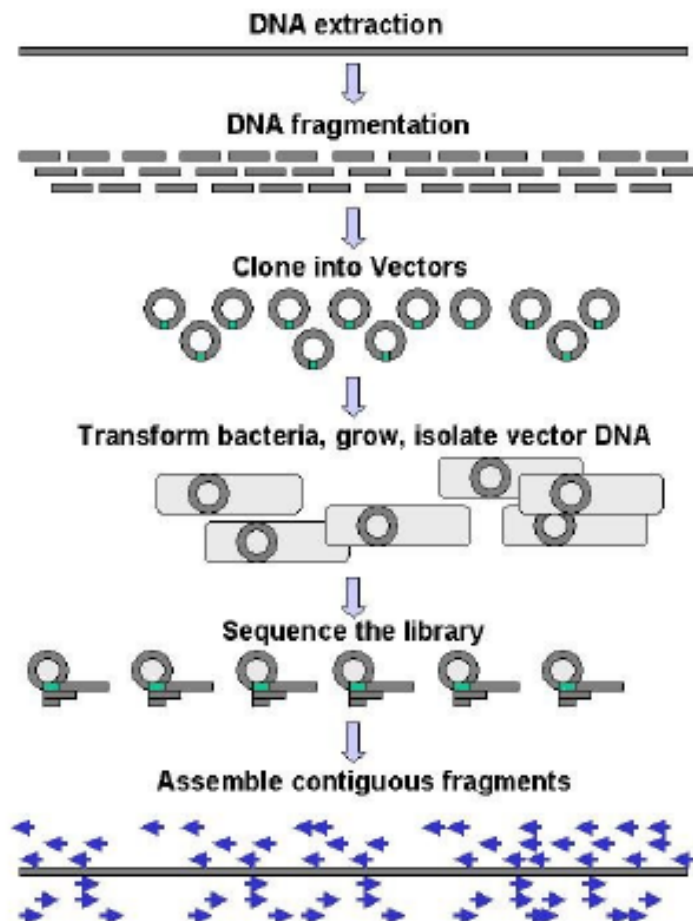
Throughput of capillary sequencing machines is limited



ABI 3730XL
(Applied Biosystems/Sanger)

up to 1,100 bases/read
96 reads/run
approx. 1 MB/day end machine

First choice for finishing projects; full length cDNA sequencing; single sample sequencing.



Method of the Year

There are events of the year, persons of the year, images of the year.... We could not resist: why not a Method of the Year?

Methods are a driving force of scientific progress. We think they should be celebrated as such. So the editors set out, a couple of months ago, to select the most notable method of 2007—not a method just off the inventor's bench but rather one that came into its own in 2007 and had a wide-ranging impact. But our discussion was quick: we soon had a clear winner in next-generation sequencing.

It is actually not one single method but several, developed in parallel. The first two next-generation sequencing methods made their official entry on the scene in 2005, with publications by the groups of Jonathan Rothberg and George Church, and two related platforms became available within two years. A feature on page 11 recounts the development of these methods and some of the events in 2007 that contributed to firmly establish them in the community.



二、新一代DNA测序技术平台

代表国际先进水平的有：

- 1、Roche (454) *GS FLX sequencer*
- 2、Illumina *Solexa* (Illumina genome analyzer)
- 3、Applied Biosystems (Applied Biosystems Genetic Analyzer) / ABI *SOLiD* (SOLiD sequencer)

4、第三代DNA测序技术平台

Heliscope、

SMRT (Single Molecule Real-time Technology)

Nanopore Sequencing Platform

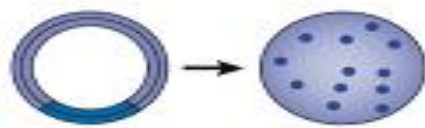
等多种测序仪器组成的基因组分析实验技术平台



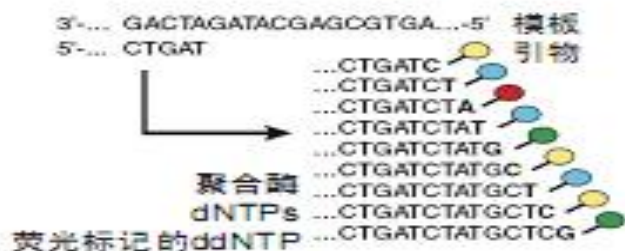
a DNA分子被随机切割成小片段



体内克隆扩增



循环测序

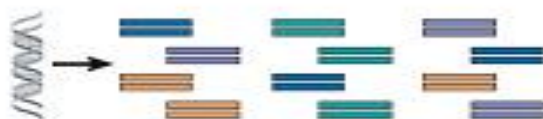


电泳

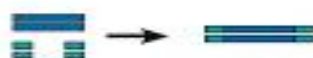
(每次电泳读取一个碱基信息)



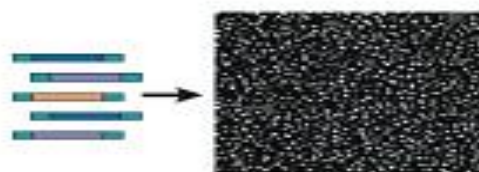
b DNA分子被随机切割成小片段



体外接头连接



制成polony芯片



芯片循环测序

(每个芯片可以获得长度超过10⁶ bp的序列)

芯片循环测序1 芯片循环测序2 芯片循环测序3



第一个碱基是什么?



第二个碱基是什么?



第三个碱基是什么?

图1 传统的Sanger测序法及新一代DNA测序技术工作流程图。

(a) 高通量鸟枪Sanger测序法。首先基因组DNA被随机切割成小片段分子，接着众多小片段DNA被克隆入质粒载体，随后转化到大肠杆菌中，最后培养大肠杆菌提取质粒，进行测序。每一个测序反应都在只有几微升的反应体系中完成。测序后获得一系列长短不一的末端标记有荧光的片段，最后通过对每一个延伸反应产物末端荧光颜色进行识别来读取DNA序列。(b) 鸟枪循环芯片测序法。首先将基因组DNA随机分割成小片段DNA分子，然后在这些小片段DNA分子的末端连接上普通的接头，最后用这些小片段DNA分子制成polony芯片，每一个polony中都含有一个小片段DNA分子的许多拷贝，许多这样的polony集合在一起就形成了polony芯片，这样一次测序反应就可以同时对众多的polony进行测序，然后与sanger法中一样，通过对每一个延伸反应产物末端荧光颜色进行识别来读取DNA序列，重复上述步骤就能获得完整的序列。



Overview



**Applied Biosystems
ABI 3730XL
1 Mb / day**



**Roche / 454
Genome Sequencer FLX
100 Mb / run**



**Illumina / Solexa
Genetic Analyzer
2000 Mb / run**



**Applied Biosystems
SOLiD
3000 Mb / run**



1、Roche (454) GS FLX sequencer

1998年《Science》介绍了该测序的基本流程

(Ronaghi et al., 1998)

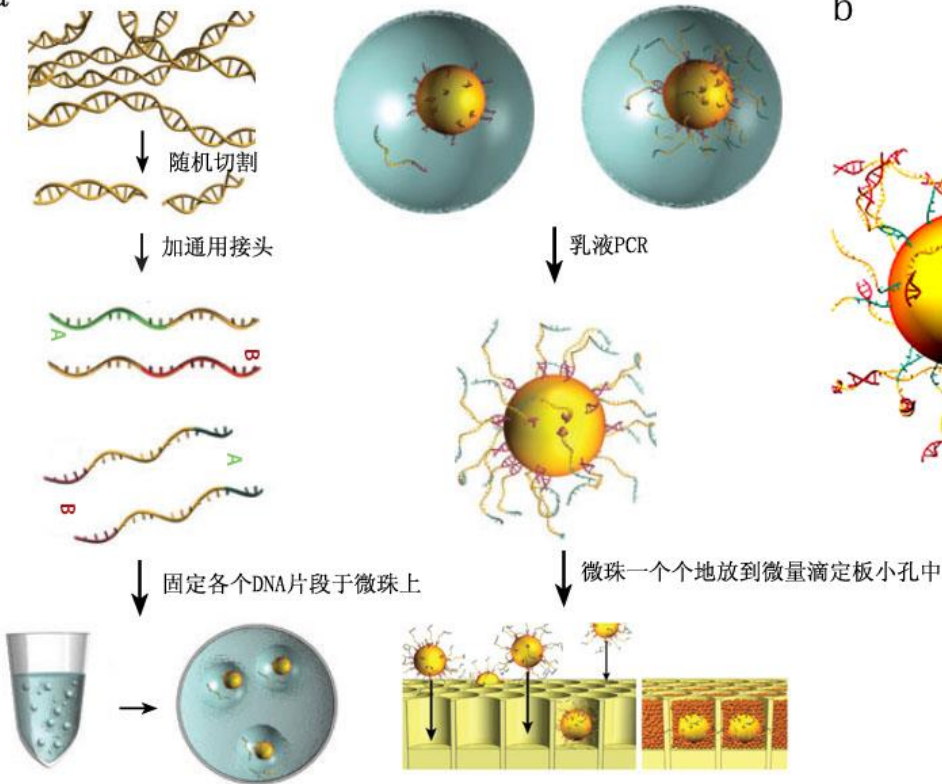
2005年《Nature》对该测序仪的应用做了报道

(Margulies et al., 2005)

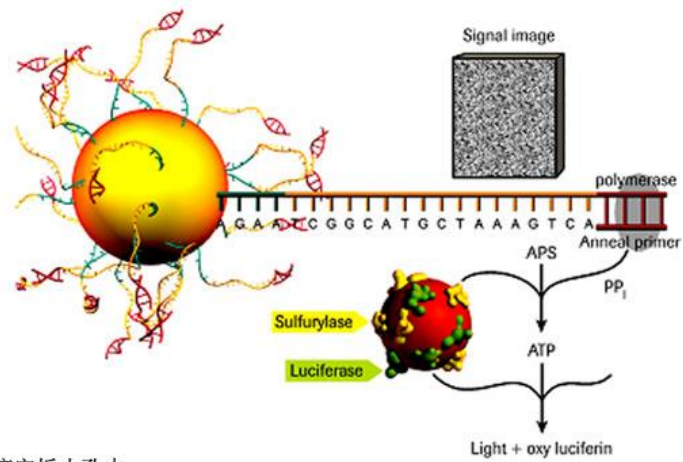
454的技术平台是将一条DNA片段与一个微珠结合，利用PCR技术，每个微珠分别获得几百万DNA片段的拷贝。在微珠进入微滴度板后，每个DNA碱基（A、C、T、G）在微板上被分别洗出，并边合成边测序。CCD摄像机记录了在焦磷酸测序反应中10万个微球同一时间添加的每一个碱基。



a



b



454测序技术原理示意图

- a. 文库制备、模板制备和测序准备过程
- b. 焦磷酸测序基本原理
- 该技术平台最主要的优点是测序读长较长，目前可以准确进行400个以上的碱基序列分析



焦磷酸测序技术的原理

1987年Nyren 等发展一种新型的DNA测序技术:

焦磷酸测序技术(**Pyrosequencing**)

是由四种酶催化的同一反应体系中的酶级联反应

四种酶分别为:

DNA 聚合酶 (DNA polymerase)

ATP 硫酸化酶(ATP sulfurylase)

荧光素酶(luciferase)

双磷酸酶(apyrase)

反应底物为:

5'-磷酸硫酸(adenosine 5'-phosphosulfate ,APS)

荧光素(luciferin)

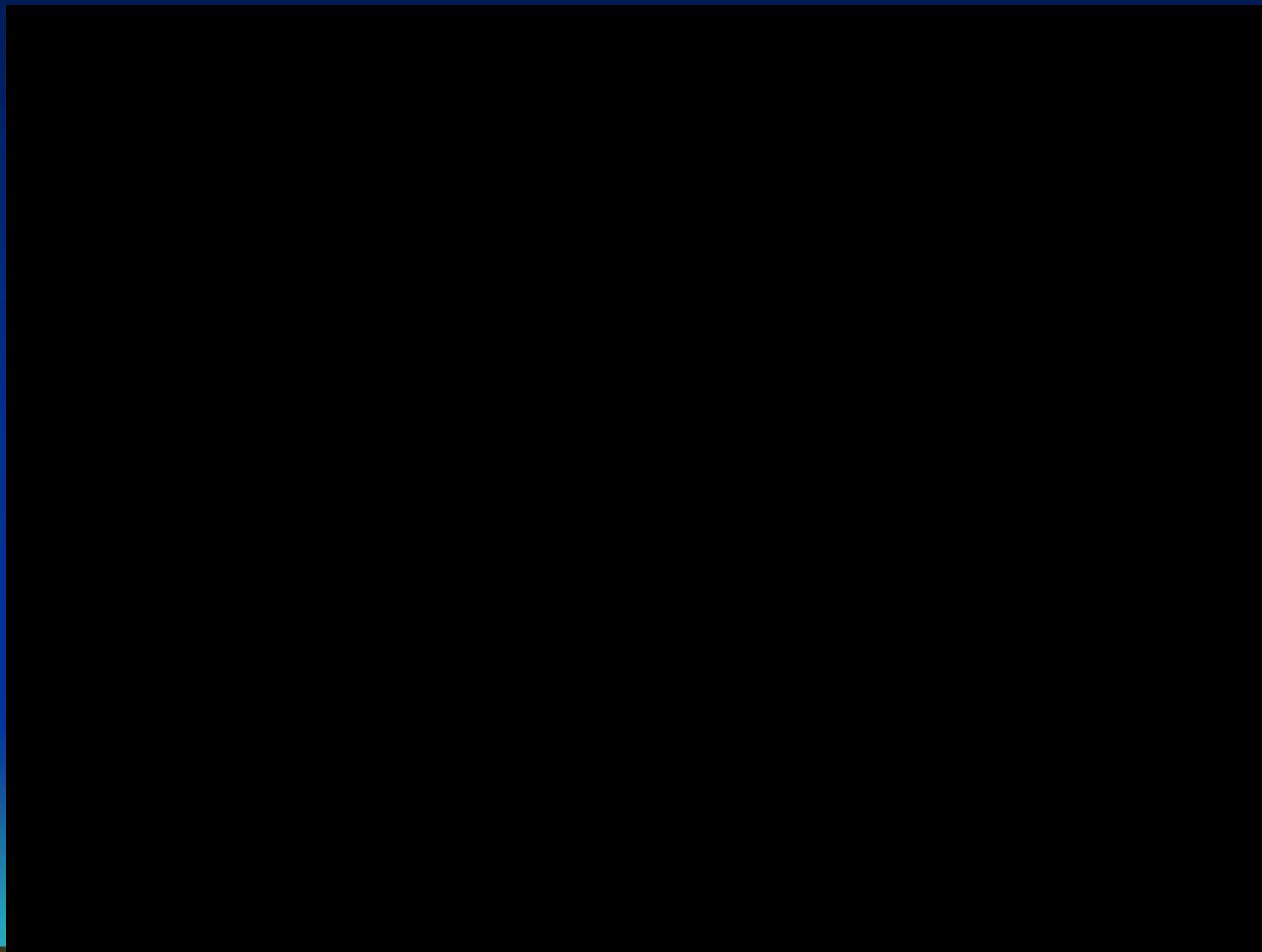
反应体系还包括待测序DNA单链和测序引物。



武汉大学

Wuhan University

DNA library preparation



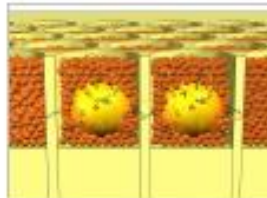
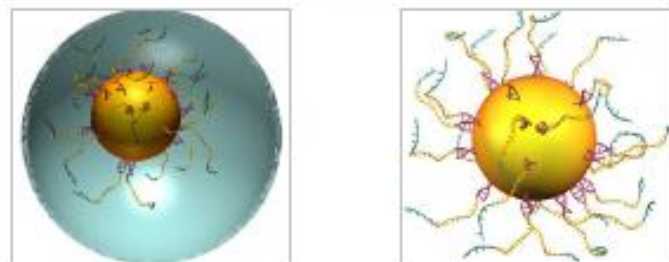
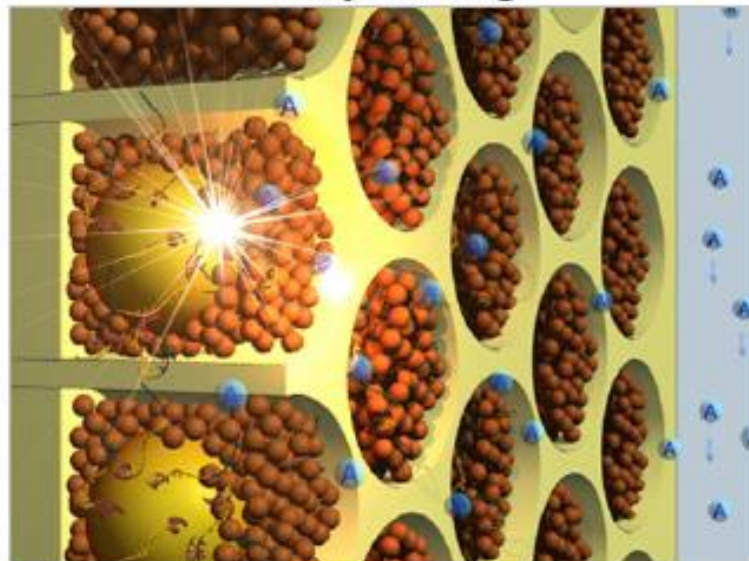
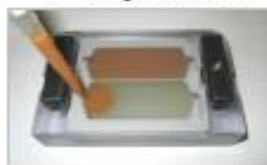
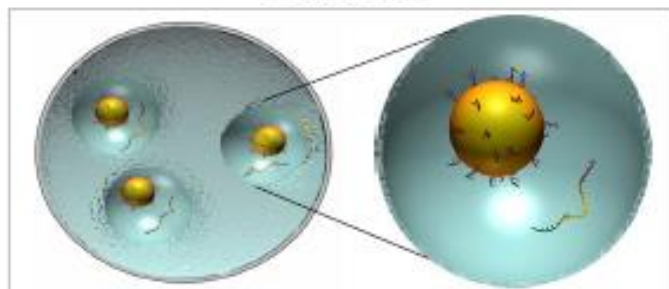
Sequencing Workflow

Short hands-on time, quick turn-around to result

emPCR

Prep Run

Sequencing



Hands-on Time 4 h
Total Time 8 h

2 h
2 h

0 h
7 h

Total Hands on Time	6 h
Total Time	17 h



emPCR



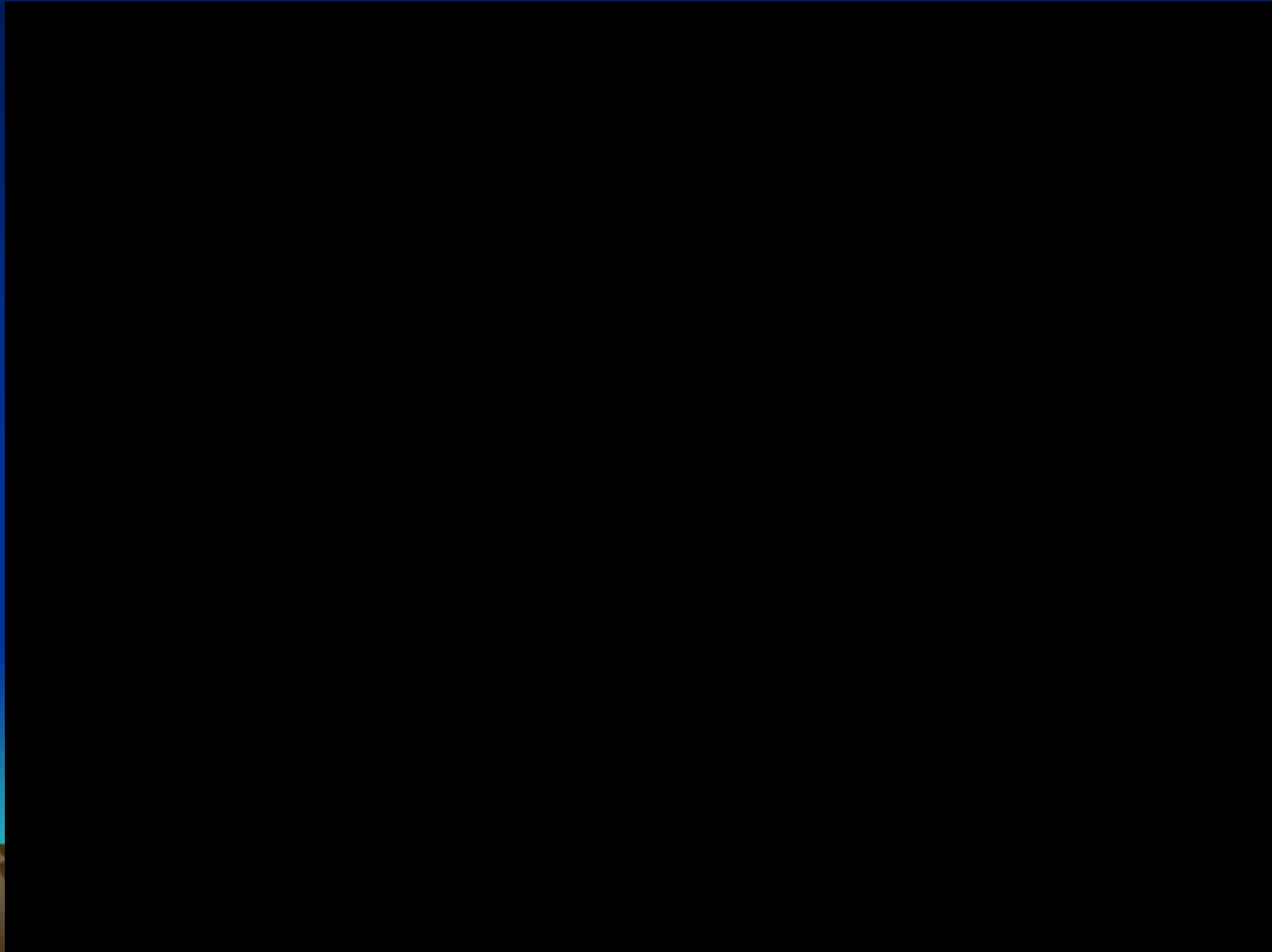


Sequencing



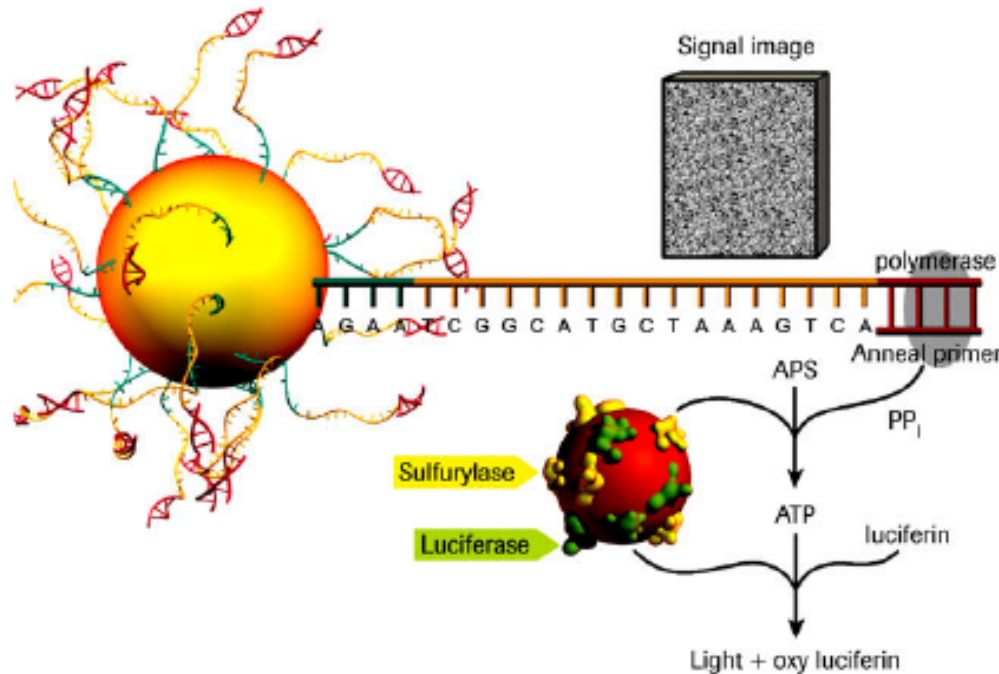


Data Analysis



Sequencing Workflow

Sequencing by Synthesis



- Bases (TACG) are flowed sequentially and always in the same order (100 times for a large GS FLX run) across the PicoTiterPlate™ device during a sequencing run.
- A nucleotide complementary to the template strand generates a light signal upon incorporation.
- The light signal is recorded by the CCD camera.
- The signal strength is proportional to the number of nucleotides incorporated.





Summary





2、Illumina **Solexa** (Illumina genome analyzer)

Solexa技术最早由两位剑桥大学的化学家创立，利用专利核心技术“DNA簇”和“可逆性末端终结”，达成自动化样本制备及基因组数百万个碱基大规模平行测序。

Illumina公司于2007年收购 Solexa，并推出成熟商业产品 *Genome Analyzer* (**基因组分析系统**)。 *Genome Analyzer* 为新一代革新性技术分子生物学综合技术平台，具有高准确性，高通量，高灵敏度，和低运行成本等突出优势。可以同时完成传统基因组学研究（测序和注释）以及功能基因组学（基因表达及调控，基因功能，蛋白/核酸相互作用）研究。



Solexa技术的基本原理

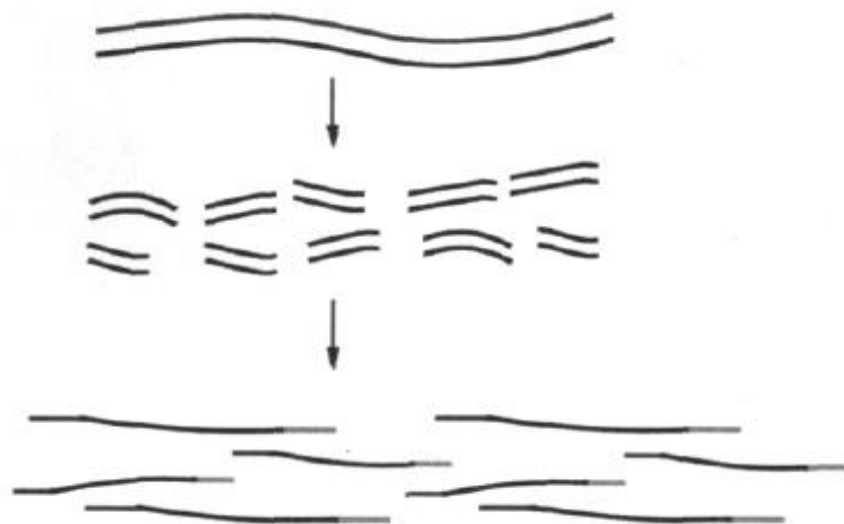
1. 将基因组DNA打成几百个碱基（或更短）的小片断，在片断的两个末端加上接头(adapter)
2. 利用专利的芯片:其芯片表面连接有一层单链引物（Primer）,DNA片断成单链后通过芯片表面的引物碱基互补被一端“固定”在芯片上
3. 引物扩增使得单链DNA成为双链
4. 该双链变性后成为单链，其一端“固定”在芯片上，另外一端（5'或3'）随机和附近的另外一个引物互补，被“固定”住，形成“桥“(bridge)
5. 这样的反应在上千万DNA单分子上发生
6. 形成的单链桥，以周围的引物为扩增引物，在芯片表面进行扩增形成双链
7. 双链经变性成单链，再次形成桥，成为下一轮扩增的模板继续扩增反应
8. 反复30轮扩增，每个单分子得到1000倍扩增，成为单克隆“DNA簇群”
9. “DNA簇群”在Genome Analyzer 综合分析仪上进行序列分析



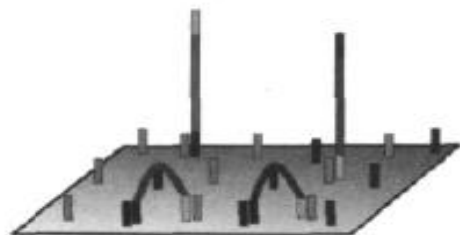
Solexa 测序流程

- (1)添加接头。用物理方法将待测样品DNA打碎，在碎片两端加上接头；
- (2)表面结合。Solexa 的测序时利用微注射系统将已经加过接头的待测片断添加到玻璃Flow cell 内，每一个Flow cell又被分为8条Lane，每条Lane的内表面上能与共价键的形式随机固定单链接头序列和带接头的单链待测DNA片断；
- (3)桥型扩增循环获得多拷贝待测DNA片断。在Flow cell 内加入未被标记的dNTP 和酶起始固相桥型扩增。所有的单链桥型待测片断被扩增成为双链桥片断，通过变性，释放出互补的单链，锚定到附近的固相表面。通过不断循环，将会在Flow cell 的固相表面上获得上百万条成簇分布的双链待测片断；
- (4)测序。加入DNA 聚合酶和被荧光标记的dNTP 和接头引物进行扩增，在每一个测序簇延伸互补链时，每加入一个被荧光标记的dNTP就能释放出相对应的荧光，测序仪通过捕获荧光信号，并通过计算机软件将光信号转化为测序峰，从而获得待测片段的序列信

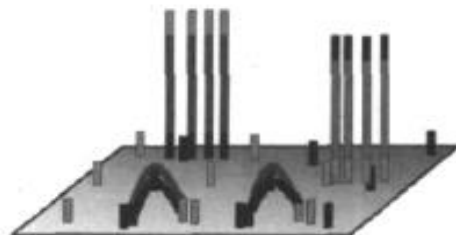
(1) 测序文库的构建
(1) Library construction



(2) 锚定桥接
(2) Surface attachment and bridge amplification



(3) 预扩增
(3) Denaturation and complete amplification



(4) 单碱基延伸测序
(4) Single base extension and sequencing

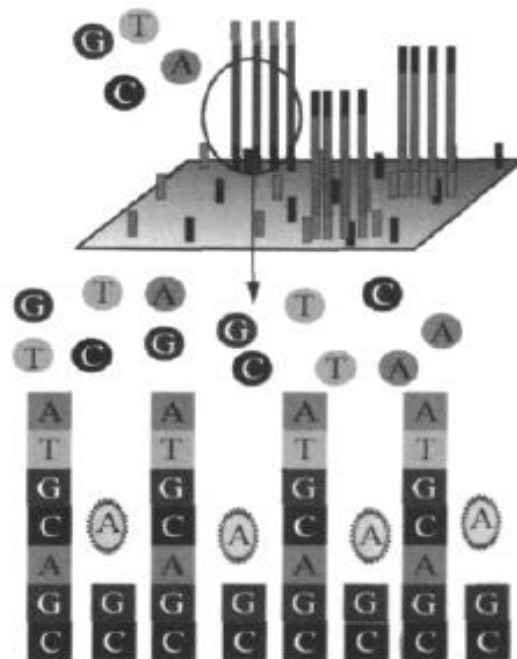


图 3 Illumina genome analyzer (Solexa)测序流程(Mardis, 2008)

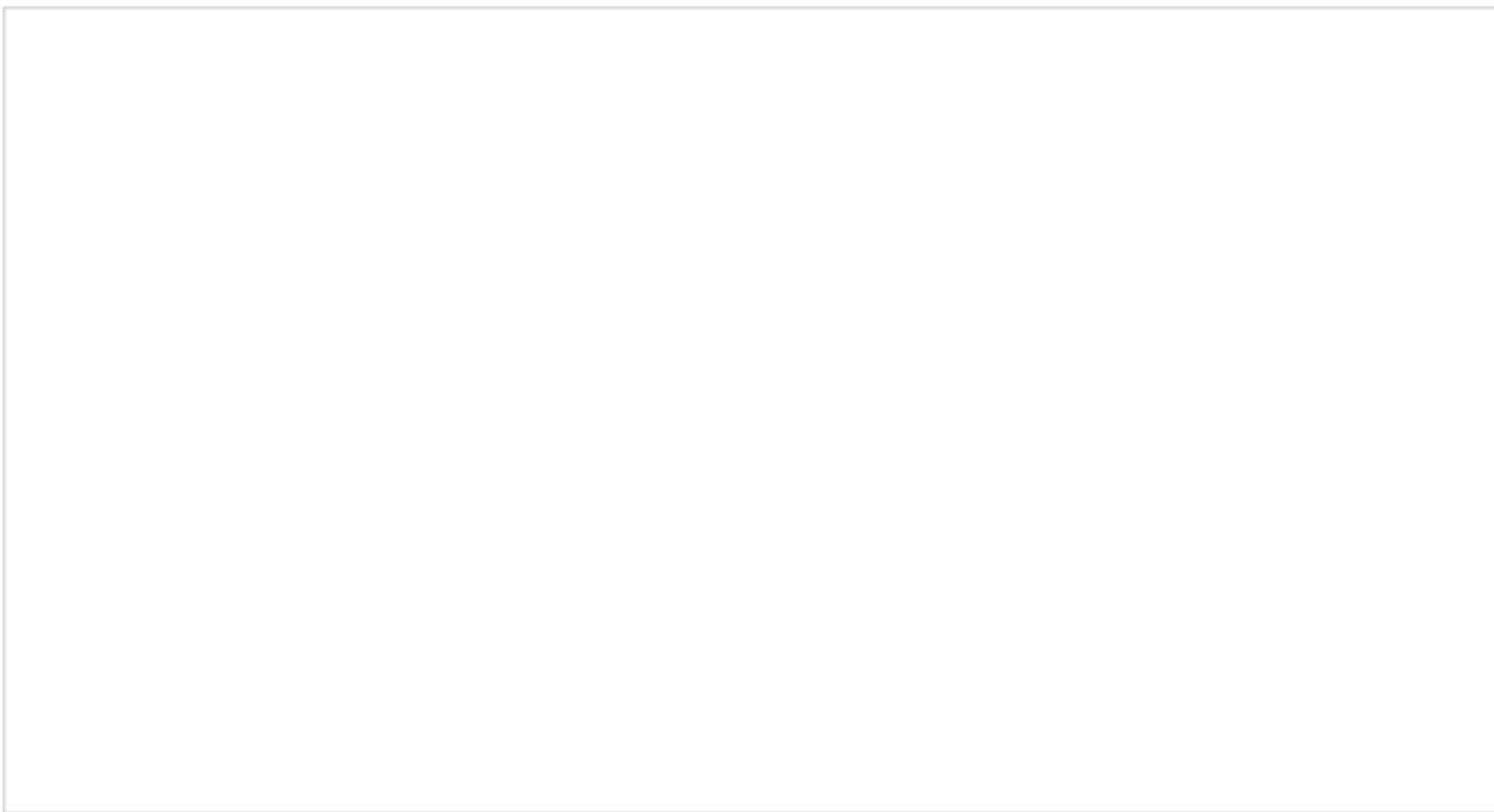
注: (1)测序文库的构建. 将基因组 DNA 随机打断后在单链 DNA 两端加上接头; (2)锚定桥接. 每一个带接头的 DNA 片断与测序通道上的接头引物随机结合, 添加未标记的 dNTP 和普通 *Taq* 酶进行固相桥式 PCR 扩增; (3)预扩增. 通过变性和桥式扩增循环在测序通道表面获得数百万条密集成簇的待测 DNA 片断; (4)单碱基延伸测序. 将 4 种被标记的 dNTP、引物和 DNA 聚合酶添加到测序通道内以启动地一个测序循环. 通过激光的激发, 从每个测序通道的测序簇里面产生出对应的荧光, 通过判断捕获的荧光颜色记录待测序簇的第 1 个碱基(<http://www.illumina.com/pagesnrn.ilmn?ID=70>)



武汉大学

Wuhan University

Solexa 测序原理及流程





技术特色突出表现在

1. 每张芯片有8个通道，每个通道可单独运行一个样品，也可以把多个样品混合在一起检测
2. 一次实验可读取10亿个碱基/芯片
3. 可精确读取重复序列如
AAAAAAAAAAAAAAAAAAAA, TTTTTTTTTTTTTTTT
4. 实验费用低，读取1百万碱基的成本最低至4美元
5. 无需建库，自动化样品制备，简单，成本低
6. 样本使用效率极高，所以对少量样本也可以极灵敏精确地检测（1 ug DNA即可以进行末端双向测序反应）
7. 可以进行35碱基长度的末端双向测序反应



SoIexa技术在应用中有如下特点

1. 不受物种限制，对人，动物，微生物，植物都可进行研究
2. 高灵敏度，精确性，及重复性
3. 无需预先知道模式物种基因组序列，无须合成探针，可直接进行全基因组表达研究
4. 不需要实验假设的支持
5. 可以检测到单拷贝分子的变化
6. 没有传统方法的荧光背景噪音的干扰



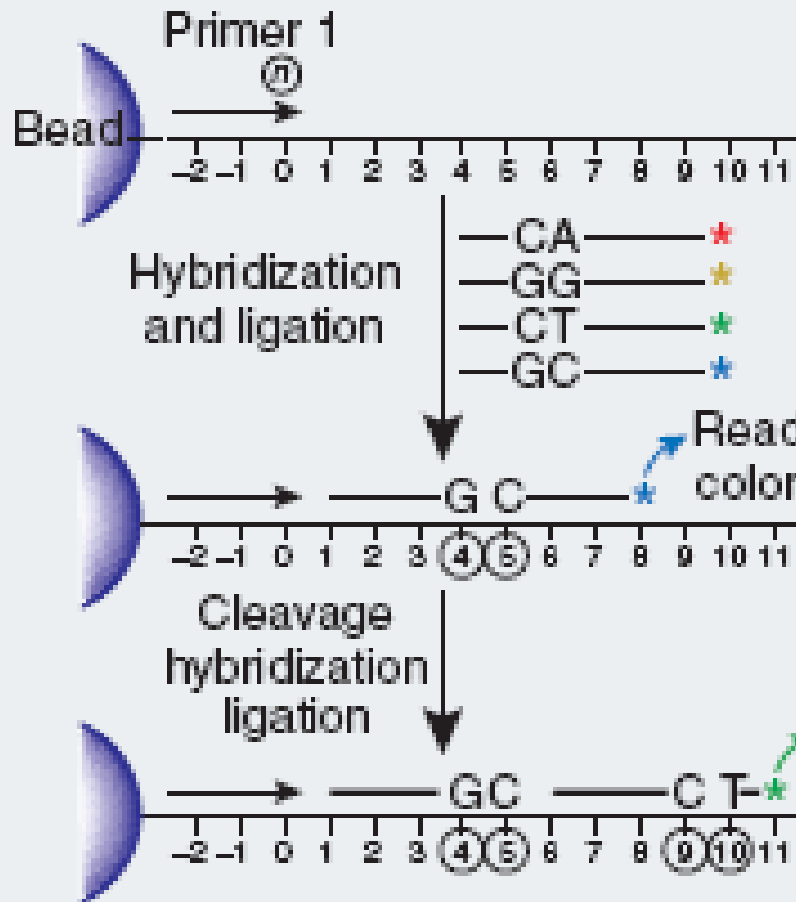
3、ABI SOLiD (SOLiD sequencer)

Applied Biosystems SOLiD (Supported Oligo Ligation Detection) (Sequencing by Oligonucleotide Ligation / Detection) sequencer 与其他新一代测序仪不同的是，它不是利用DNA 聚合酶在合成过程中读取序列，而是利用DNA连接酶在连接过程读取序列。

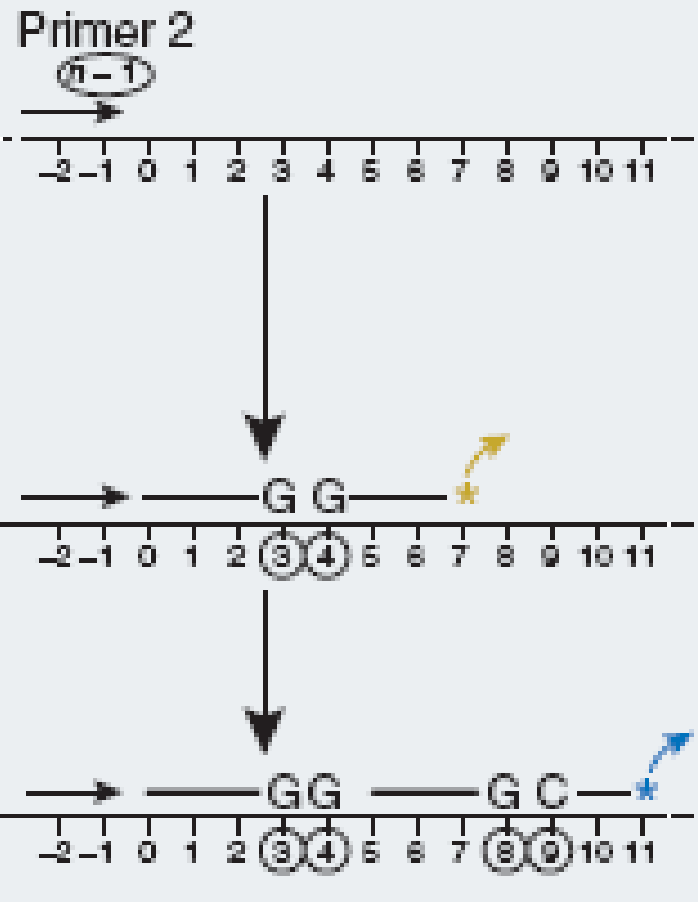
SOLiD 测序的基本流程为：

- (1) 测序文库：在DNA 碎片两端加上一对接头P1、P2。
- (2) 乳滴PCR
- (3) 边连接边测序

First round



Second round



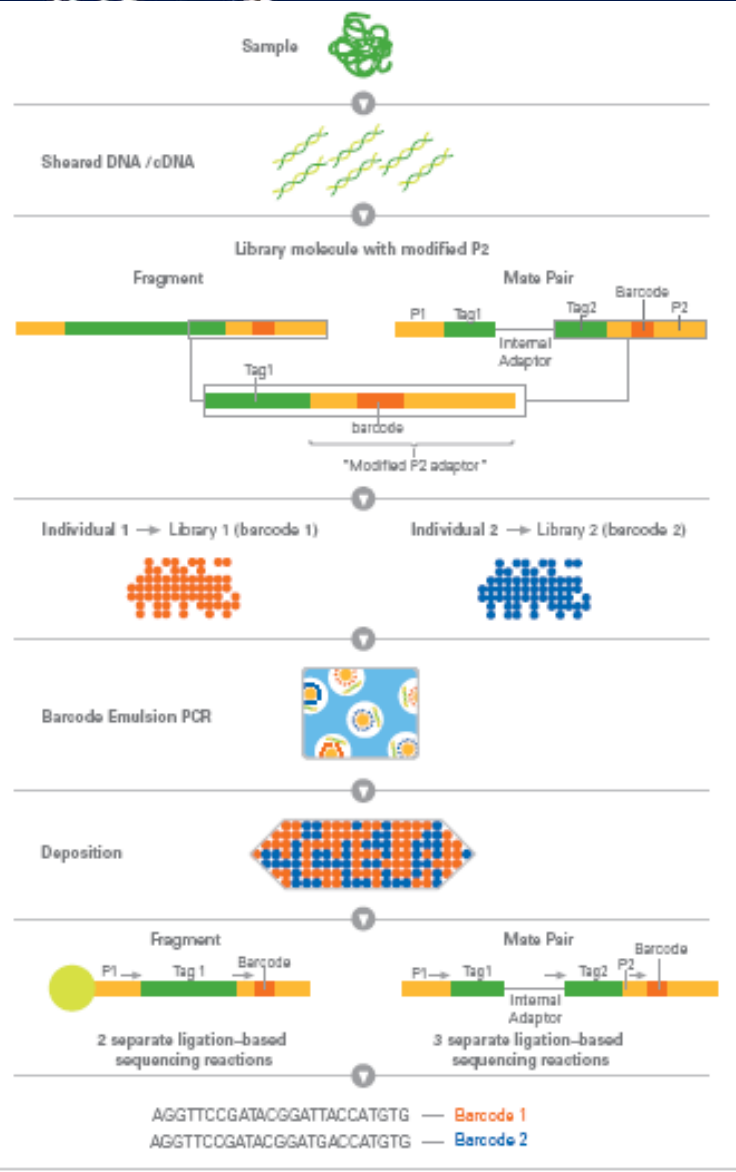
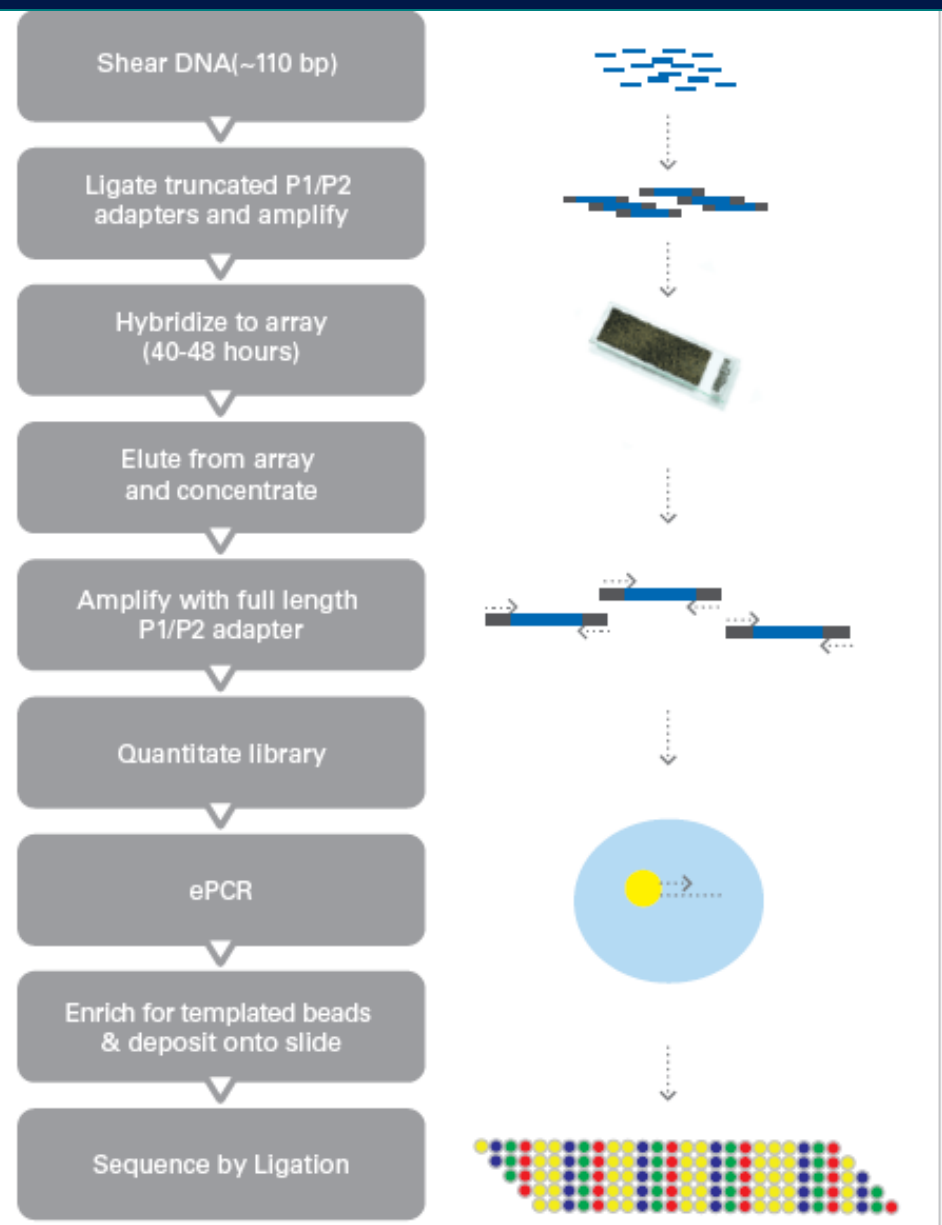


Figure 1. Integration of SOLiD™ System barcodes into the library construction workflow.

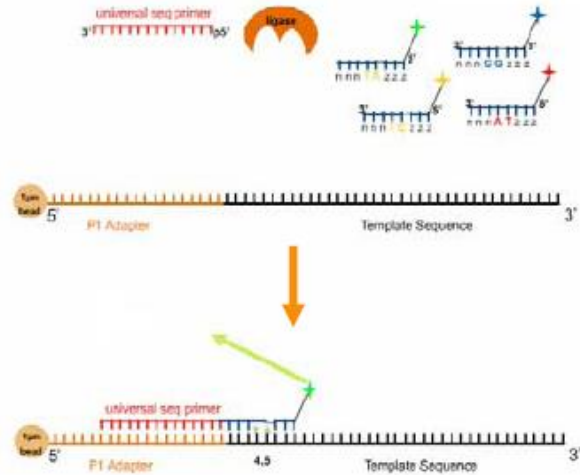


Workflow for array-based enrichment of targeted genomic regions prior to sequencing with the SOLiD™ System

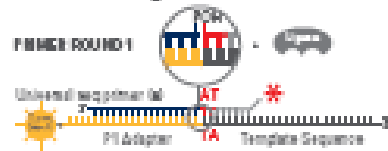


ABI SOLiD Overview

SOLiD™ Chemistry System 4-color ligation Ligation reaction



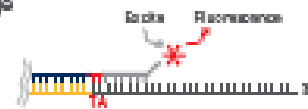
1. Prime and Ligate



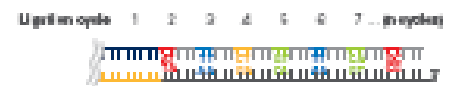
4. Cleave off Fluor



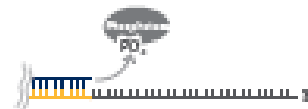
2. Image



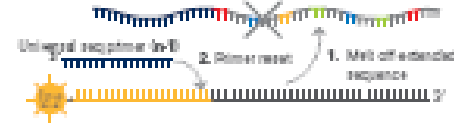
5. Repeat steps 1-4 to Extend Sequence



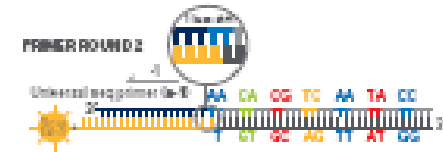
3. Cap Unextended Strands



6. Primer Reset



7. Repeat steps 1-5 with new primer



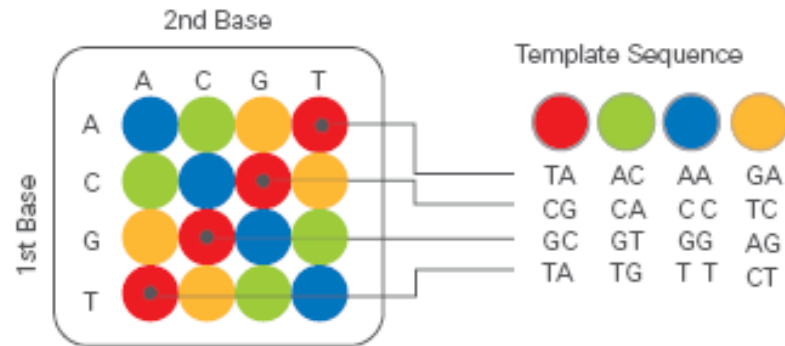
8. Repeat Reset with n-2, n-3, n-4 primers

Primer Round	Read Position	Read Position															
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	Universal seq primer (n)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
2	Universal seq primer (n-1)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
3	Universal seq primer (n-2)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
4	Universal seq primer (n-3)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
5	Universal seq primer (n-4)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•

• indicates positions of interrogation. Ligation Cycle: [Blue] [Orange] [Green] [Red]

Figure 1: SOLiD® System – Sequencing by ligation using di-base labeled probes

Possible Dinucleotides Encoded By Each Color



Double Interrogation

With 2 base encoding each base is defined twice



Figure 2. SOLiD Color Space Code — Four dyes encode for sixteen potential two base combinations

TABLE 1. COLOR SPACETRANSITIONS

	Blue	Green	Yellow	Red
Blue	BB	BG	BY	BR
Green	GB	GG	GY	GR
Yellow	YB	YG	YY	YR
Red	RB	RG	RY	RR

All 2 color possibilities are represented for the four possible dyes. Four sets of allowable changes are classified by color. Changes that fall outside a particular set can be classified as possible errors for further analysis.



武汉大学

Wuhan University

SOLiD测序原理与步骤

- [SOLiD video final.wmv](#)





Sequencing Platform Comparison

	454FLX(Roche)	Solexa(Illumina)	AB SOLiD
Read length	250bp (400)	36bp (50) or 2x36bp	25bp (30)
Number of reads/run	420K	~40M	40M
Data output	~100Mb	~1Gb	~2Gb
Reagent cost per run	~\$8,000	~\$3,200	~\$6,000
Reagent cost per 1Gb	~\$80,000	~\$3,200	~\$3,000
Error	0.5%	0.2%	0.1%
Run time to 1Gb	10 days	2-3 days	7-10 days
Easy of use	Most difficult	Least difficult	454-like difficulty
Base calling	Flow Space	Nucleotide Space	Color Space
Applications			
Whole transcriptome shotgun	Yes	Yes	Yes
Whole genome shotgun	Yes	Yes	Yes
Paired-read	Yes	Yes	No
Small RNA	No	Yes	No
ChIP-Sequencing	No	Yes	No
Expression	Yes	Yes (1st application)	Yes

表11 新一代测序技术的应用

应用范围	应用举例
全基因组测序 (Complete genome resequencing)	人类个体基因组多态性及突变的全面检测
约化表示测序法 (Reduced representation sequencing)	大规模多态性检测
靶向再测序 (Targeted genomic resequencing)	靶向多态性及突变检测
末端配对测序 (Paired end sequencing)	遗传及获得性结构变异检测
环境基因组测序 (Metagenomic sequencing)	传染性 & 共生菌群检测
转录组测序 (Metagenomic sequencing)	定量基因表达及选择性剪切; 转录注释; 转录 SNPs或体细胞突变检测
小RNA测序 (Small RNA sequencing)	microRNA表达谱
酸性亚硫酸盐标记DNA测序 (Sequencing of bisulfite-treated DNA)	基因组DNA中胞嘧啶甲基化模式的测定
染色质免疫沉淀测序 (ChIP-Seq)	全基因组蛋白质与DNA相互作用图谱
核酶片段及测序 (Nuclease fragmentation and sequencing)	核小体定位
分子条码 (Molecular barcoding)	多个体来源样品的多通路测序



四、第三代测序技术

1、Helicos Biosciences

Heliscope

2、Pacific Biosciences

(Single Molecule Real-time Technology)

SMRT

3、Oxford Nanopore Technologies





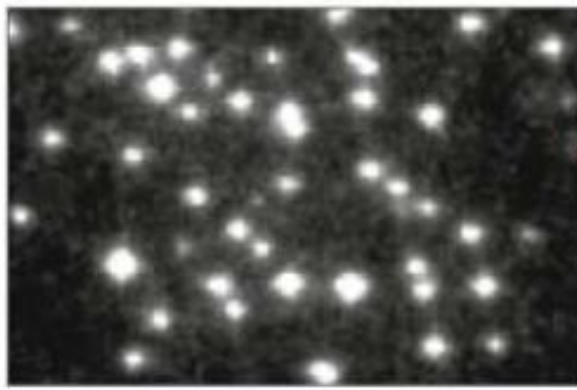
4.1 HeliScope测序仪

HeliScope测序仪是由Quake团队设计开发的，也是一种循环芯片测序设备。其最大的特点是无需对测序模板进行扩增，它使用了一种高灵敏度的荧光检测仪直接对单链DNA模板进行合成法测序。

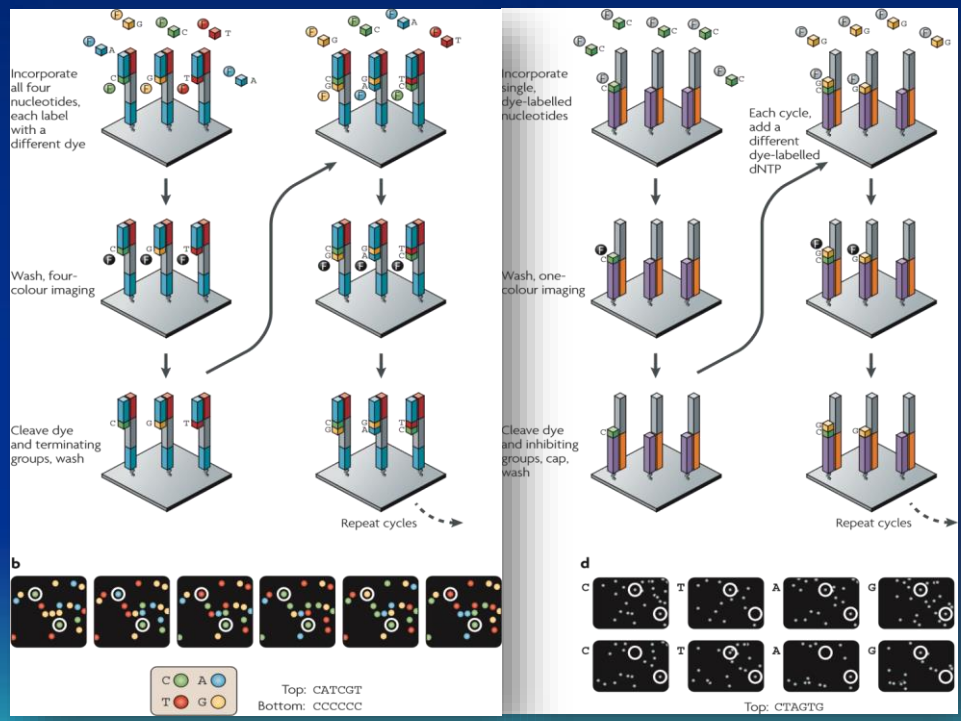
首先，将基因组DNA切割成随机的小片段DNA分子，并且在每个片段末端加上poly-A尾。然后通过poly-A尾和固定在芯片上的poly-T杂交，将待测模板固定到芯片上，制成测序芯片。最后借助聚合酶将荧光标记的单核苷酸掺入到引物上。采集荧光信号，切除荧光标记基团，进行下一轮测序反应，如此反复，最终获得完整的序列信息。根据最近的报道，经过数百轮这种单碱基延伸可以获得25bp或更长的测序长度。

HeliScope测序仪的其它特点见下表





- 1、Cy5-dATP, 采集信号, 切除荧光标记
- 2、Cy5-dGTP, 采集信号, 切除荧光标记
- 3、Cy5-dCTP, 采集信号, 切除荧光标记
- 4、Cy5-dTTP, 采集信号, 切除荧光标记



HeliScope测序仪：单核酸分子不经扩增直接测序。Poly-A尾被添加到DNA文库片段末端，通过与固定在芯片上的Poly-T互补杂交将模板链固定到芯片，制成测序芯片。模板上标记有Cy3以标识出它们在芯片上的位置。DNA聚合酶将荧光标记的单核苷酸掺入到引物上，采集荧光信号获得序列信息，切除荧光基团，重复上述步骤，完成测序。

Pushkarev D, Neff N F, Quake S R, 2009



武汉大学

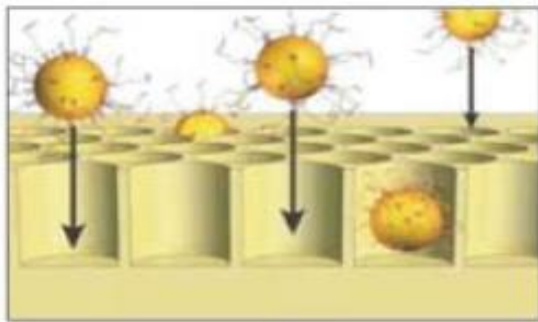
Wuhan University

HeliScope单分子测序过程

[heliscope.swf](#)



a



3'-AGATACTATATCGAGATCCAG...-5'-[微珠]
5'-TCTATGATA

(a) 454测序仪使用的方法

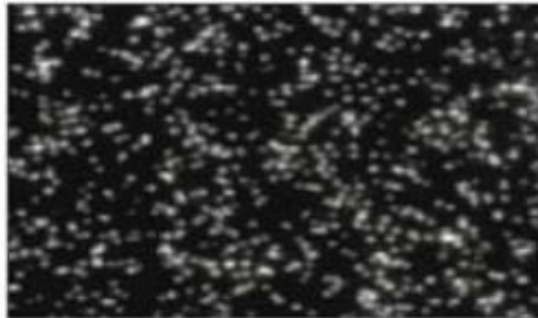
焦磷酸 + 腺苷酰硫酸

ATP + 荧光素

氧化荧光素 + 发光

- 1、dATP α S, 底物, 腺苷三磷酸双磷酸酶洗涤
- 2、dGTP, 底物, 腺苷三磷酸双磷酸酶洗涤
- 3、dCTP, 底物, 腺苷三磷酸双磷酸酶洗涤
- 4、dTTP, 底物, 腺苷三磷酸双磷酸酶洗涤

b

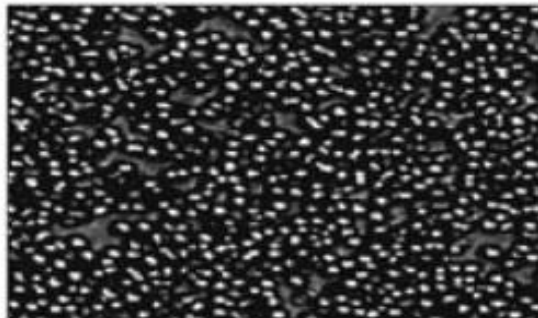


(b) Solexa测序仪使用桥式PCR

3'-GAGCAGGAGGACATATCAGAG...-5'-[载体表面]
5'-CTCGTCTTC

- 1、荧光基团1-dATP-(阻碍物基团)+ 荧光基团2-dGTP-(阻碍物基团)+ 荧光基团3-dCTP-(阻碍物基团)+ 荧光基团4-dTTP-(阻碍物基团)
- 2、四个通道中荧光成像
- 3、化学法切除荧光标记基团及阻碍物终止基团

c

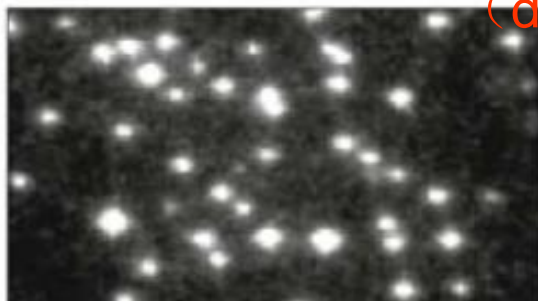


3'-GAACATACGACTATAGACCTA...-5'-[微珠]
TCTGGAT...-3'
FLx
5'-zzzXXnnn-3'

(c) SOLiD测序仪

- 1、连接上特定的8bp探针
- 2、采集信号、切除, 重复5次
- 3、重新开始新一轮测序

d



(d) HeliScope测序仪, 单核酸分子不经扩增直接测序。

Cy3
3'-...AAAAAAAGACATACTATGAG-5'
[载体表面]- 5'-...TTTTTTT

- 1、Cy5-dATP, 采集信号, 切除荧光标记
- 2、Cy5-dGTP, 采集信号, 切除荧光标记
- 3、Cy5-dCTP, 采集信号, 切除荧光标记



3. 新一代测序技术的前景

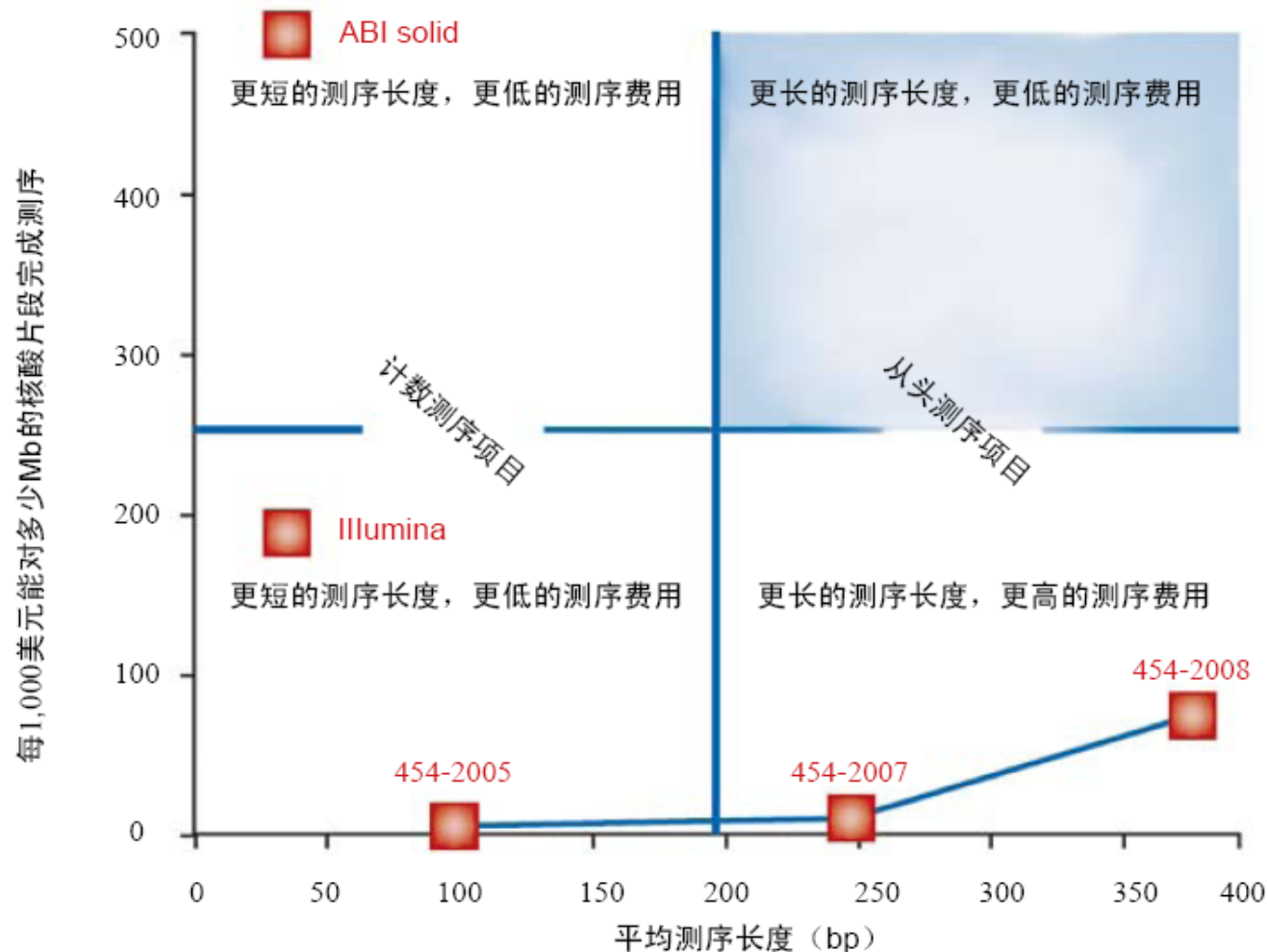


图7 新一代测序仪的比较情况。

对两种不同测序仪从测序长度和测序费用两方面进行比较。Illumina公司的Solexa测序仪和ABI公司的SOLiD测序仪的测序费用数据和测序长度数据来自许多渠道，包括各自生产厂家的网页和最新的技术资料。454公司从2005年至2008年上半年的数据反应了它们在测序长度和测序费用方面的进步。

表10 新一代DNA测序平台

新一代测序仪名称	模板扩增方法	合成测序方法中使用到的酶和试剂	测序费用 (美元) /Mb	设备价格 (美元/台)	是否末端配对	发生1° 错误的突变类型	测序长度
454	微乳液PCR	聚合酶 (焦磷酸测序)	~60	500,000	是	插入或缺失突变	250bp
Solexa	桥式PCR	聚合酶 (可被切除的终止基团)	~2	430,000	是	替换突变	36bp
SOLiD	微乳液PCR	连接酶 (由双碱基编码技术设计的8bp长度探针)	~2	591,000	是	替换突变	35bp
Polonator	微乳液PCR	连接酶 (九聚物)	~1	155,000	是	替换突变	13bp
HeliScope	单个分子检测	聚合酶 (非同步延长)	~1	1,350,000	是	缺失突变	30bp

注：DNA测序领域的快速发展使得对各类测序方法的价格及读长的评估在很短时间内便失去意义。Roche Applied Science、Illumina及Applied Biosystems公司目前都在不断推出新的产品。表中列出的测序费用只是对使用的反应试剂费用的一个估算。测序长度指的是单链长度。



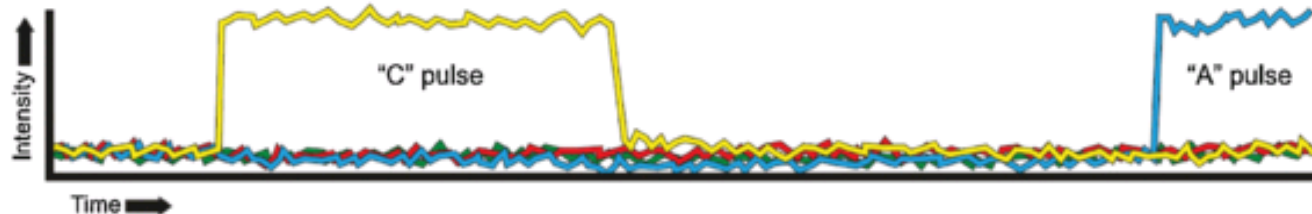
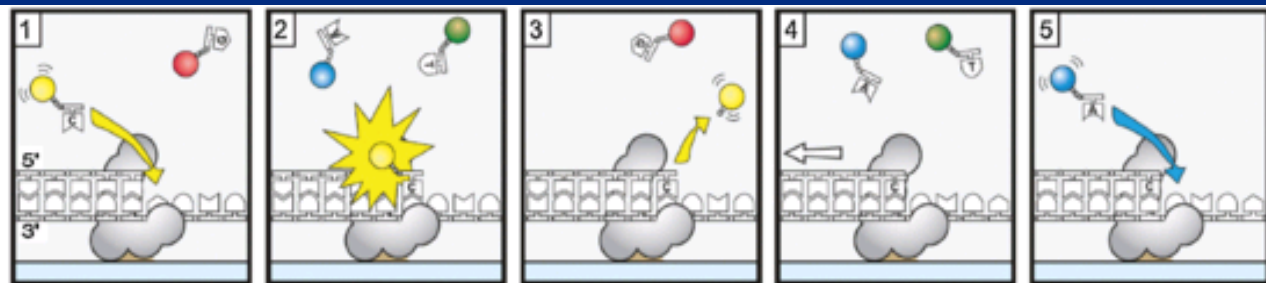
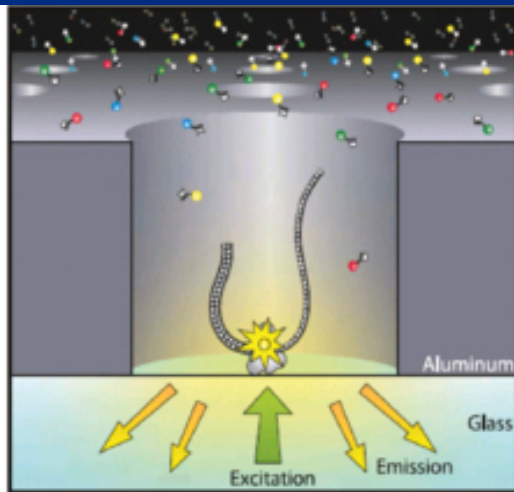
武汉大学

Wuhan University

Pacific Biosciences

5、Single Molecule Real-time Technology: SMRT

单分子实时DNA测序



Flusberg B A, Webster D R, Lee J H, et al., 2010

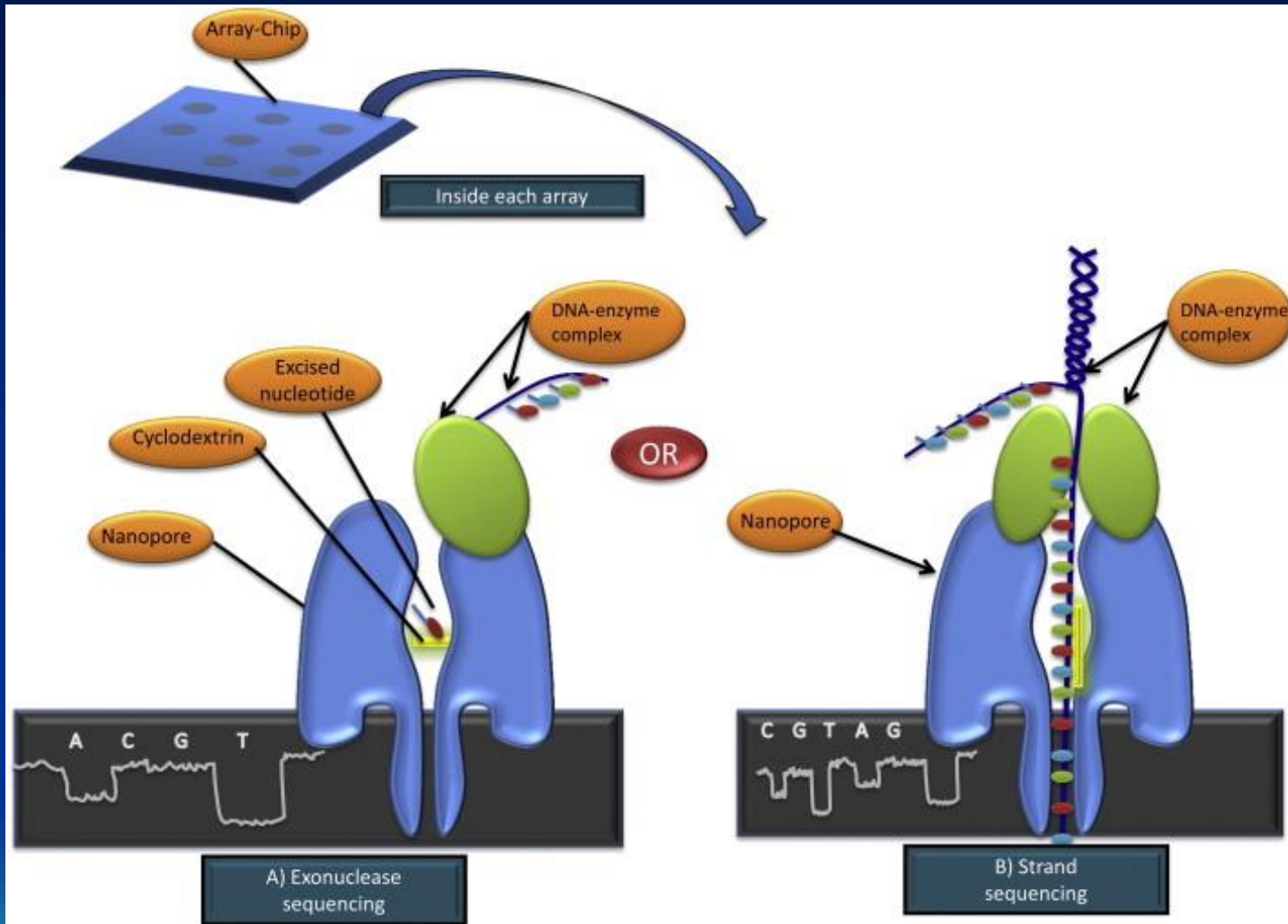


NANOPORE SEQUENCING PLATFORM

其核心是将在某一面上含有一对电极的特殊的脂质双分子层置于一个微孔之上，该双分子层中含有很多由 α 溶血素蛋白组成的直径约1.5nm的纳米孔中，结合一个核酸外切酶。当DNA模板进入孔道时，孔道中的核酸外切酶会“抓住”DNA分子，顺序剪切掉穿过纳米孔道的DNA碱基，每一个碱基通过纳米孔时都会产生一个阻断，根据阻断电流的变化就能检测出相应碱基的种类，最终得出DNA分子的序列。

人们使用硅和其他电子材料来制作纳米孔，使用微小的晶体管来控制通过小孔的电流，当4种不同的碱基通过小孔时，它们对电信号影响的方式不同，导致电压上升或者下降，可以依据这种微小的改变来确定通过DNA上碱基的类型。

纳米孔单分子测序技术的一大优势就是仪器构造简单使用成本低廉；因为它不需要对核苷酸进行标记，也不需要复杂的光学探测系统（如激光发射器和CCD信号采集系统等），能直接对RNA分子进行测序。同时由于它是直接检测每一个碱基的特征性电流，因而能对修饰过的碱基进行测序，这一点对于表观遗传学研究具有极高的价值；缺点就是它采用的是水解测序法，不能进行重复测序，因而无法达到一个满意的测序精确度



Wheeler D A, Srinivasan M, Egholm M, et al. 2008



Nanopore.swf

Nanopore Technologies



	测序方法/ 平台	方法/酶	测序长度	每个循环的 数据产出量	每个循环耗 时	主要错误来 源
第一代测序 技术	Sanger/ABI37 30 DNA Analyzer	Sanger法 /DNA聚合酶	1000 bp	56 Kb		
第二代测序 技术	454/GS FLX Titanium Series	焦磷酸测序 法 /DNA聚合酶	400 bp	400-600 Mb	10 h	插入、缺失
	Solexa/Illumin a Genome Analyzer	边合成边测 序 /DNA聚合酶	2*75 bp	20.5-25 Gb	9.5 d	替换
	SOLiD/SOLiD 3 system	连接酶测序 /DNA连接酶	2*50 bp	10-15 Gb	6-7 d	替换
第三代测序 技术	Heliscope/Hel icos	边合成边测 序 /DNA聚合酶	30-35 bp	21-28 Gb	8 d	替换
	SMRT	边合成边测 序 /DNA聚合酶	100000 bp			插入、缺 失
	Nanopore Sequencing	电信号测序/ 核酸外切酶	无限长			缺失
	FRET Sequencing	FRET/DNA聚 合酶				缺失
	The Ion Torrent Sequencing	半导体/DNA 聚合酶				缺失





新一代测序方法的应用主要有

(1) **基因组 *de novo* 测序** 基因组从头测序 (*de novo sequencing*) 是在没有参考基因组的情况下, 对某种物种进行基因组测序及组装, 从而构建出该物种的全基因组组装序列图谱, 即完整的物理图谱。

(2) **基因组重测序 重测序 (*Re-sequencing*)** 是对已知基因组序列的物种进行不同个体的基因组测序, 并在此基础上对个体或群体进行差异性分析。

(3) **简化基因组测序** 简化基因组测序 (*Reduced-representation sequencing*) 技术是指利用生物信息学方法, 设计标记开发方案, 筛选特异性长度片段, 应用高通量测序技术获得海量标签序列来充分代表目标物种全基因组信息的测序方法。目前简化基因组测序技术主要有:



①利用限制性位点关联的**DNA标志物测序** (**restriction-site associated DNA tags, RAD-seq**) 技术进行简化基因组测序, 通过筛查两侧带有特定核酸内切酶识别位点的DNA短片段 (RAD标签) 进行遗传变异分析, 从而降低基因组测序的复杂度, 快速鉴定出成千上万的单核苷酸多态性 (SNPs) 标记。RAD-seq 技术在模式和非模式生物的基因型-表型关联图谱、系统进化、群体遗传等遗传分析研究领域具有广泛的应用前景

②以高通量测序为基础的**SLAF-seq** (**specific-locus amplified fragment sequencing/ specific length amplified fragment sequencing**) 技术, 以较低的花费实现全基因组范围的标记开发和分型, 是目前最热门的标记分型技术之一。利用SLAF-seq对个体或者群体DNA混池进行标记分型, 可以构建高密度遗传图谱、进行集群分离分析, 从而实现性状相关标记筛选、基因定位的目的等。



(4) 转录组测序 转录组是指特定细胞在某一功能状态下所能转录出来的所有RNA的总和。转录组研究是基因功能及结构研究的基础。转录组测序技术是把mRNA、smallRNA和非编码RNA等用高通量测序技术把它们的序列测出来，全面快速地获取某一物种特定器官或组织在某一状态下的几乎所有转录本，反映出它们的表达水平。



基因组图谱的应用

- (1) **寻找新的基因**：从基因组序列中进行基因搜寻、分析与基因相关的序列，这是解读整个基因组图谱的基础。
- (2) **基因的克隆与分离**：根据饱和的基因组图谱，可以对基因进行克隆和分离。
- (3) **基因功能的预测**：利用基因组图谱还能够对基因功能进行预测，观察基因组作为一个整体如何行使其功能。基因功能的预测可以通过计算机预测和实验确定两种途径进行。
- (4) **利用基因组图谱数据进行比较基因组学研究**：通过基因组比较作图可以揭示染色体或染色体片段上同线性或共线性的存在，从而对不同物种的基因组结构及基因组进化历程进行精细分析。
- (5) **基因定位**：借助基因组图谱，可使基因定位在精度、深度、广度等方面有极大的提高。



武汉大学

Wuhan University

本节 结束

谢谢!





焦磷酸测序技术的原理

1987年Nyren 等发展一种新型的DNA测序技术:

焦磷酸测序技术(**Pyrosequencing**)

是由四种酶催化的同一反应体系中的酶级联反应

四种酶分别为: DNA 聚合酶 (DNA polymerase)

ATP 硫酸化酶(ATP sulfurylase)

荧光素酶(luciferase)

双磷酸酶(apyrase)

反应底物为:

5'-磷酸硫酸(adenosine 5'-phosphosulfate ,APS)

荧光素(luciferin)

反应体系还包括待测序DNA单链和测序引物。

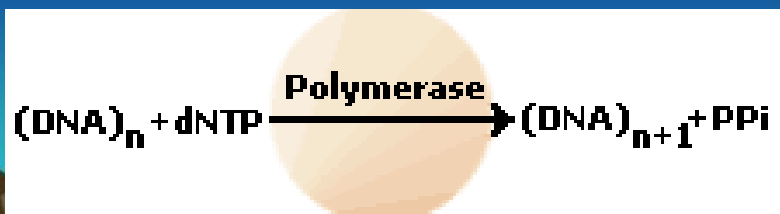


具体原理:

引物与模板DNA 退火后,在上述四种酶的协同作用下,每一个dNTP 的聚合与一次荧光信号的释放偶联起来,以荧光信号的形式实时记录模板DNA的核苷酸序列。

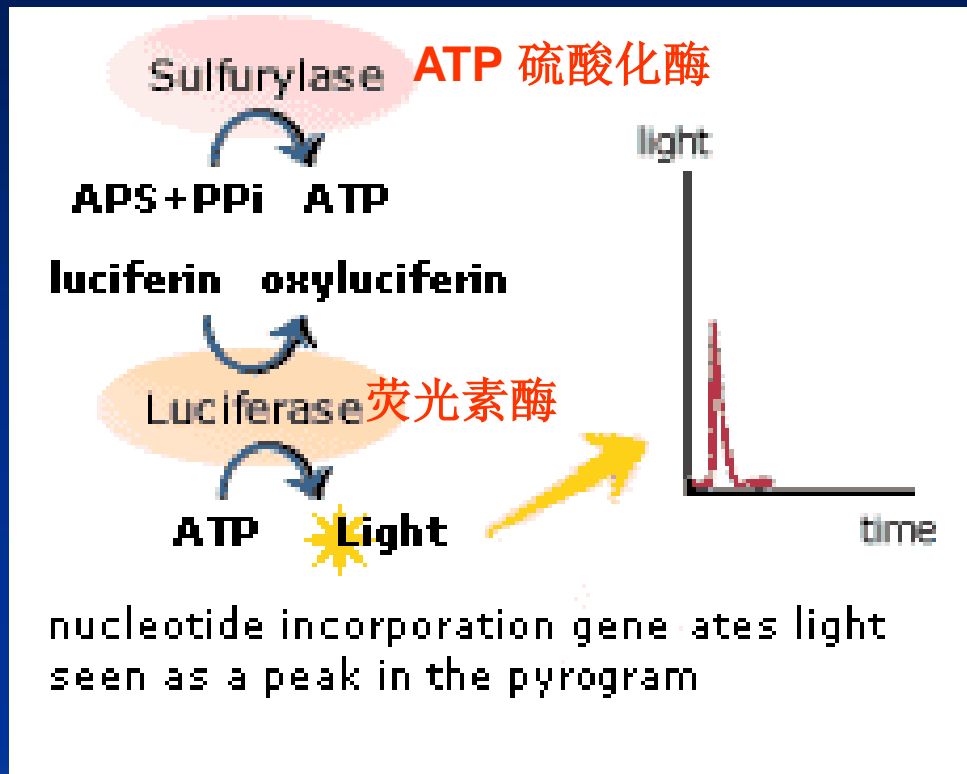
第1步: 将特异性的测序引物杂交到PCR扩增的模板上,然后加入DNA 聚合酶、ATP 硫酸化酶、荧光素酶和双磷酸酶及其反应底物:5'- 磷酰硫酸(APS)、荧光素,组成一个反应体系一起孵育。

第2步: 在每一轮测序反应中,只能加1种dNTP,如该dNTP 与待测模板配对,聚合酶就可以催化该dNTP 掺入到引物链中并释放出等摩尔数的焦磷酸基团(PPI) 。

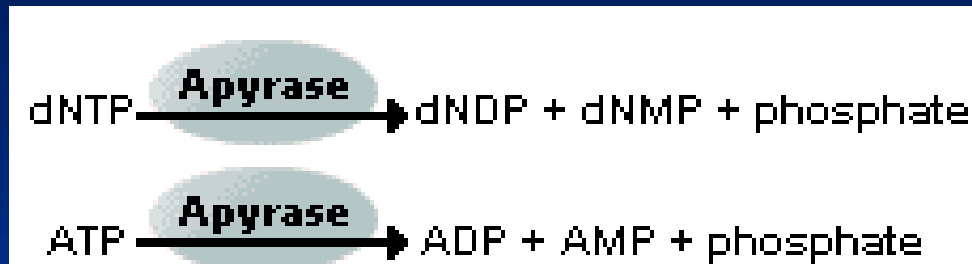




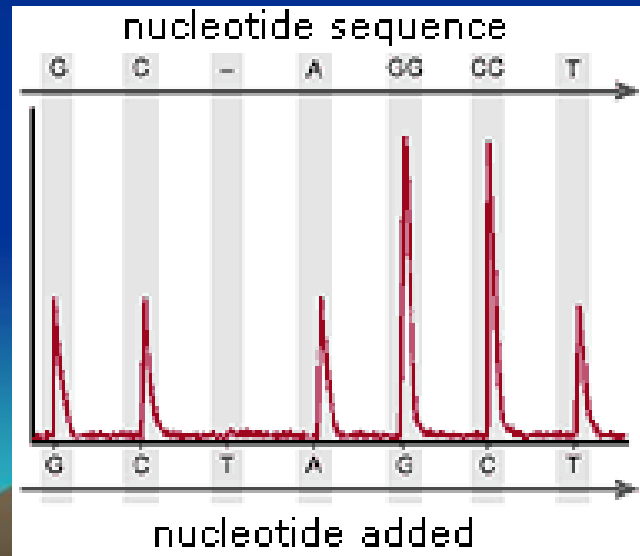
第3步：在ATP硫酸化酶的作用下，生产的ppi可以和APS结合形成ATP；在荧光素酶的催化下，生产的ATP又可以和荧光素结合形成氧化荧光素，同时产生可见光，通过CCD光学系统即可获得一个特异的检测峰，峰值的高低则和相匹配分碱基数成正比。



第4步：反应体系中剩余的dNTP和残留的少量ATP在Apyrase的作用下发生降解：



第5步：加入另一种dNTP,使第2—第四步反应重复进行，根据获得的峰值同时可读取准确的DNA序列：



ACGTGGGCCTATAGCTACTCGGACACCTACGCATATCGCCCG

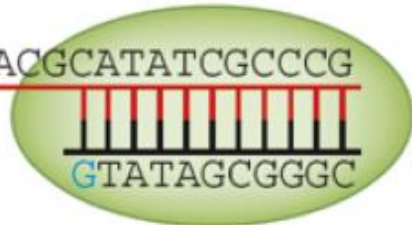


图3 焦磷酸测序反应原理。
图中红色单链表示模板链，引物以黑色表示，DNA聚合酶以绿色的椭圆形表示。每当掺入一个碱基（如图中蓝色的G）时，就会释放出焦磷酸（PPi），然后被磷酸化酶（sulfurylase，图中蓝色箭头所示）转化成ATP。然后荧光素酶（图中红色箭头所示）就可以在ATP参与下将荧光素转变成氧化荧光素，同时发光也会被测序仪检测到。

