

第12章 基因组学

第1节 基因组学概述

第2节 基因组图谱

第3节 基因组测序

第4节 后基因组学

第1节 基因组学概述

- 一、基因组学的概念
- 二、生物体基因组特征
- 三、人类基因组计划
- 四、模式生物基因组研究
- 五、水稻基因组计划

一、基因组学的概念

- **基因组(Genome)**: 又称染色体组，是指一个物种单倍体的染色体数目，是生物体全部遗传物质的总和。
- **基因组学(Genomics)**: 对生物体所有基因进行基因组作图(包括遗传图谱、物理图谱、转录图谱)、核苷酸序列分析、基因定位和基因功能分析的一门科学。
- **最终目标**: 获得生物体全部基因组序列，注解基因组所含的全部基因，鉴定所有基因的功能及基因间相互作用关系，并阐明基因组的复制及进化规律。

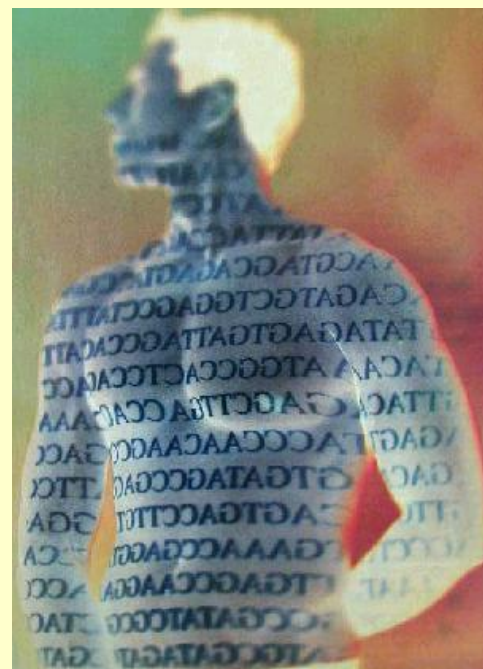
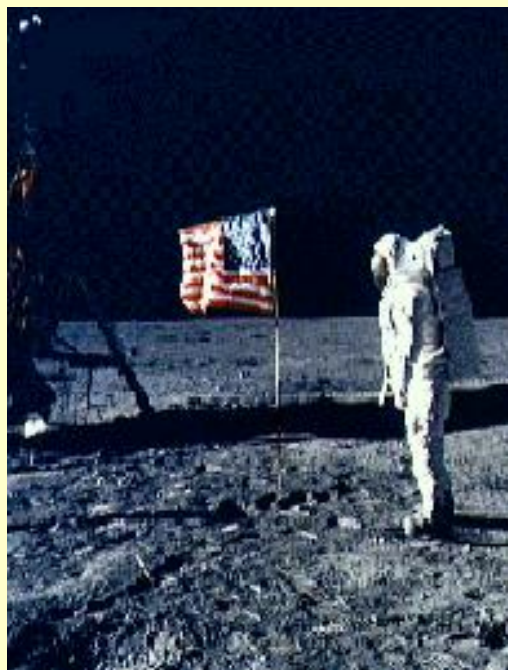
生物	基因组大小 (bp)
T4噬菌体 <i>T4 phage</i>	2.0×10^5
大肠杆菌 <i>Escherichia coli</i>	4.2×10^6
酵母 <i>Sccharomyces cereviside</i>	1.5×10^7
拟南芥 <i>Arabidopsis thaliana</i>	1.0×10^8
线虫 <i>Caenorhbditis elegans</i>	1.0×10^8
果蝇 <i>Drosophila melanogaster</i>	1.65×10^8
水稻 <i>Oryza sativa</i>	4.3×10^8
小鼠 <i>Mus musculus</i>	3.0×10^9
人类 <i>Homo sapiens</i>	3.3×10^9
玉米 <i>Zea mays</i>	5.4×10^9
小麦 <i>Triticum aestivum</i>	1.6×10^{10}

二、基因组学的发展

高参考价值的真题、答案、学长笔记 辅导班课程，访问：www.kaoyancas.net

1. 人类基因组计划

与曼哈顿原子计划、阿波罗登月计划并称的人类科学史上的重大工程。于**1990**年首先在美国启动，后有德、日、英、法、中等国的科学家先后正式加入。



完整版，请访问www.kaoyancas.net 科大科院考研网，专注于中科大、中科院考研

▲美国从70年代起启动了“肿瘤计划”，但是，不惜血本的投入换来的是令人失望的结果。人们渐渐认识到，包括癌症在内的各种人类疾病都与基因直接或间接相关。测出基因的碱基序列，则是基因研究的基础。

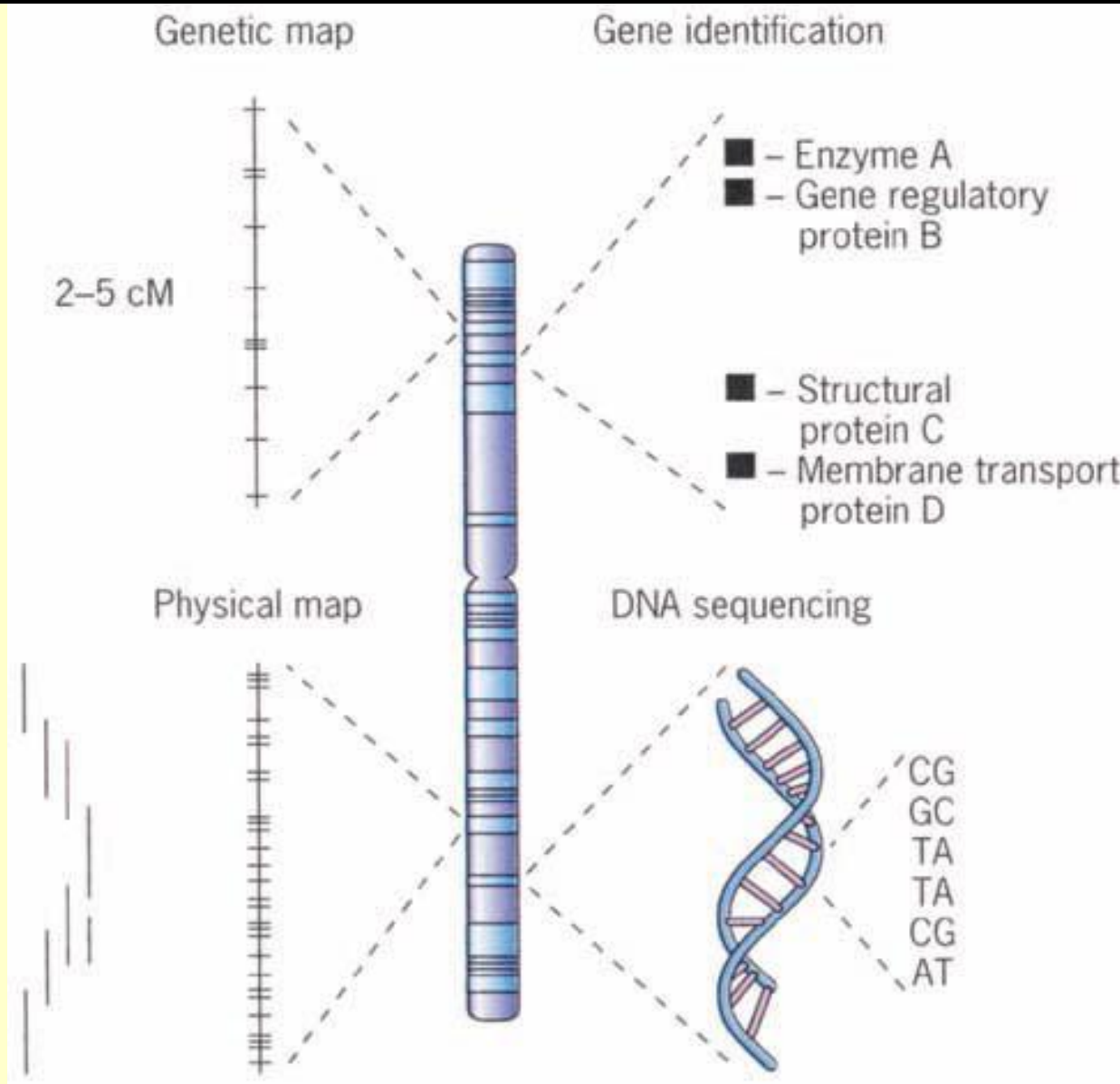
▲ 1986年3月，杜伯克在美国《科学》杂志上发表了一篇题为《癌症研究的转折点：测序人类基因组》的文章，这篇短文后来被称为人类基因组计划的“标书”。

▲ 1990年，美国国会批准美国的“人类基因组计划”在10月1日正式启动。其总体规划是准备在15年内（1990—2005）至少投入30亿美元，分析人类的基因组30 亿个碱基对。

▲ 2003年，6国科学家宣布人类基因组序列图绘制成功，HGP的所有目标全部实现。覆盖人类基因组所含基因区域的99%，精确率达到99.99%，比原计划提前两年多，耗资27亿美元。

四张图——物理图、转录图、遗传图、序列图

高参考价值的真题、答案、学长笔记、辅导班课程，访问：www.kaoyancas.net



完整版，请访问www.kaoyancas.net 科大科院考研网，专注于中科大、中科院考研

☆人类基因组的物理图有两个要素：一是序列，二是位置。在如此长的序列中，物理图就像地图一样标明各个序列的路标。通过分子杂交的办法，利用**DNA**双链互补特点，一个**DNA**片段杂交在这个位置，“说明这个位置的结构与它相似，就是这个位置的标记，这是以序列作为标记的位置。”

☆物理图重要的作用，有了前面标志的序列位置，就可以将克隆的**DNA**片段，一个一个接起来。

☆转录是表达的第一阶段，**DNA**转录后的**mRNA**携带信息，转录图是基因图的雏形。

☆**DNA**片段的部分序列——表达标签序列(**EST**)。转录图中的基因是有表达功能的基因；且转录本身是有组织与时间特异性的，因此可以绘制出在正常条件下基因表达的数目、种类及结构、功能的信息。有了正常和异常的转录图，就可以在此基础上构建基因表达谱。

☆一个基因在基因组中有其位点，这个位点至少有两个等位基因，一个是正常的，一个是不正常的。如果不正常的基因不表达，这个人还是正常的，仅仅是一个携带者。这个位点和全部基因组的遗传标记存在着距离。

☆采用一个遗传标记，来检查家系中这个遗传标记的位点是否与致病位点发生交换。靠物理图就可以在这个遗传标志的相应距离找到这个基因。序列中的差异是最好的“遗传标记”。

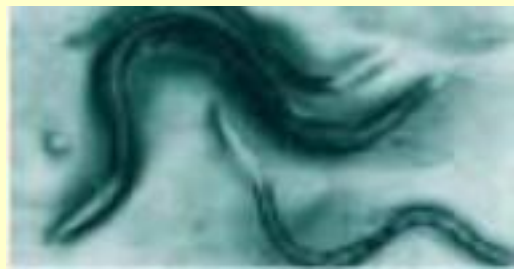
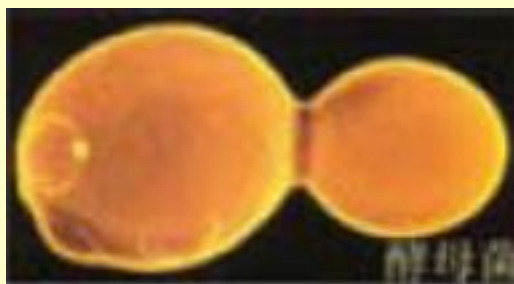
物理图、转录图和遗传图都是序列前计划，这些图的绘制，都是为人类基因组的序列图作准备，只有序列图完成了，才能将人群内序列的差异，作为密度最高的遗传标记来完善遗传图，因此序列图是人类基因组计划中的最重要部分。

2. 模式生物基因组研究

高参考价值的真题、答案、学长笔记、辅导班课程，访问：www.kaoyancas.net

为人类的基因组研究提供重要的依据。

- ☆ 1996年，酵母菌基因组测序。
- ☆ 1998年12月，线虫完整基因组序列的
- ☆ 2000年3月，果蝇的基因组测序
- ☆ 2001年12月14日，拟南芥基因组的完整图谱。



我国超级杂交稻（籼稻）基因组计划 2001年7月启动
2002年4月5日《Science》。

☆材料：籼稻“9311”。

☆完成单位：华大基因研究中心、中科院遗传与发育生物研究所等12个单位。

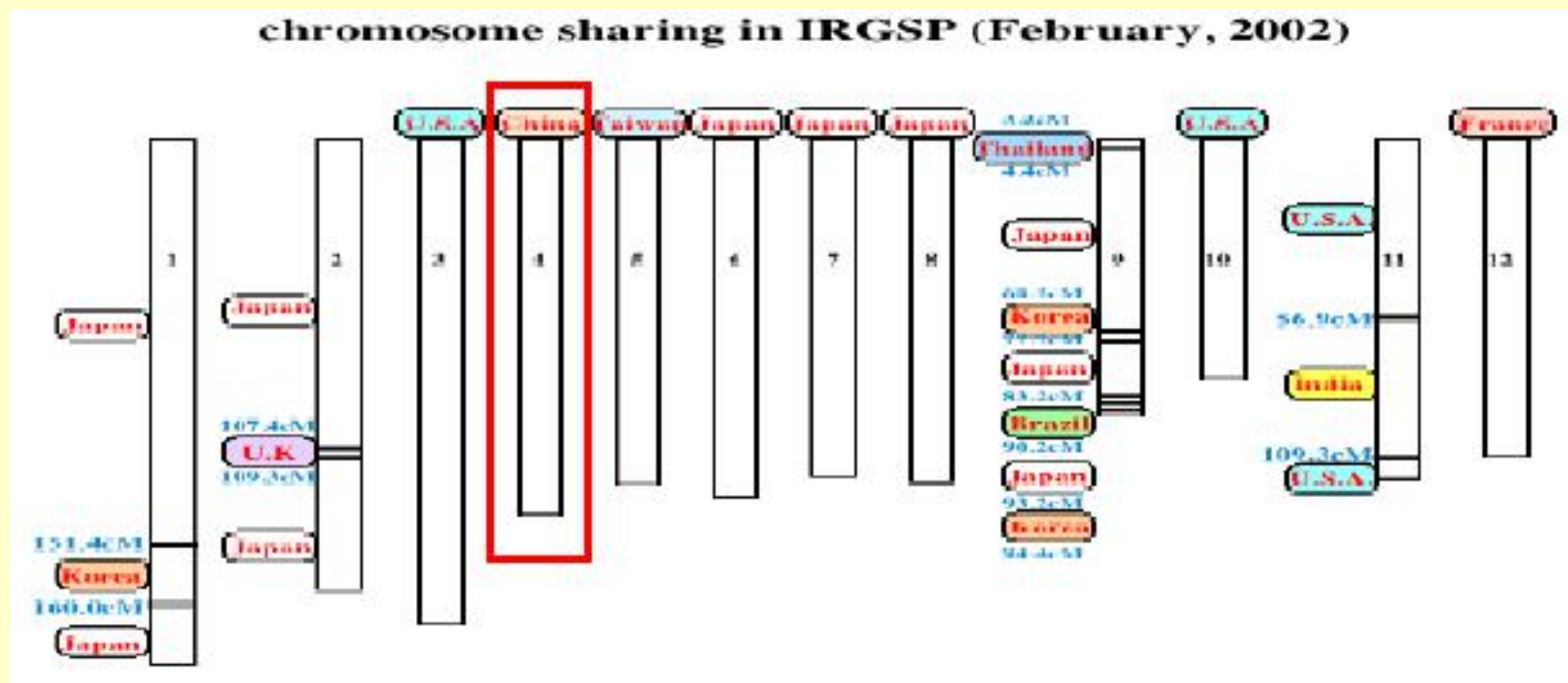
☆水平：水稻基因组的总基因数约为46022~55615个，工作框架图序列已覆盖水稻整个基因组92%以上的基因。

☆方法：“鸟枪射击法”，利用国产曙光2000、曙光3000超级计算机(1000亿次/秒)对随机DNA碎片进行排序和组装。

国际水稻基因组测序计划

高参考价值的真题、答案、学长笔记、辅导班课程，访问：www.kaoyancas.net

国际水稻（粳稻）基因组计划始于**1998**年，日本、美国、中国、法国等国家和地区参加。中国负责第**4号染色体：36 Mb (占9~10%)**。



2002年12月21日《Nature》，中国第四号染色体。

☆材料：粳稻“日本晴”。

☆完成单位：中科院国家基因研究中心等4家单位。

☆水平：第四号染色体中的总碱基数目为0.35亿碱基对，覆盖全长序列98%的区域，只剩下7个小空洞，碱基序列的精确度达到99.99%。完整测定的着丝粒序列在高等生物中属于首次。

三、生物体基因组特征

高参考价值的真题、答案、学长笔记、辅导班课程，访问：www.kaoyancas.net

(一) 人类基因组

◆核基因组DNA的总长约 3×10^9 bp，含有24条线性DNA分子，最长的有250 Mb，最短的55 Mb。30亿个碱基对。

◆线粒体基因组是长度为16569 bp的环状DNA分子，每个细胞平均含有800个线粒体，每个线粒体含10个基因组拷贝。

以每10cm 书写60个字母计算，30亿个碱基对连接的长度可达5 000 km，相当于北京到香港来回的距离。

1. 基因组小，测序容易
2. **DNA**绝大部分编码成蛋白质，重复序列少
3. 基因中一般没有内含子，结构比较简单

大肠杆菌基因组是双链环状DNA，全长 4.6×10^6 bp。含有4230个基因，编码蛋白的序列占基因组的87.7%，非编码的重复序列占0.7%，剩下的11.6%可能起调控作用。

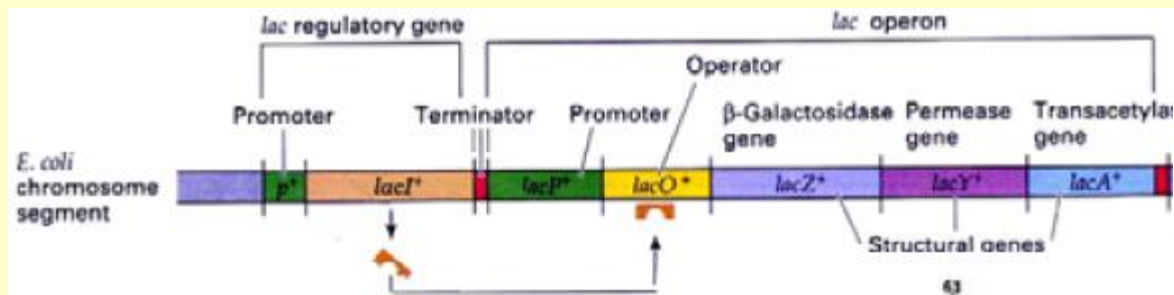


(二) 细菌和病毒基因组

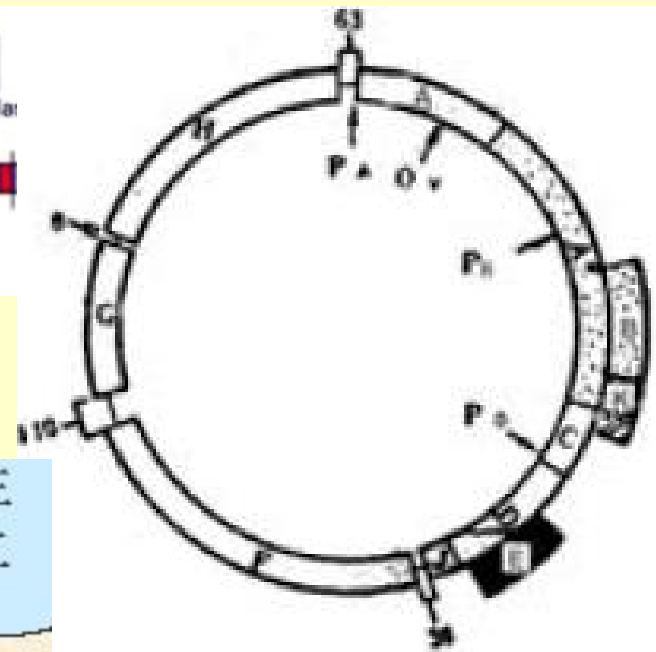
高参考价值的真题、答案、学长笔记、辅导班课程 访问：www.kaoyancas.net

4. 功能相关的几个基因排列在一起形成操纵子
如，乳糖操纵子结构

5. 存在重叠基因 如， Φ X174基因组为5386bp，
只有11个基因，其中有6个基因相互重叠。



Sanger因完成该生物体的测序工作，第二次获诺贝尔奖。

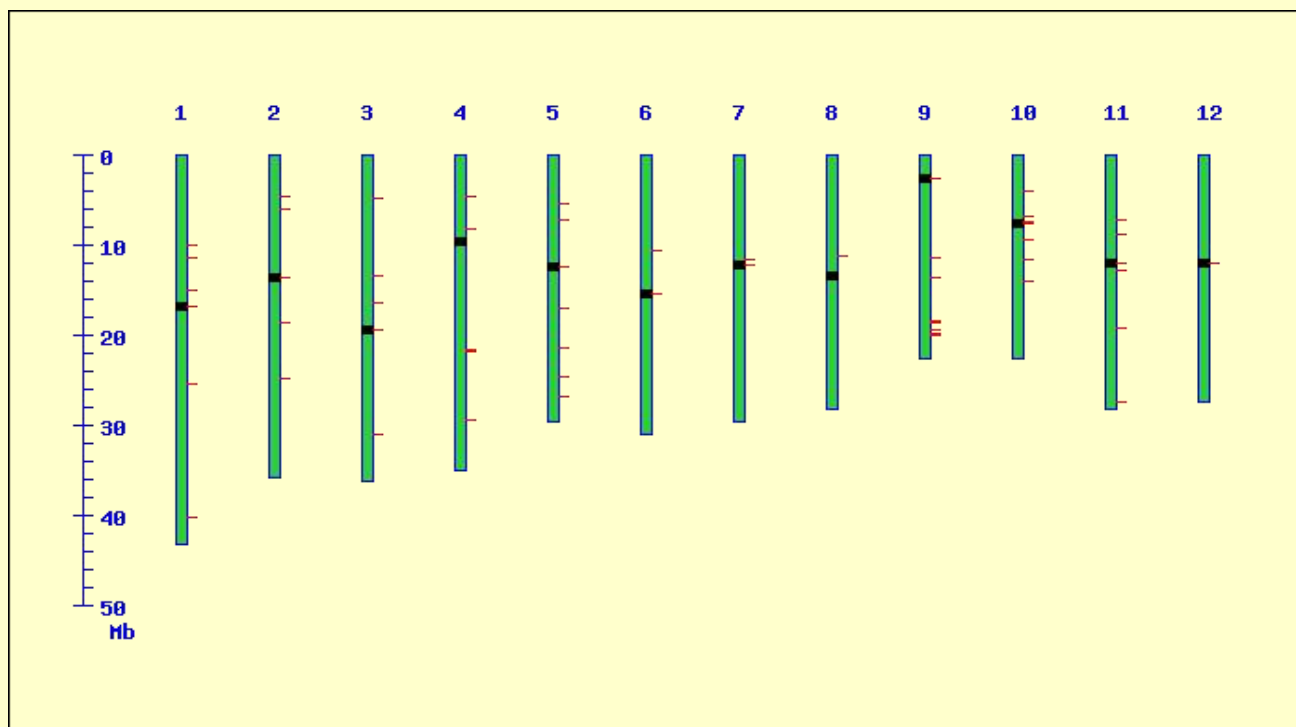


Φ X 174 的遗传图(数字表示不翻译的间隔区)

- 真核生物核基因组一般含有数目不等的线性**DNA**分子，而且**DNA**分子都与蛋白质结合形成染色体。
- 真核生物都具有环状的线粒体**DNA**，植物细胞还含有环状的叶绿体**DNA**。
- **DNA**编码部分只占基因组的少部分
大多真核生物基因组内编码部分只占基因组的1%-10%，非编码的**DNA**中存在着大量的重复序列(repeat sequence)。小基因组重复顺序较少，大基因组重复顺序急剧扩增。如玉米基因组比人类基因组大1000 Mb，其中绝大部分为重复顺序。

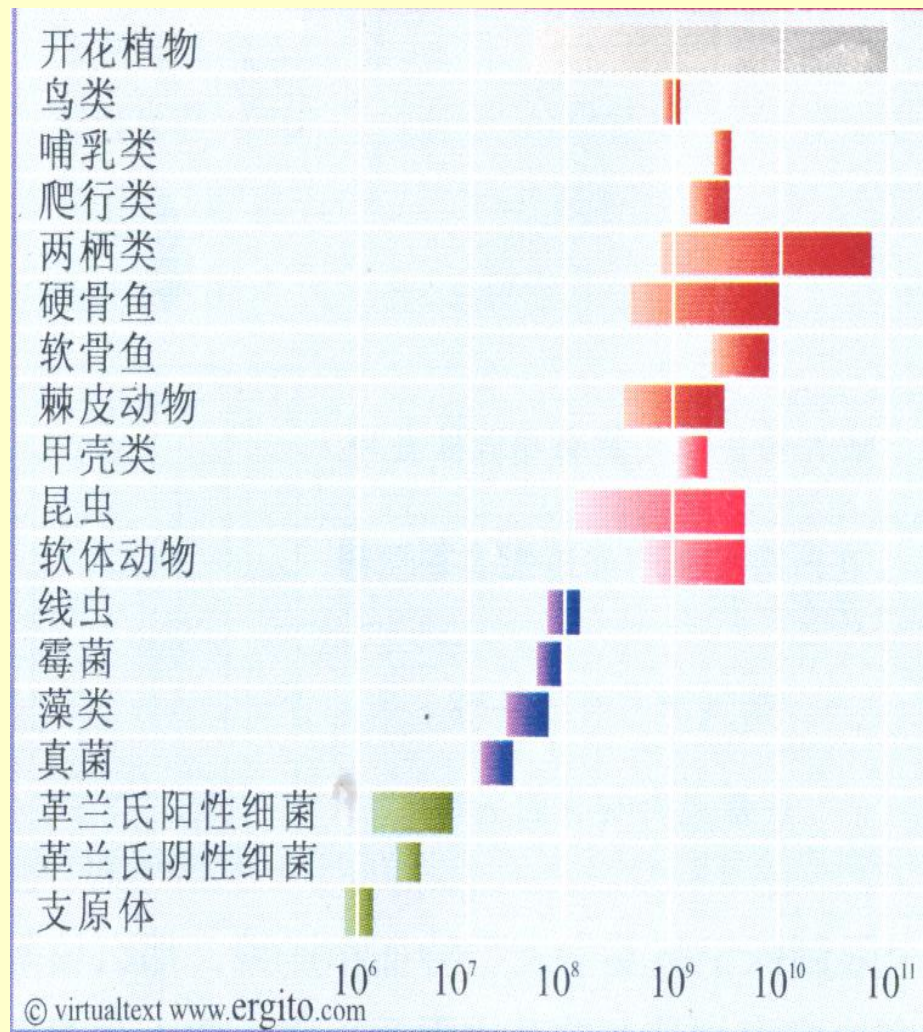
◆水稻是第一个完成基因组全序列测定的农作物，核基因组含有**12条染色体**，总长约**389Mb**，1号染色体最大为**43.2Mb**，10号染色体最小**22.6Mb**。全基因组预测约含有**4万个基因**。

◆水稻双链闭环线粒体基因组大小为**491kb**，叶绿体基因组**134.5kb**



C值：是指一个单倍体基因组中**DNA**的总量。

值悖理（C value paradox）：物种的C值和它的进化复杂性之间无严格对应关系的现象称为C值悖理，是复杂生物基因组的一个普遍特征



➤ **N值**：是指生物体所含有的基因数目。

➤ **N值悖理 (N value paradox)**：复杂性不同的生物种属所具有的基因数目与其生物结构的复杂性不成比例的现象。

如结构比较简单的线虫含有的基因数为1.9万个，比线虫更复杂的果蝇基因数为1.8万个，水稻的基因数约4万个，最复杂的人类其基因总数约3万个

（一）结构基因组学(structural genomics)

➤通过基因作图、核苷酸序列分析确定基因组成、进行基因定位的科学。

➤遗传信息在染色体上,但染色体不能直接用来测序,必须将基因组这一巨大的研究对象进行分解,使之成为较易操作的小的结构区域,这个过程就是基因作图。完成基因组图谱构建之后,就可以利用图谱进行基因组序列测定和组装。

(二) 功能基因组学(functional genomics)

◆利用结构基因组所提供的信息和产物，研究基因组功能表达的一门分支学科。

◆主要内容：

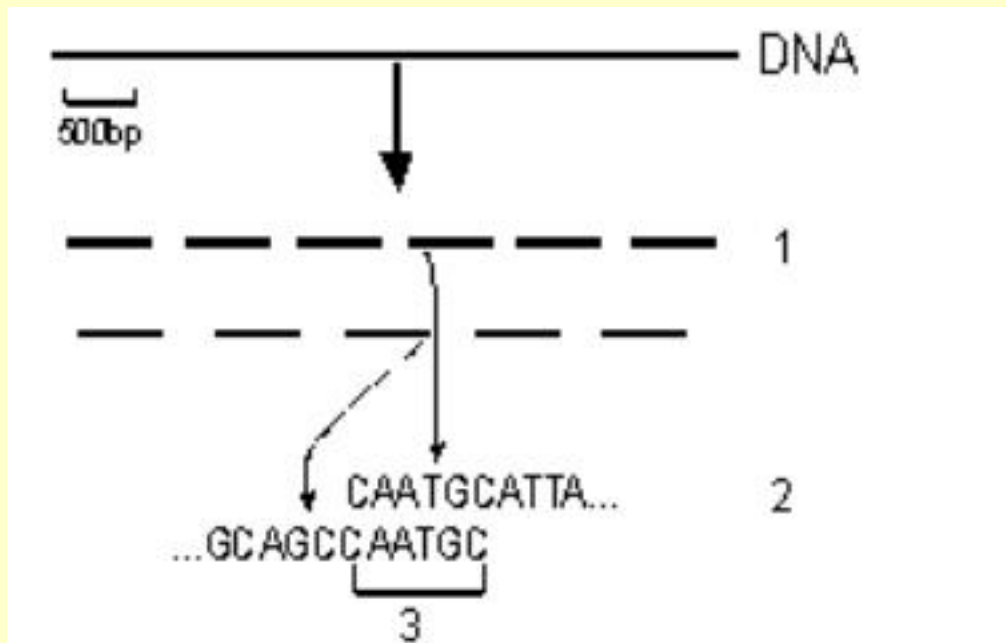
- 基因的识别、鉴定和克隆。包括新策略、新技术、新方法的创立和各种基因组数据的建立；
- 基因结构与功能及其相互关系的研究。包括基因变异体的系统鉴定和目录的绘制；基因表达谱的编制、基因结构与功能关系的鉴定、基因相互作用网络图的编制；
- 基因表达调控的研究

（三）蛋白质组学(proteomics)

- 研究细胞内蛋白质组成及其活动规律。旨在阐明生物体全部蛋白质的表达模式及功能模式,内容包括鉴定蛋白质表达、存在方式、结构、功能和相互作用方式等。
- 基因是遗传信息的携带者,而全部生物功能的执行者却是蛋白质,仅仅从基因的角度来研究是远远不够的。

第2节 基因组图谱绘制

基因组计划的目的是获得全基因组序列，并对其进行解读。**DNA**测序每次反应仅能读取**1000bp**的长度，因此，基因组测序的基础是基因组图谱的构建。



鸟枪射击法
(shotgun) 基因
组序列测定

图 9—25 鸟枪射击法序列测定及装配

1、待测序的 DNA 片段；2、测得的 DNA

序列。完整版，请访问www.kaoyancas.net 科大科院考研网，专注于中科大、中科院考研

基因组测序策略

☆**重叠群法** 相互存在重叠序列的一组克隆。根据重叠群的相对位置讲各个克隆首尾相连，长度可达百万级bp。对单个重叠群，采用鸟枪法测序，然后进行组装。这是由上而下（**up to down**）的测序策略。

☆**直接鸟枪法** 首先进行全基因组鸟枪法测序，再用分子标记为起点强鸟枪DNA片段组装。这是由下而上（**bottom to up**）的测序策略。这种方法依赖于高密度分子标记基因组图谱。

基因组图谱分为遗传图谱和物理图谱。

一、遗传图谱绘制

(一) 遗传标记

(二) 遗传图谱的构建

(一) 遗传标记

高参考价值的真题、答案、学长笔记、辅导班课程，访问：www.kaoyancas.net

◆遗传标记就是遗传物质的特殊的易于识别的多态性表现形式，它包括形态标记、细胞学标记、生化标记和分子标记。

➤形态标记：主要指可以观察到的一些性状,如种皮颜色、眼色、株高等。

➤细胞学标记：细胞学标记是指能明确显示遗传多态性的细胞学特征。

➤生化标记：主要是同工酶及种子贮藏蛋白，有时又称蛋白质标记。

➤分子标记：主要指DNA水平上的标记。

DNA 标记

☆以DNA为基础的分子标记主要包括

◆基于杂交的分子标记，如RFLP。

◆基于PCR的分子标记，如RAPD、AFLP、SSR (又称microsatellite)、AFLP 等。

◆基于DNA序列和芯片的分子标记，如SNP (single nucleotide polymorphism)。

☆ RFLP技术是基于Southern杂交的分子标记的方法。它是指用限制性内切酶酶切不同个体基因组DNA后，酶切片段长度的差异。

☆ 这一技术用放射性同位素标记DNA片段作为同源序列探针（RFLP标记），与经限制酶消化并转移到支持膜上的基因组总DNA杂交，通过放射性自显影（或非同位素技术）来显示酶切片段的大小，检测不同遗传位点的多态性

RAPD由Williams等（1990）和Welsh等（1990）分别发展起来的分子标记技术。这一技术是以基因组DNA为模板，采用随机设计的单个寡核苷酸序列（一般为10bp）为引物，通过PCR扩增，产生不连续的DNA产物，用于检测DNA序列的多态性。

SSR (simple sequence repeats)
或微卫星 (microsatellite)

高参考价值的真题、答案、学长笔记 辅导班课程 访问: www.kaoyancas.net

☆ 重复序列

◆ 串联重复序列 (tandem repeated sequence)，其重复单位首尾相连，成串排列 (Flavell 1986)。

◆ 散布重复序列 (interspersed repeated sequence)，其重复单位与其它无关序列或单拷贝序列相间排列。

☆微卫星DNA序列或SSR又称短串联重复序列（short sequence repeat, STR），它是由几个核苷酸（一般1~6个）为重复单位簇集而成的串联重复序列，可随机的分布在整个基因组的不同位置上。微卫星长度具有高度变异性，并且这种多态性常常表现复等位性，两端的序列多是相对保守的单拷贝序列，因而可以根据两端的序列设计一对特异引物，扩增每个位点的微卫星序列，从而揭示其长度的多态性（simple sequence length polymorphism, SSLP）。

ISSR是一种新型的分子标记。与**SSR**相反，直接用同位素标记**SSR**序列，扩增**2**个**SSR**间的单拷贝序列。为了增加扩增的特异性，在引物的**5'**和**3'**端分别加入**1~2**个选择性碱基，引物长度**16~18bp**。

AFLP (***Amplicon fragment length polymorphism***)

☆ AFLP结合了RFLP和RAPD技术的优点。AFLP的基本原理是基于PCR的扩增基因组DNA限制性片段多态性。基因组DNA先用限制性内切酶切割，然后将双链接头（adapter）连接到DNA片段的末端，通过选择在3'端分别添加1~3个选择性碱基的不同引物，选择性地识别具有特异配对顺序的酶切片段并与之结合，从而实现特异扩增。

☆引物主要有3部分组成：核心序列、酶特异序列和选择性延伸序列，如*EcoRI*和*MseI*引物（N表示任何碱基）

	核心序列	酶特异序列	选择性延伸序列
<i>EcoRI</i>	5-GACTGCGTACC	AATTC	NNN-3
<i>MseI</i>	5-GATGAGTCCTGAG	TAA	NNN-3

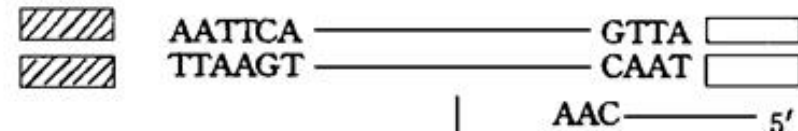
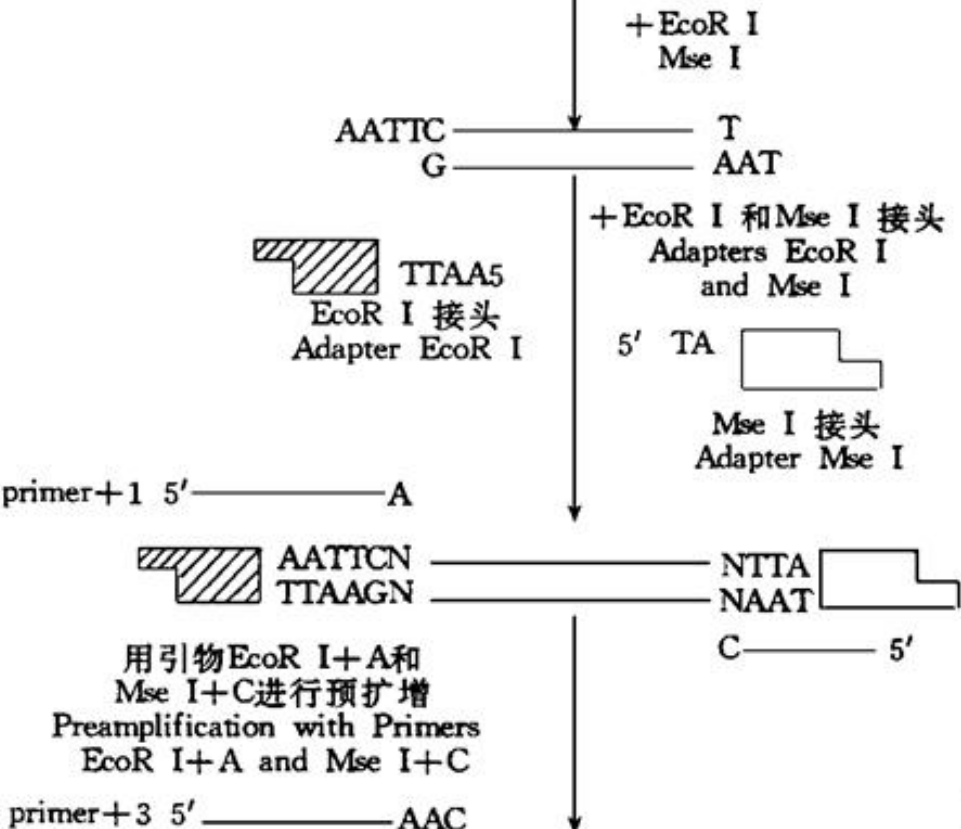
☆接头由2部分组成：核心序列和酶特异序列，如*EcoR* /接头

<i>EcoRI-adapter</i>	5-CTGGTAGACTGGGTACC CATCTGACGCATGGTTAA-5
<i>MseI-adapter</i>	5-GACGATGAGTCCTGAG TACTCAGGACTCAT-5

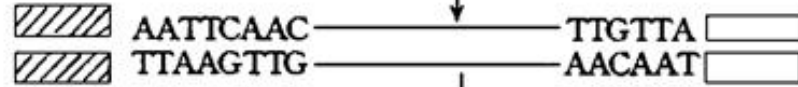
5' ————— GAATTC ————— TTAA — 3'
 3' ————— (高参考价值的真题、答案、学长笔记) — 5'

辅导班课程，访问：www.kaoyancas.net

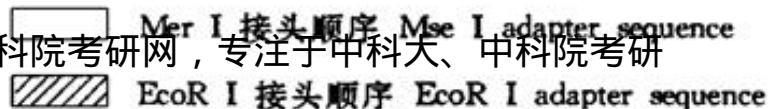
AFLP反应过程示意图



用带有3个选择性核苷酸的引物进行
 选择性扩增
 Selective amplification with primer+3



变性聚丙烯酰胺凝胶电泳
 Denatured polyacrylamide gel electrophoresis



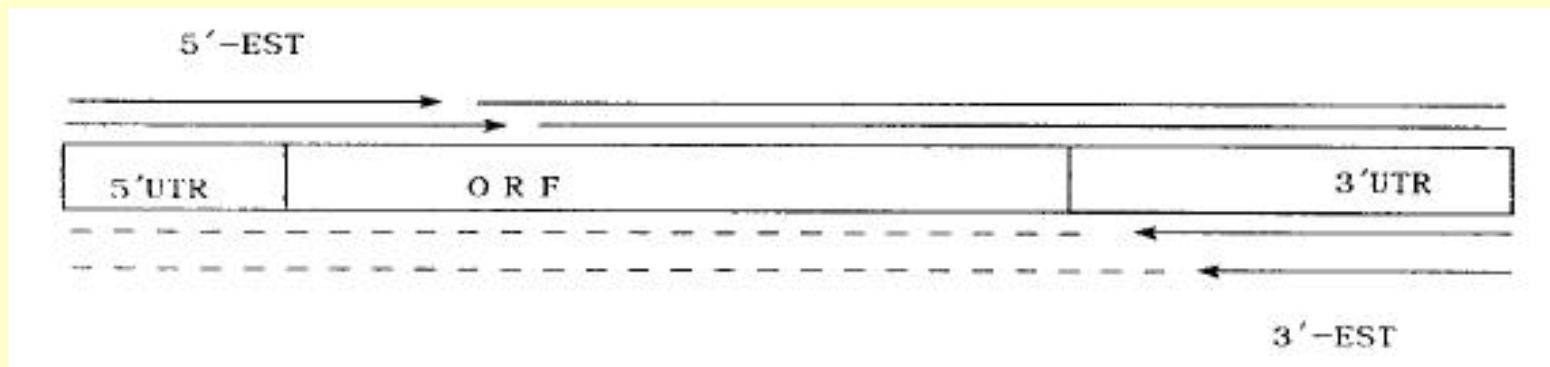
☆基因组大的生物，增加末端选择性核苷酸的数目，可以减少扩增的多态性片段数。每增加一个核苷酸，扩增片段数目就减少4倍，当引物有4个选择性核苷酸时，易引起错配。

☆对于复杂基因组DNA，AFLP扩增一般需要分2步进行，第1步预扩，AFLP引物只选择1个选择性核苷酸，扩增后稀释；第2步——选扩模板DNA，第2次扩增引物有3个选择性核苷酸。

☆ AFLP既能检测酶切位点不同造成的多态性，又能检测随机选择碱基造成的多态性，其多态性检测能力远高于RFLP。

☆遗传信息由DNA → mRNA → 蛋白质。

☆一个典型的真核生物mRNA分子：5' - UTR (5'端转录非翻译区)，ORF (开放阅读框架)，3' - UTR (3'端转录非翻译区)，polyA



☆任何一个基因，cDNA 的5'端或3'端的有限序列即可特异性地代表生物体某种组织某个时期的一个表达基因。EST 的数目可以显示所代表的基因的拷贝数

☆从组织细胞中提取总mRNA，构建成标准cDNA文库，然后从中挑取大量克隆，利用载体通用引物测出插入载体的cDNA片段5'端或3'端300 - 500 碱基的序列。

☆将测序所得的EST与dbEST等数据库中的数据进行比较分析,根据核酸或蛋白质序列的同源性比较，可以鉴定出哪些EST代表已知基因，哪些EST代表未知基因。

☆序列标签位点（**sequence tagged site**）是一小段**DNA**序列。每个基因组仅**1**个拷贝，很容易分辨。**STS**要满足**2**个条件：

◆是一段已知的序列，可据此涉及PCR引物来检测不同DNA片断中是否存在 这一序列。

◆ STS在染色体上必须是独一无二的。如果在基因组中有多个位点出现，作图数据将含混不清。

☆常见的寻找**STS**的方法：

EST、 SSLP、 随机基因组序列

☆单核苷酸多态性是指基因组DNA序列中由于单个核苷酸(A, T, C, G)的替换而引起的多态性。通常SNPs不包括碱基的插入、缺失以及重复序列拷贝数的变化。这种标记只有两种等位基因。

人类基因组的编码基因中有20万个SNPs，在非编码区的数目可能还要多10倍以上。

☆单倍型：当前常用术语“haplotype”(单倍型)代替术语“allele”(等位基因)。在给定的一条染色体的紧密连锁的位点上多个等位基因的集合,通常3~4个相邻等位基因彼此靠近而构成的单倍型可作为一个整体而遗传(称为单倍型块(haploblock))

特性	RFLP	RAPD	SSR	ISSR	AFLP
分布	普遍存在	普遍存在	普遍存在	普遍存在	普遍存在
遗传	共显性	多数显性	共显性	多数显性	多数显性
多态性	中	高	高	高	非常高
等位检测	是	不是	是	不是	不是
检测位点数	1~3	1~10	1~5	0~50~更多	20~100
样品信息量	低~中	高	高	高	非常高
基因组区域	底拷贝编码	整个基因组	整个基因组	整个基因组	整个基因组
技术难度	中等	简单	简单	简单	中等
重复性	高	中等	高	高	高
DNA样品量	2~30μg	1~100ng	50-100ng	2~50ng	100ng
反射线	一般是	不是	不是	不是	一般是
耗费时间	慢	快	快	快	中等
可靠性	高	中等	高	高	高

(二) 遗传图谱的构建

高参考价值的真题、答案、学长笔记、辅导班课程，访问：www.kaoyancas.net

1 人类基因组遗传图谱的构建

◆人类的遗传图谱是利用家系分析法，在对8个家系的134个成员的分析中(186个减数分裂)，主要根据5264个STR标记绘制而成的。

◆利用这些家系的资料绘制第1至22号染色体图谱。对于X染色体图谱，还利用了来自另外12个家系，170个成员(105个减数分裂)的资料绘制而成。

◆最后，将5264个标记定位在2335个位点（其中有些标记相距很近而作为作为一个位点）。

◆作图群体

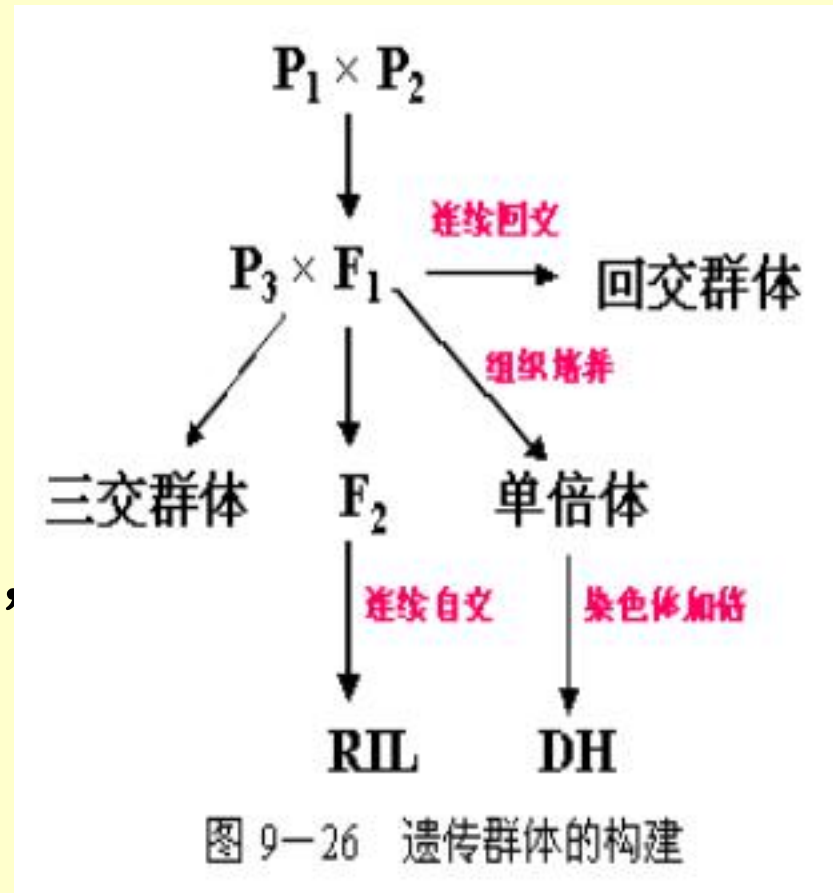
常用的遗传作图群体有 F_2 群体、回交群体、加倍单倍体

（double haploid, DH）群体、重组近交系（recombinant inbred lines, RIL）群体、近等基因系（near—isogenic lines, NIL）群体等（徐云碧，1994）。

◆遗传标记的染色体定位

◆标记间的连锁分析

LINKAGE、Mapmaker、JoinMap



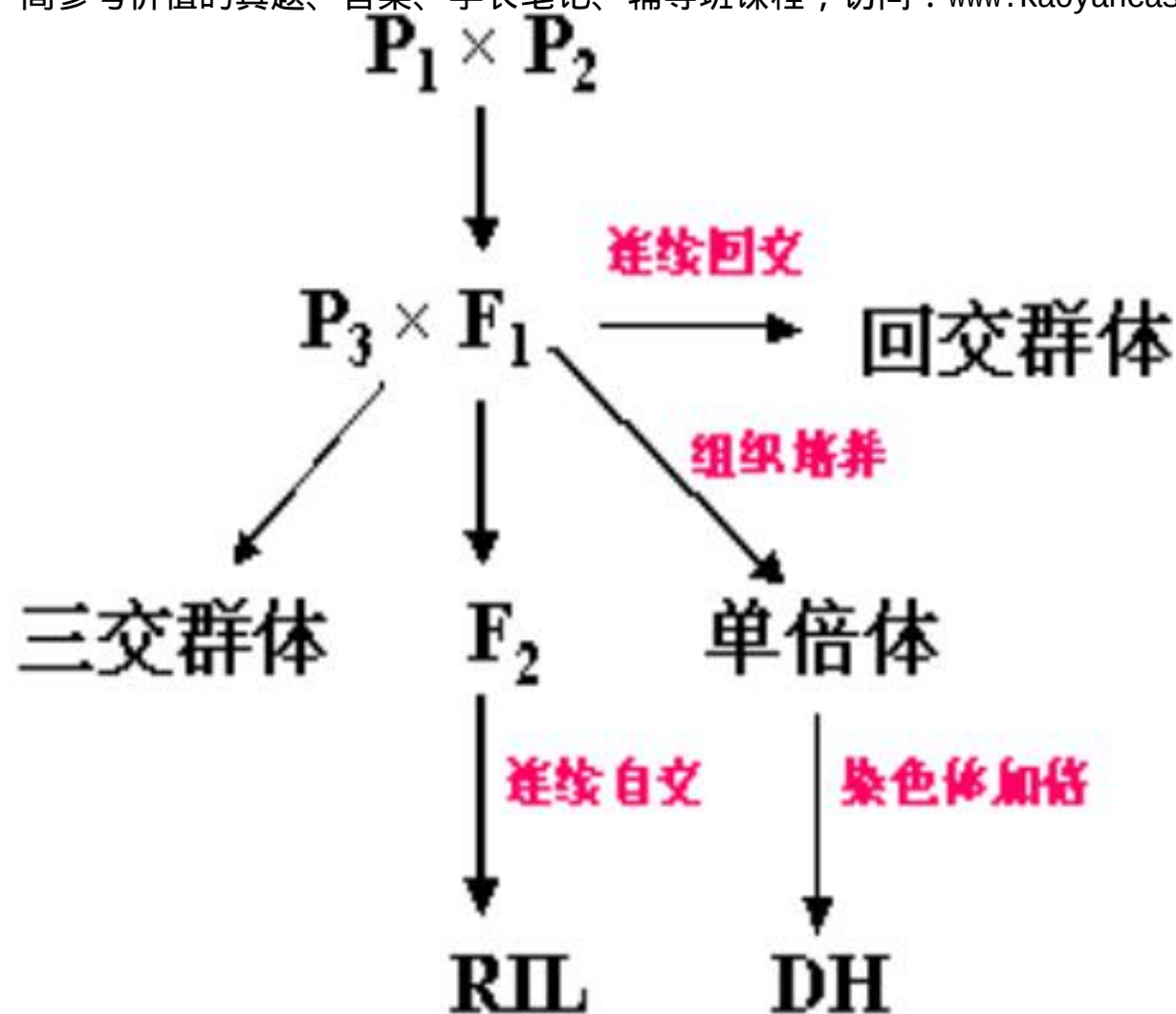


图 9—26 遗传群体的构建

二、物理图谱绘制

- (一) 限制性作图 (参见基因工程)
- (二) 基于克隆的基因组作图
- (三) 原位杂交
- (四) 序列标签位点 (STS) 作图
- (五) 人类基因组图谱

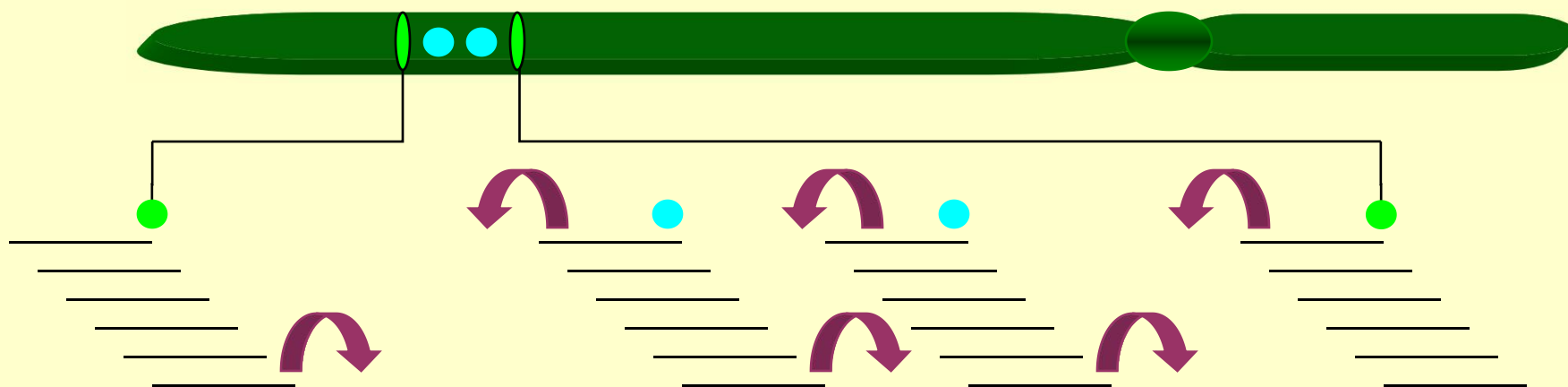
(二) 基于克隆的基因组作图

高参考价值的真题、答案、学长笔记、辅导班课程，访问：www.kaoyancas.net

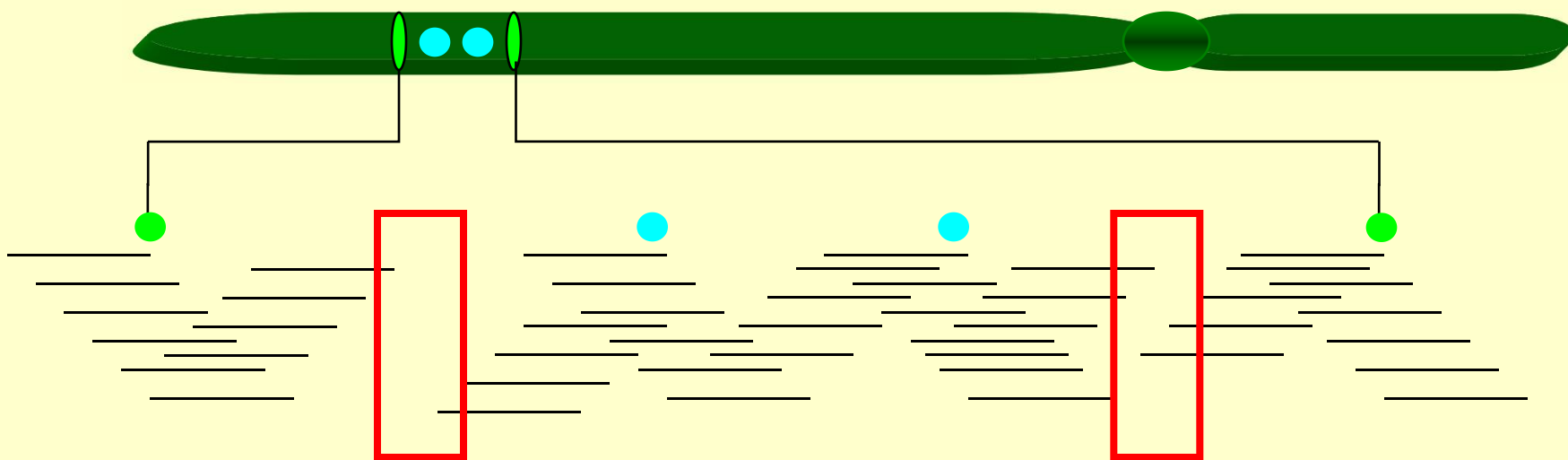
重叠群 (*contiguous DNA clones, contigs*)

- ◆ 从1个感兴趣的位置开始，利用第1个元件的末端部分来辨别第2个元件，沿染色体“行走” (walk)。通过鉴别目标位点两测的2个DNA标记，从1个标记向另1个标记的行走。
- ◆ 沿染色体鉴别一系列重叠群是大规模研究的基础。在特定区域的染色体行走可以提供分离通过遗传图谱定位在该区域基因的方法。集中全部染色体的重叠群，可以为以后研究提供有效克隆来源。
- ◆ 染色体行走考虑的根本是每“步”的大小，较大的步加快积聚相邻克隆的进程。

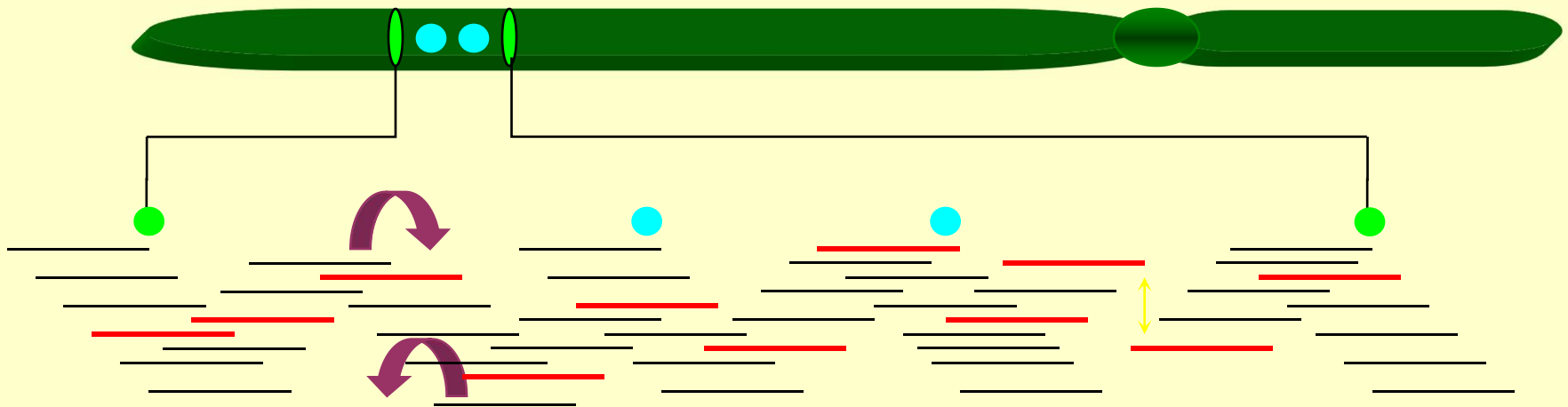
区域作图 *Regional mapping*



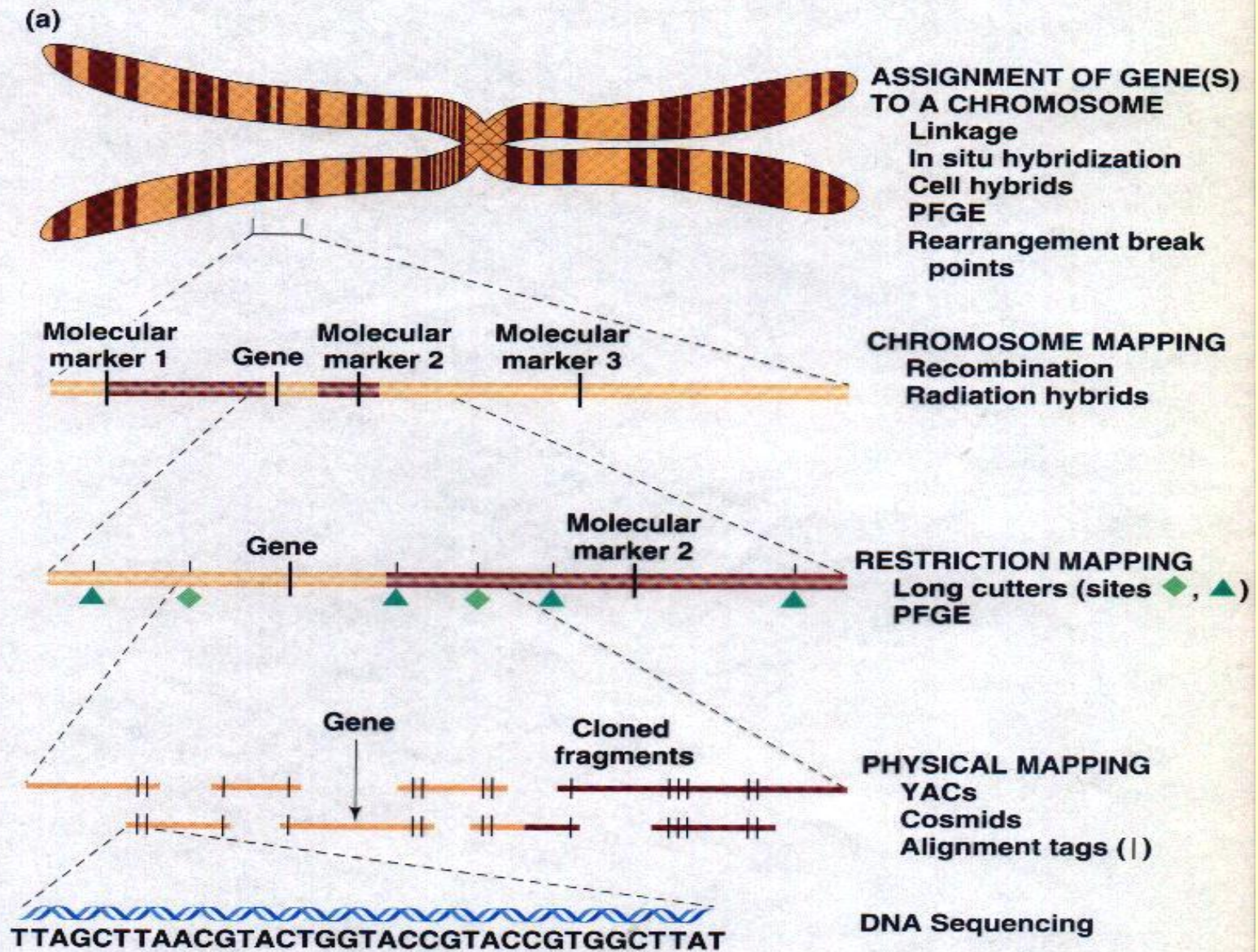
区域作图 *Regional mapping*

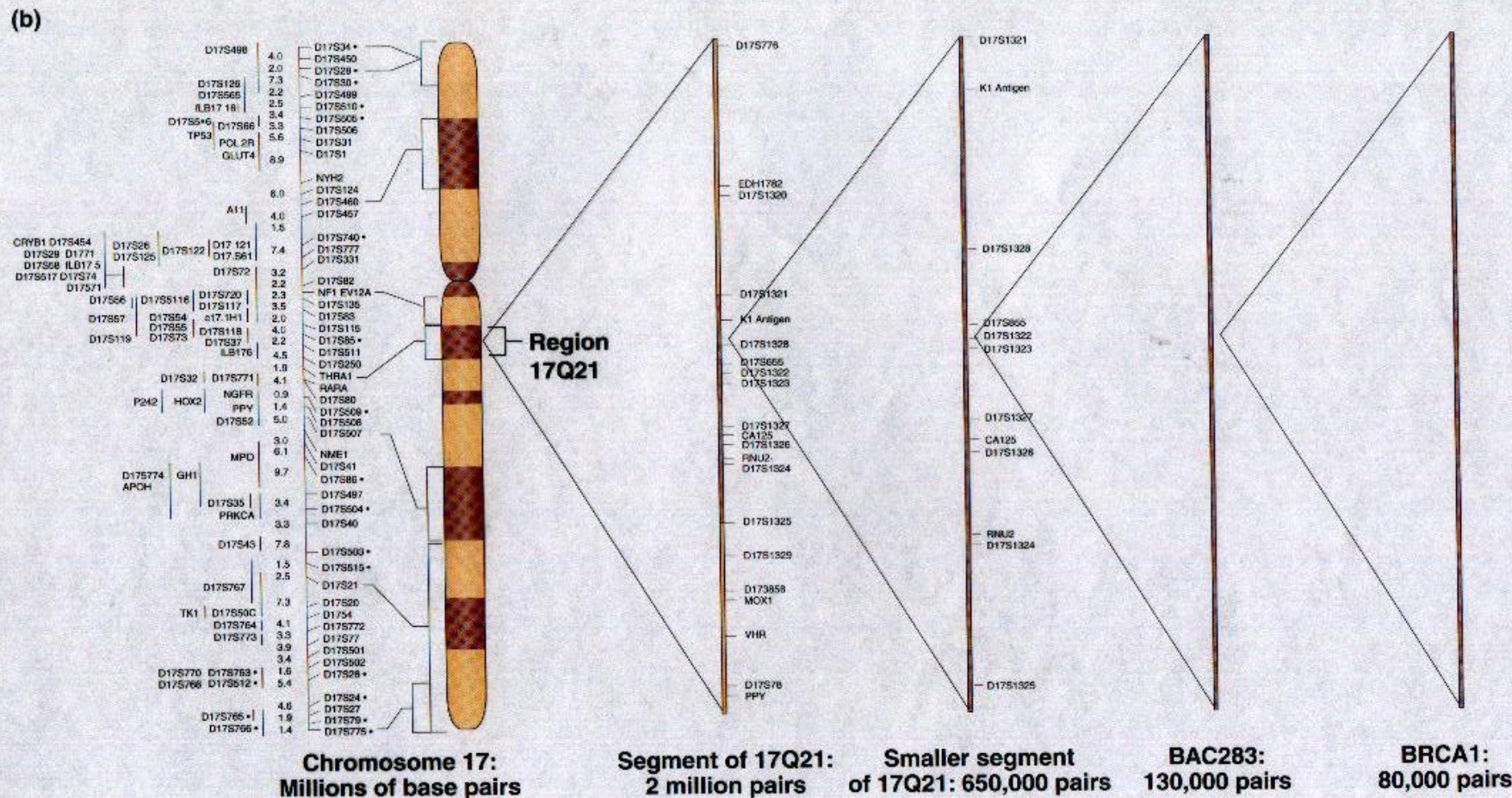


区域作图 *Regional mapping*



Minimal tiling path selected for sequencing.





(三) 原位杂交

高参考价值的真题、答案、学长笔记、辅导班课程

☆ 荧光原位杂交 (fluorescent in situ hybridization, FISH)

☆ 基因组原位杂交 (genome in situ hybridization, GISH)

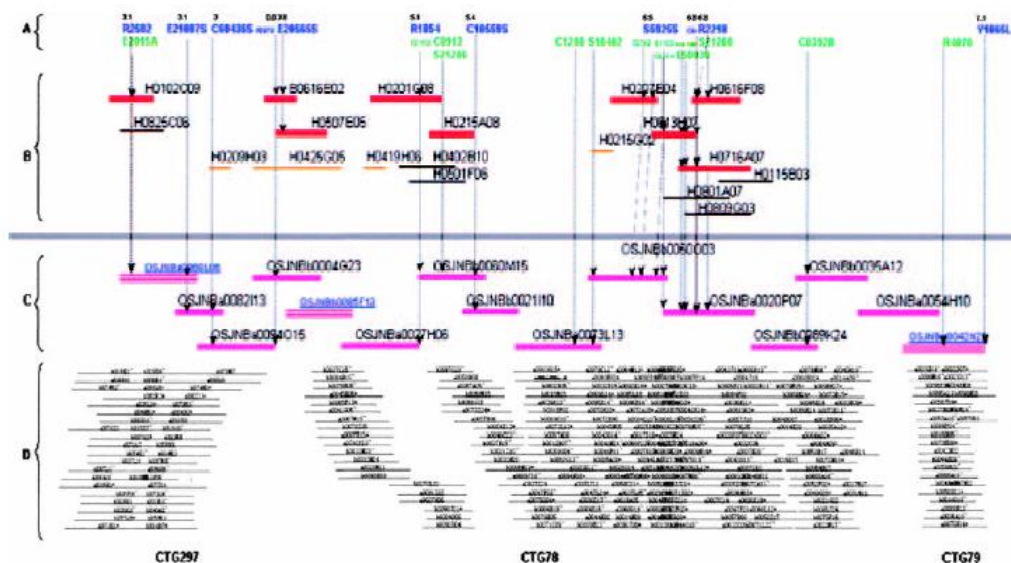


Figure 2 The BAC contig map of CTG4-1 near the short arm telomere together with its collinear Guangluai 4 region. (A) Markers associated with the map. Fifteen EST (in green) and 11 RFLP (in blue) markers used to anchor contigs are shown together with their genetic positions (numbers in cM). (B) Anchored Guangluai 4 BAC contigs by the markers. Fully sequenced BACs (red lines), BACs identified using the Nipponbare sequence scaffold against the Guangluai 4 BAC ends (orange lines), and other clones (black lines) in the contigs are indicated. (C) Fifteen tiled and completed sequenced Nipponbare BACs. Three bridging clones OSJNBa0094015 and OSJNBb0004G23 for joining CUGI contigs 297 and 78 and OSJNBa0054H10 for CUGI contigs 78 and 79 are shown. (D) Display of three original fingerprinted CUGI contigs 297, 78, and 79. CTG4-1 is composed of 295 fingerprinted CUGI BACs and three extended bridge clones. Contig 79 was anchored by marker Y1065L by in silico anchoring. Contigs 297 and 78 were anchored by the Guangluai 4 sequence scaffolds as indicated in syntetic positions.

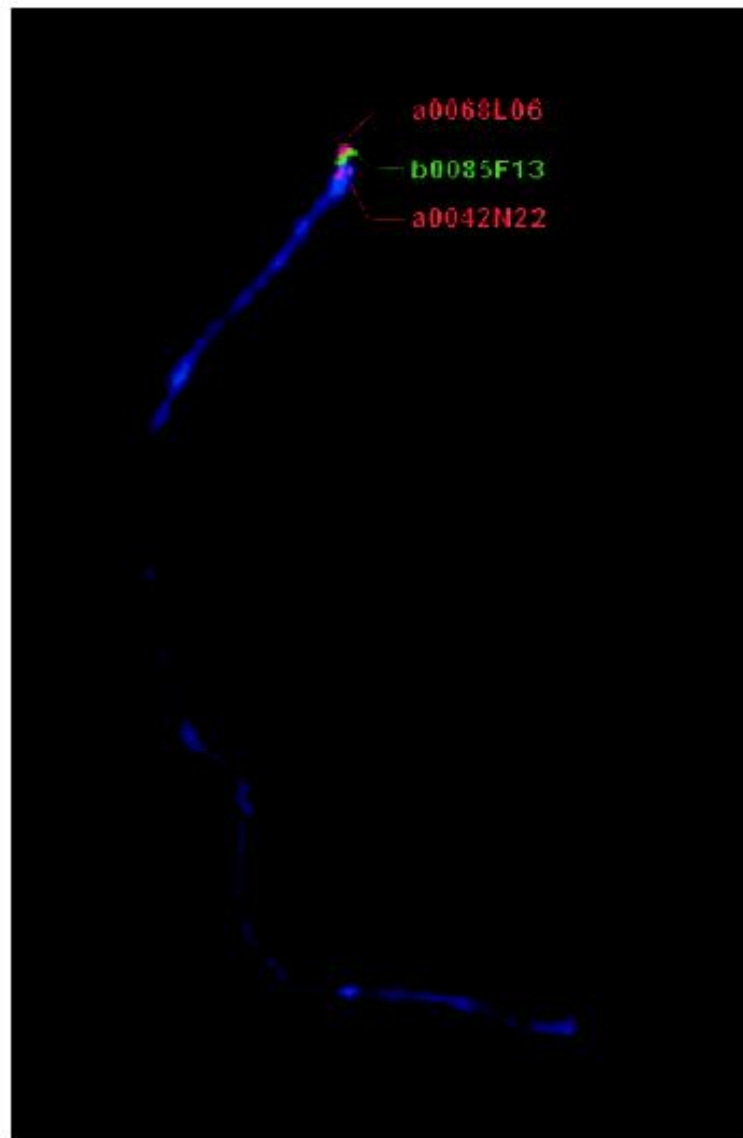


Figure 4 FISH analysis of CTG4-1. Three BAC clones, OSJNBa0068L06 (red signal), OSJNBb0085F13 (green signal), and OSJNBa0042N22 (pink signal), within CTG4-1 were chosen for FISH analysis to confirm the contig localization and orientation. RFLP markers R2502 (3.1 cM) and R2373 (0 cM) matched OSJNBa0068L06 and OSJNBb0004G23, respectively. Because OSJNBb0004G23 and OSJNBb0085F13 are overlapped, the order of the two markers was confirmed by the physical map and FISH analysis. Therefore, their genetic order is reversed on the physical map.

☆ **辐射杂交系** 是含有另一种生物染色体片断的啮齿类细胞。带有人类染色体片段的辐射杂交系

☆ **DNA库 YAC/BAC**

☆ **克隆作图** 获得大分子**DNA**克隆文库以后，用**PCR**的方法检测**STS**，根据重叠的**STS**标记绘制克隆连锁图。

➤当两个片段含有同一**STS**顺序时，则这两个片段彼此重叠。如果它们彼此邻接，这两个**STS**总会同时出现在相同片段上。如果它们相距甚远，有时会在同一片段，有时则在不同片段。

➤要将一组**STS**作图定位，必需收集来自同一染色体或整个基因组随机断裂的**DNA**片段。不同**DNA**片段之间有各种可能的重叠，可以覆盖整个作图区段。依次采用单个**STS**挑出它们所在的**DNA**片段，根据它们彼此的重叠关系可以逐段绘**DNA**物理图。

遗传图

- **1987**，**RFLP**连锁图，含**393**个**RFLP**和**10**个其它多态性标记。平均密度为**10 Mb**。
- **1994**年发表的基因组图含有**5800**个标记，其中包括**4000**多个**SSR**，密度达到每**0.7 Mb**一个标记。遗传图最后的一个版本发表于**1996**年，在**1994**年的基础上增加了**1250**个**SSR**。。

物理图

- 90年代初，33000个YAC组成的全基因组重叠群物理图。
- 1995年，STS图含有15 086个STS，密度为199 kb。1996年又添加了2 0104个STS，其中大多数是EST。密度为每100 kb一个标记。
- 包括7000个多态性SSR的物理位置，这些SSR标记已定位于遗传图上。由于同一STS标记分别定位在遗传图和物理图上，可以直接使克隆重叠图与遗传图彼此衔接，产生完全整合的基因组图，成为人类基因组计划DNA测序阶段的工作框架。

(五) 人类基因组图谱

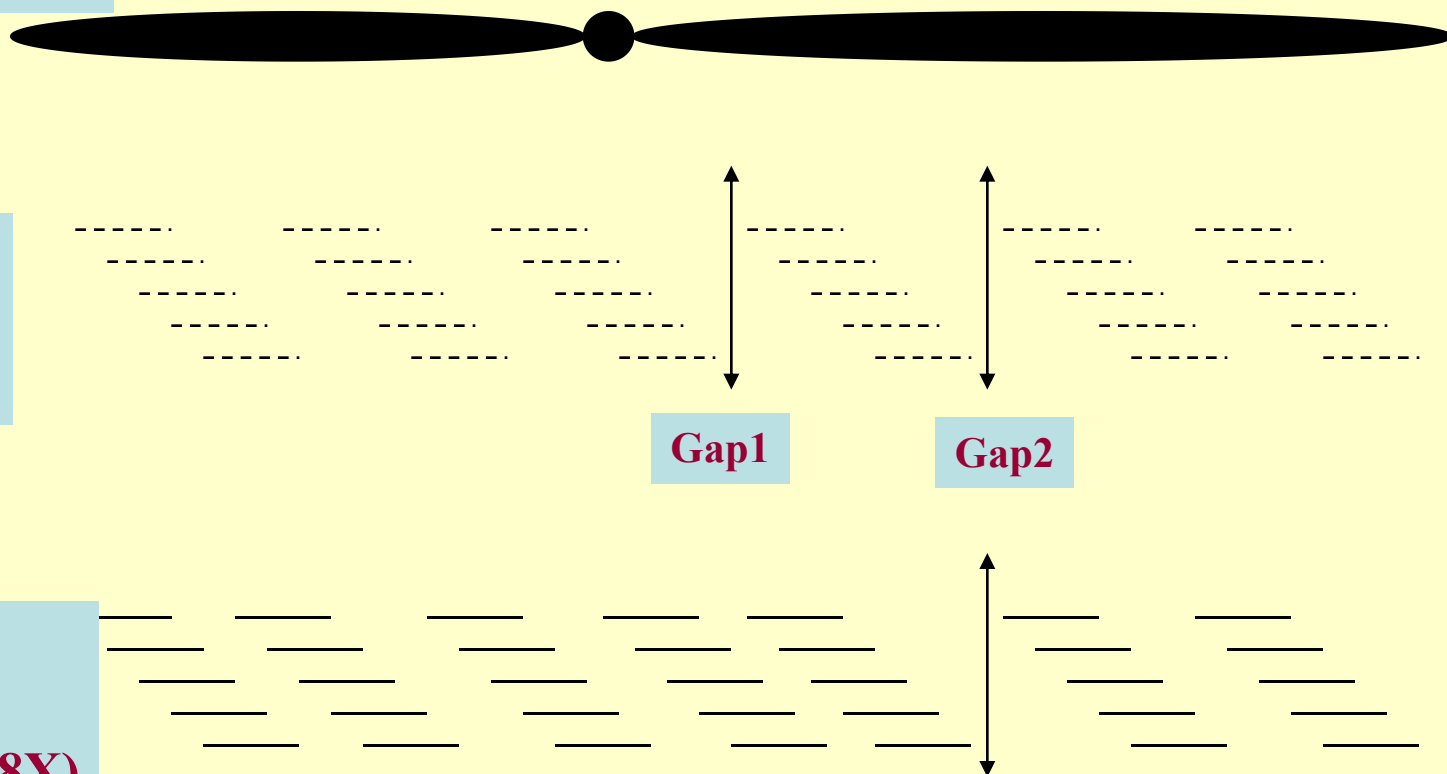
高参考价值的真题、答案、学长笔记、辅导班课程，访问：www.kaoyancas.net

人类基因组工作草稿（框架图）与完成图

Chromosome

“Working Draft”
(90%; 4X)

Finished Genome
(99.99%; 8X)



(五) 人类基因组图谱

高参考价值的真题、答案、学长笔记、辅导班课程，访问：www.kaoyancas.net

Contig12994 of Humanmap. Clones 61 of 192. Markers 17 of 17. Sequenced 18. Length 124

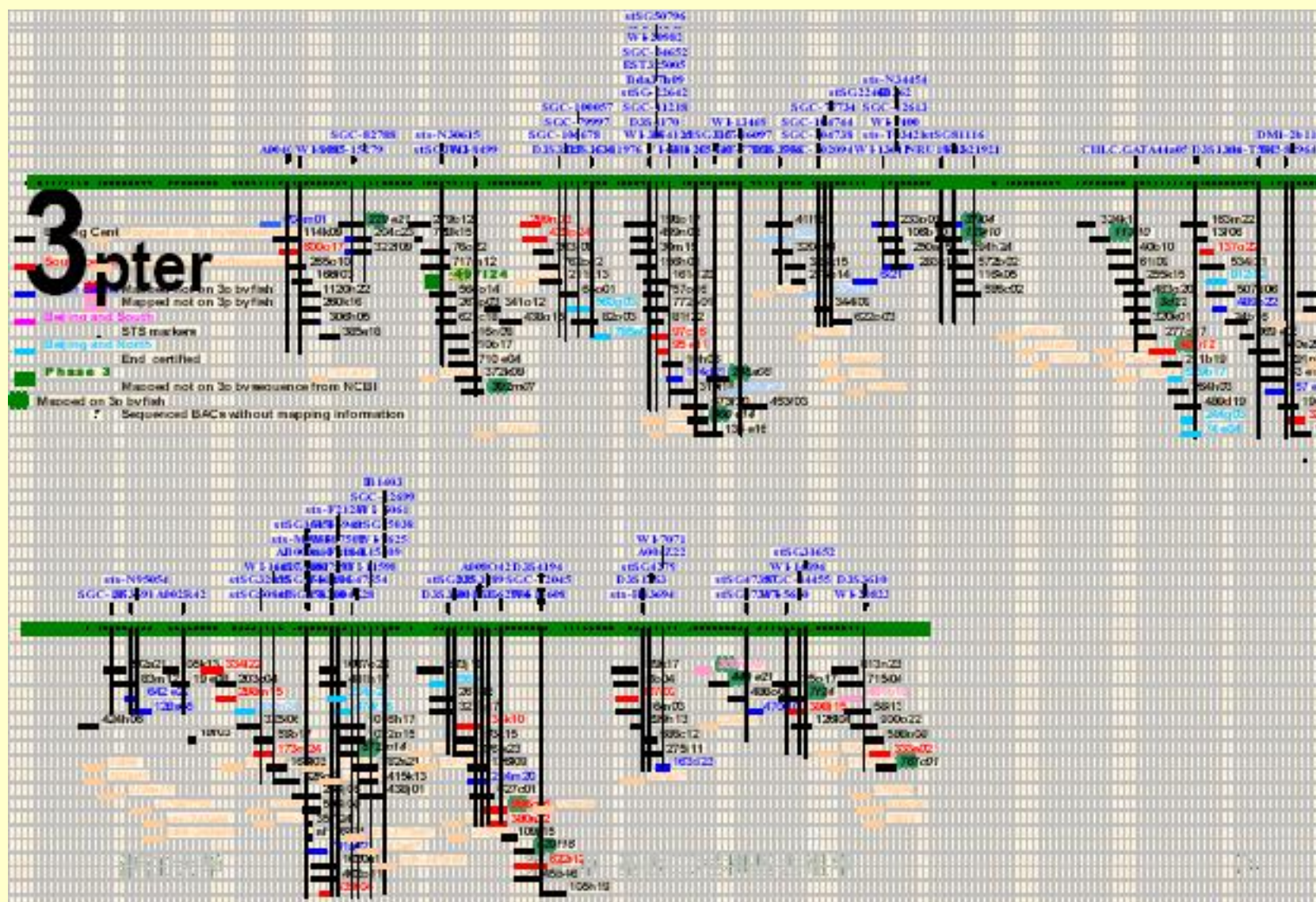
11-SGC35198	3-st5662575	3-4I-14870	3-SGC32064	3-A005P03
11-sts-66010085			3-SGC-32064	3-A006542
11-st5654043			3-AFM238wb12	3-sts-T79204
11-4I-22370			1-A009J41	3-st5650515
			1-st568010	3-st5662549
			1-st5612651	
	<u>H0367F20*</u>		<u>H0304C15*</u>	
<u>H0064023</u>	<u>H0386M06</u>	<u>H0192H02</u>	<u>H0824H09</u>	<u>H0473C13*</u>
<u>H091B922</u>		<u>H0434E03*</u>		<u>H0384922</u>
<u>H0639F01*</u>	<u>H0638H18*</u>	<u>H0474F22*</u>	<u>H0697I06*</u>	<u>H0708J04</u>
<u>H0654H07*</u>	<u>H0384D17</u>	<u>H0504K21*</u>	<u>H0627H03*</u>	<u>H0801I08*</u>
<u>H0483G17*</u>	<u>H0163G22*</u>	<u>H0712G15</u>	<u>H0753B01*</u>	<u>H0392922*</u>
<u>H0793K13*</u>	<u>H0477J14*</u>	<u>H0217E01*</u>	<u>H0710K15*</u>	<u>H0134A15*</u>
<u>H0180N16</u>	<u>H0224P21</u>	<u>H0294L13*</u>	<u>H0305K05</u>	<u>H0201I23*</u>
<u>H02015H18*</u>	<u>H0683L08*</u>	<u>H0712C11*</u>	<u>H0491B17*</u>	<u>H0078P16*</u>
<u>H0465C18*</u>	<u>H0819L13*</u>	<u>H0817H15*</u>	<u>H0738C03*</u>	<u>H0088J03*</u>
<u>H0610H19</u>	<u>H0470A11*</u>	<u>H0370G10*</u>	<u>H0324H21*</u>	<u>H0409C07</u>
<u>H0413J16</u>	<u>H0801H24</u>	<u>H0340C06*</u>	<u>H0791E06*</u>	<u>H0781I24</u>
<u>H0451G04</u>	<u>H0802D20</u>	<u>H0674E16*</u>	<u>H0083B02*</u>	<u>H0277D23*</u>
			<u>H0221J21</u>	

contig克隆错位来组装DNA

完整版，请访问www.kaoyancas.net 科大科院考研网，专注于中科大、中科院考研

(五) 人类基因组图谱

高参考价值的真题、答案、学长笔记、辅导班课程，访问：www.kaoyancas.net



测序、组装和连接成连续的DNA分子

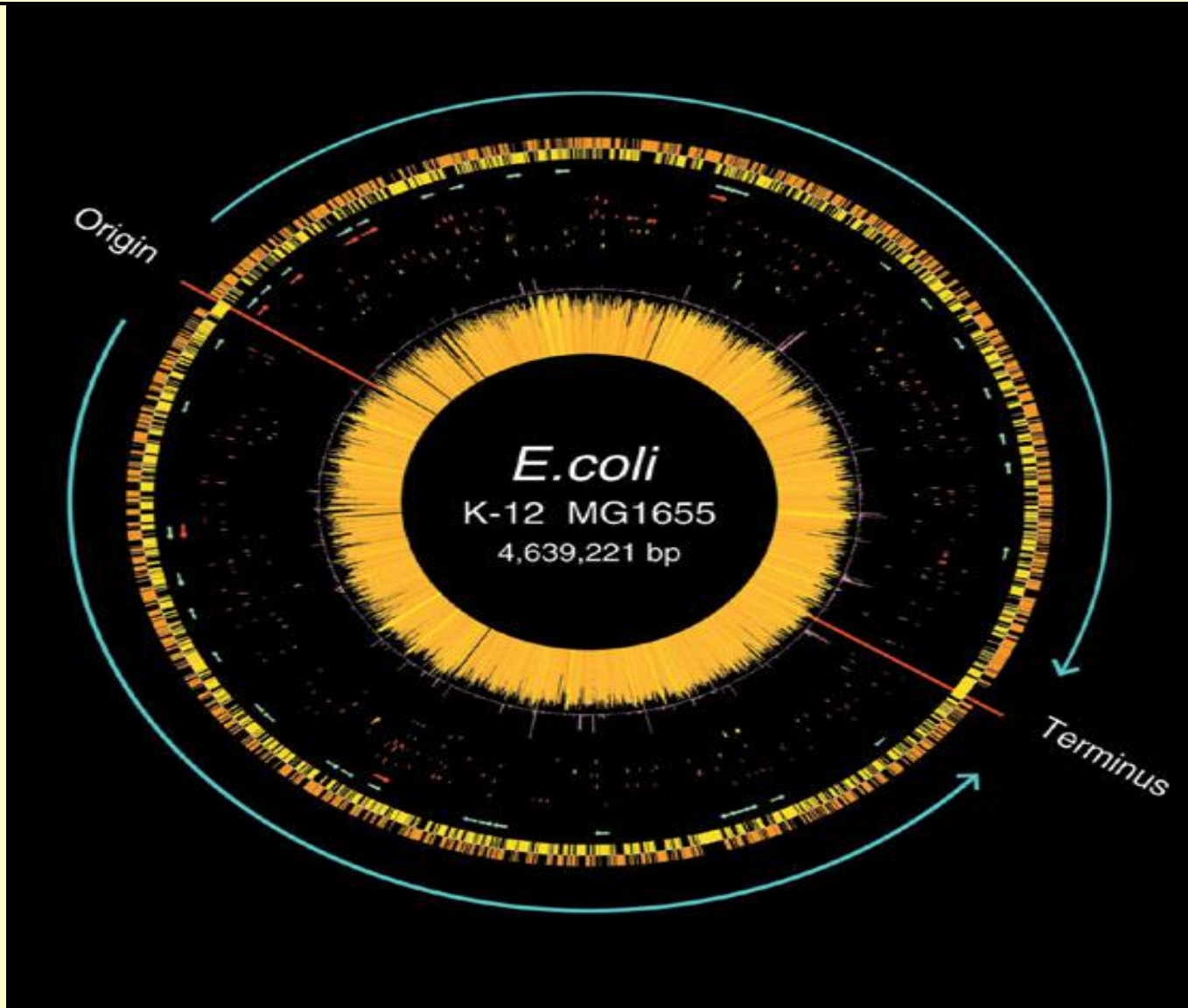
完整版，请访问www.kaoyancas.net 科大科院考研网，专注于中科大、中科院考研

高参考价值的真题、答案、学长笔记、辅导班课程，访问：www.kaoyancas.net

人类部分染色体物

完整版，请访问www.kaoyancas.net 科大科院考研网，专注于中科大、中科院考研

E.coli 物理图谱



- ◆ **基因定位** 借助基因组图谱，使基因定位在精度、速度、广度等方面有极大的提高，在QTL定位分析方面，也取得了很大进展。
- ◆ **基因组比较分析** 从分子水平了解物种间的同源性，研究基因组的进化和染色体的演变。
- ◆ **标记辅助选择(MAS)** 根据图谱间接选择目的基因，大大加速目的基因的转移与利用，提高回交育种的效率。此外，有助于克服轮回选择过程中早期选择的盲目性。
- ◆ **基因的克隆与分离** 根据饱和的基因组图谱，可以找到一个与目的基因紧密连锁的分子标记，作为染色体步行的起始点，进行基因的克隆和分离

第3节 基因组测序

- 一、双脱氧链终止法
- 二、自动化测序
- 三、序列组装

一、双脱氧链终止法

高参考价值的真题、答案、学长笔记、辅导班课程，访问：www.kaoyancas.net

☆DNA的复制需要：**DNA聚合酶**，**单链DNA模板**，带有**3' -OH**末端的**单链寡核苷酸引物**，**4种dNTP（dATP、dGTP、dTTP和dCTP）**。

聚合酶用模板作指导，不断地将dNTP加到引物的3'-OH末端，使引物延伸，合成出新的互补DNA链。

☆如果加入一种特殊核苷酸，**双脱氧核苷三磷酸（ddNTP）**，**ddNTP**在脱氧核糖的**3'位置**缺少一个羟基，故不能同后续的**dNTP**形成磷酸二酯键。

一、双脱氧链终止法

高参考价值的真题、答案、学长笔记、辅导班课程，访问：www.kaoyancas.net

◆例如，存在**ddCTP**、**dCTP**和三种其他的**dNTP**（其中一种为 ^{32}P 标记）的情况下，将引物、模板和**DNA**聚合酶一起保温，即可形成一种全部具有相同的5'-引物端和以**ddC**残基为3'端结尾的一系列长短不一片段的混合物。

◆经变性聚丙烯酰胺凝胶电泳分离制得的放射性自显影区带图谱将为新合成的不同长度的**DNA**链中**C**的分布提供准确信息，从而将全部**C**的位置确定下来。

一、双脱氧链终止法

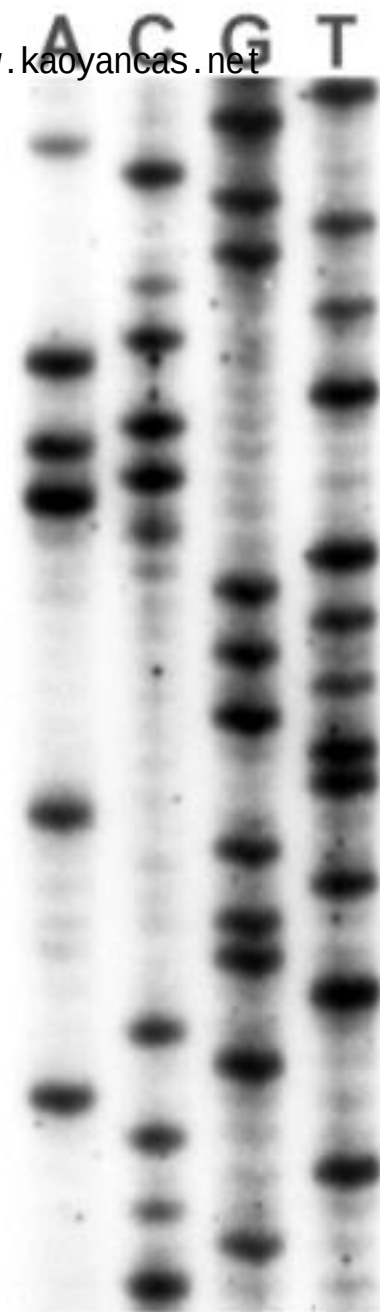
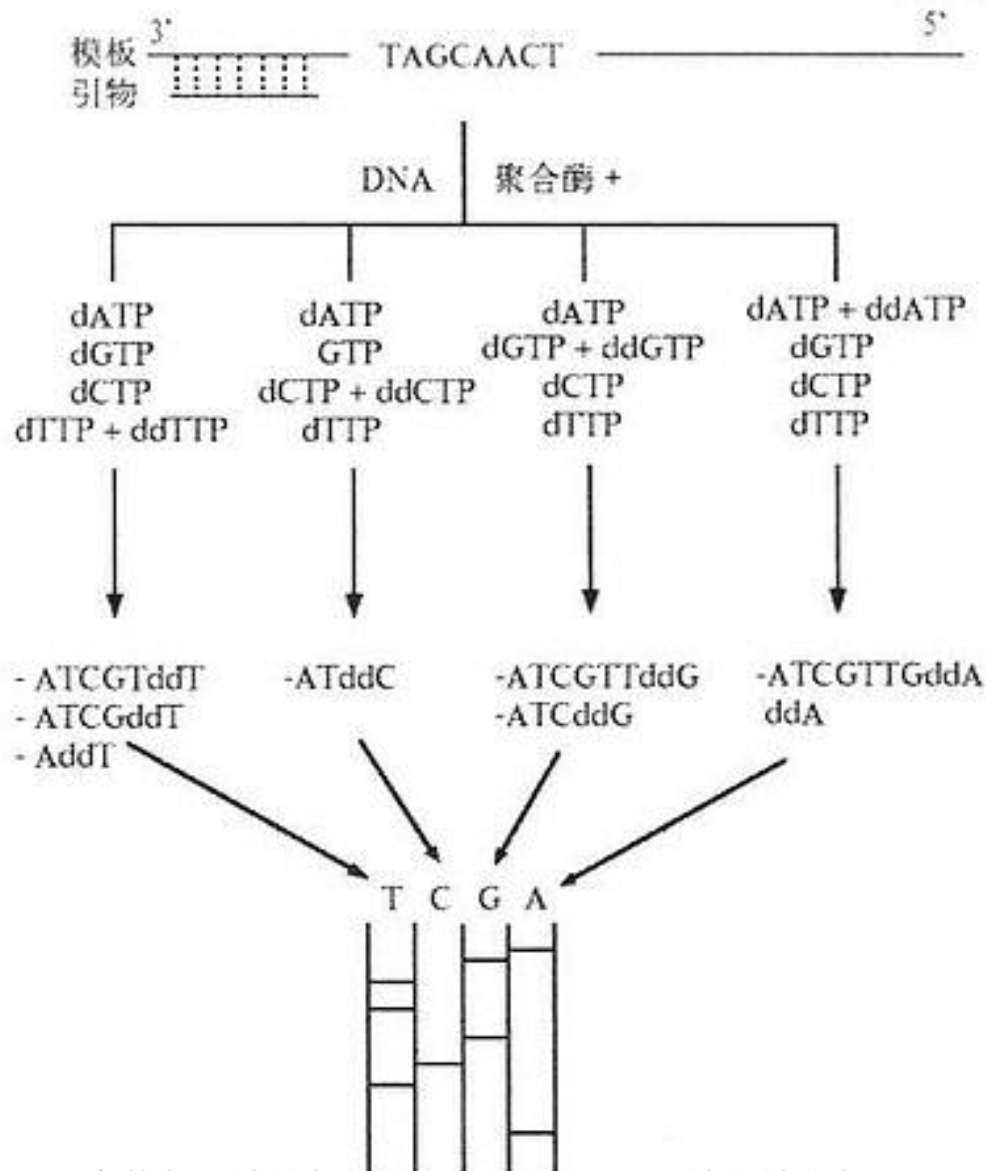
高参考价值的真题、答案、学长笔记、辅导班课程，访问：www.kaoyancas.net

☆采用类似的方法，在**ddATP**、**ddGTP**和**ddTTP**存在的条件下，可同时制得分别以**ddA**、**ddG**和**ddT**残基为3'端结尾的三组长短不一的片段。

☆将制得的四组混合物全部平行地点加在变性聚丙烯酰胺凝胶电泳板上进行电泳，从而制得相应的放射性自显影图谱。从所得图谱即可直接读得**DNA**的碱基序列。

一、双脱氧链终止法

高参考价值的真题、答案、学长笔记、辅导班课程，访问：www.kaoyancas.net



完整版，请访问www.kaoyancas.net 科大科院考研网，专注于中科大、中科院考研

图 9-14 DNA 序列测定。

荧光标记

如ddATP红色荧光

ddCTP蓝色荧光

ddTTP绿色荧光

ddGTP黄色荧光

将4种标记混合在一个反应中，聚合终止链反应完成后，电泳，分别获得末端带有**A、C、T、G**的**DNA**单链，仪器检测。

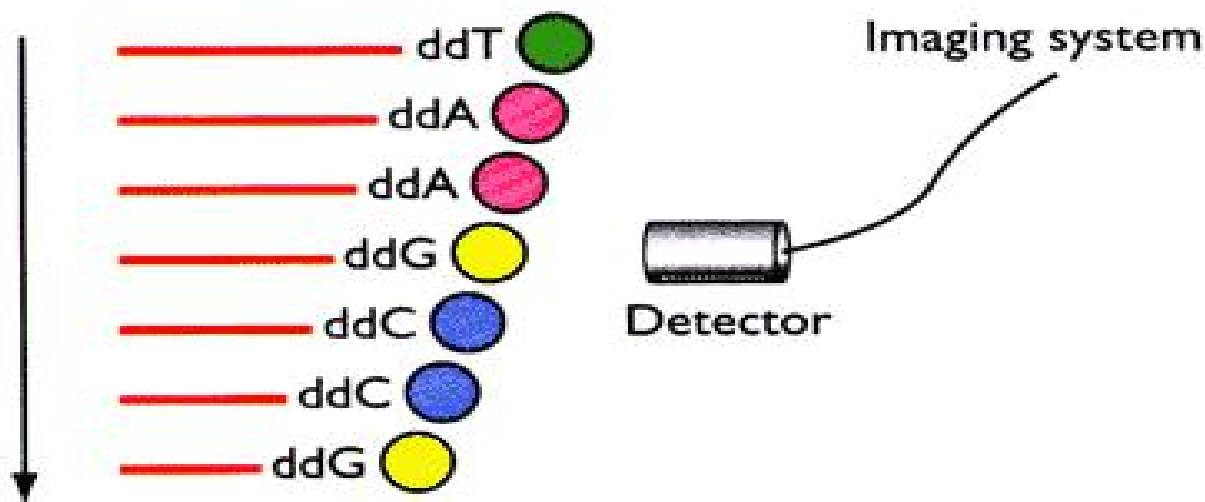
二、自动化测序

高参考价值的真题、答案、学长笔记、辅导班课程，访问：www.kaoyancas.net

(A)

ddA  ddC  ddNTPs – each with a
different fluorescent label
ddT  ddG 

Sequencing reactions,
fractionation of products



Fluorescent bands
move past the detector

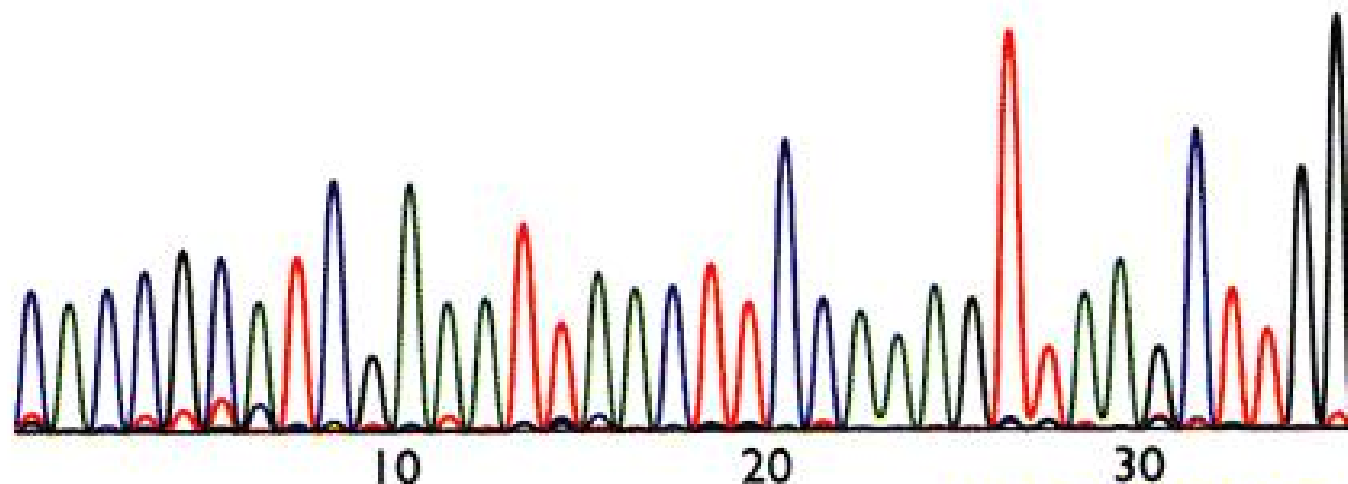
完整版，请访问www.kaoyancas.net 科大科院考研网，专注于中科大、中科院考研

二、自动化测序

请参考价值的真题、答案、学长笔记、辅导班课程，访问：www.kaoyancas.net

(B)

CACCGCATCGAAATTAACTTCCAAAGTTAAGCTTGG



跑得快的先检测到

跑得慢的后检测到

跑得快慢表示长短不同

同一个起点开始复制，短的表示在前头，长的表示在后头

根据荧光定序列。

将终止链法阅读的小的**DNA**片段组装称真实的排列顺序，是一项繁浩而精细的工作。

（一）随机测序与序列组装

☆直接鸟枪法序列组装：直接从已经测序的小片段中，寻找彼此重叠的测序克隆，依次向两侧邻接的序列延伸。不需预先了解任何基因组情况，即缺少遗传图和物理图也可以完成这个基因组序列的组装。

三、序列组装

高参考价值的真题、答案、学长笔记、辅导班课程，访问：www.kaoyancas.net

☆流感嗜血杆菌：将DNA打断成2kb的片断→连接到质粒载体（1.6~2kb）→挑19 687个克隆，进行28 643次测序，11 631 485bp序列（相当于其基因组的6倍）→得到140个重叠群（尚有空隙，gap）

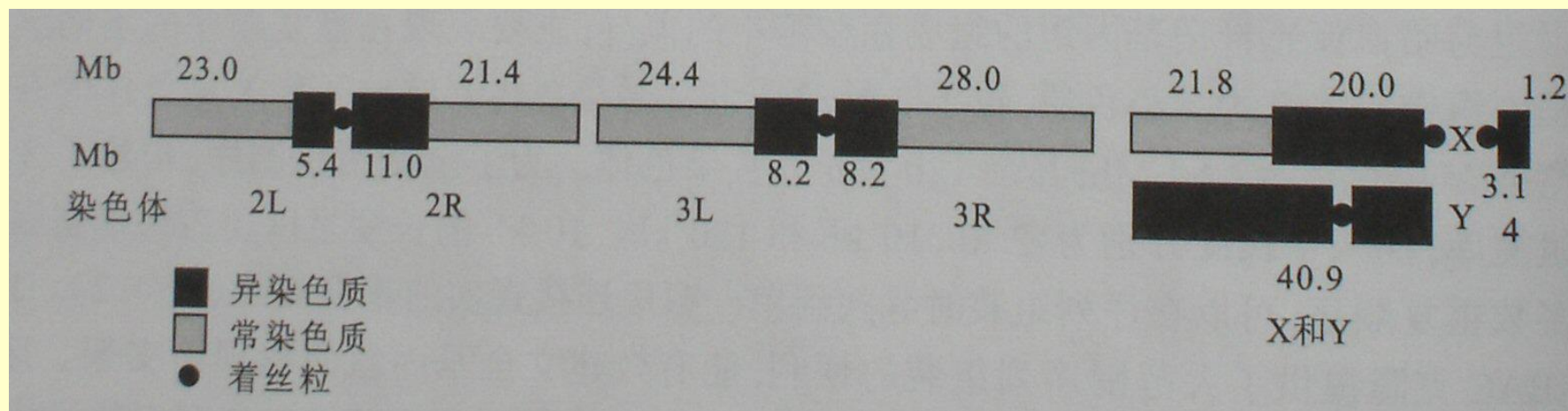
◆顺序空隙：测序时遗漏的顺序。用已知序列钓取（99个）

◆物理空隙：建库时载体和宿主均选用不当而丢失的顺序。重新构建文库，间隙两侧序为探针或制备PCR引物（23）

四、序列组装

高参考价值的真题、答案、学长笔记、辅导班课程，访问：www.kaoyancas.net

- ☆果蝇: 180Mb, 60Mb异染色质, 120Mb常染色质。
- ☆构建了2kb, 10kb, 130kb等3个基因组文库
- ◆10kb在序列组装时, 寻找单一序列连接的重叠克隆。
- ◆130kb的BAC克隆用于建立全基因组物理图支架。
- ☆利用已有的STS标记筛选BAC克隆, 构建的物理图谱覆盖95%的常染色体区域、根据测序结果, 组建的克隆重叠群和支架群总长120Mb, 包括了97.5%的基因。



(二) 限定测序与序列组装

高参考价值的真题、答案、学长笔记、辅导班课程，访问：www.kaoyancas.net

☆大基因组**DNA**分解成较大的片段，然后进行随机克隆，构建基因组文库。对每个大片段克隆进行独立鸟枪测序，或组建重叠群后测序。重叠群经分子标记定位到遗传或物理图上，并用该区段的其他标记（**STS**，**EST**等）验证组装结果。

➤拟南芥测序完全依赖克隆重叠群，先进行**BAC**克隆随机测序，在进行组装。国际水稻基因组测序也采用这个策略。

➤人类基因组计划主要采用**BAC/YAC**文库，已有300 000个**BAC**克隆，大多数定位到物理图谱中。采用鸟枪法完成**BAC**测序，再按照物理图进行序列组装。

(三) 指导测序与序列组装

高参考价值的真题、答案、学长笔记、辅导班课程，访问：www.kaoyancas.net

☆人类基因组测序 如果完成**7 000**万次测序，每次**500bp**。可以给出**35 000Mb**的序列，覆盖**99%**以上的基因组序列。

☆对**7 000**万次测序工作安排如下：构建**2kb**和**10kb**的质粒文库，每个克隆双向可读序列**500bp**。

采用**2**个文库：减少丢失的克隆，扩大覆盖面，矫正重复序列产生的差错（多数重复序列在**5kb**和略长）。

☆序列组装参考人类基因组工作框架图，特别是大量**STS**标记，可以事随机获得的顺序重叠群准确的锚定到基因组图上。

☆重要区域优先测序

☆ **EST**测序

☆在完成基因组图谱构建以及全部序列测定的基础上，研究全基因组的基因功能、基因之间相互关系和调控机制为主要内容的学科。

☆面临的最重要问题

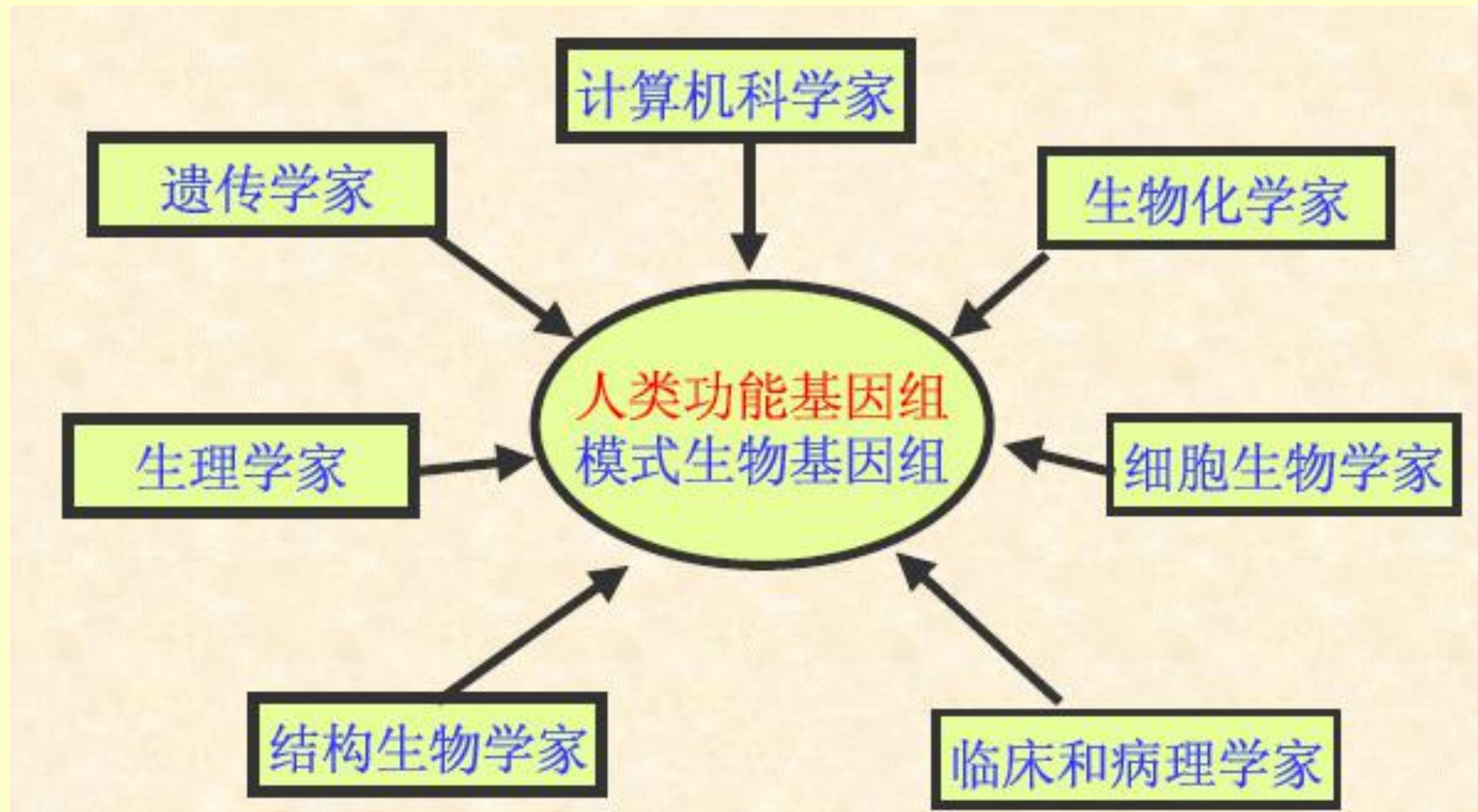
➤虽然完成了99%人类基因的序列分析，但90%以上的人类基因尚不了解功能。

➤从1996年酵母菌基因组全序列测定以来，全世界已有1000多个实验室、5000多名科学家从事酵母菌后基因组学的研究，共发表论文7000多篇，鉴定1060个新基因的功能，但仍然还有约1600个阅读框架（基因）的功能不清楚。

第4节 后基因组学(post-genomics)

高参考价值的真题、答案、学长笔记、辅导班课程，访问：www.kaoyancas.net

如何将人类和模式生物体基因序列资料转变为有用的知识，如何对这些基因加以利用，使之能够造福于人类的健康。



完整版，请访问www.kaoyancas.net 科大科院考研网，专注于中科大、中科院考研

第5节 后基因组学

高参考价值的真题、答案、学长笔记、辅导班课程，访问：www.kaoyancas.net

☆目前已衍生出了许多新兴学科：

➤疾病基因组学 发现疾病相关基因和致病基因，从疾病诊断到疾病易感性研究。

➤药物基因组学(**Pharmacogenomics**) 研究不同个体对药物敏感性的基因基础，特别是 **SNPs**。

➤功能基因组学 (**Functional genomics**) 对基因及其编码蛋白的功能进行研究。

第5节 后基因组学

高参考价值的真题、答案、学长笔记、辅导班课程，访问：www.kaoyancas.net

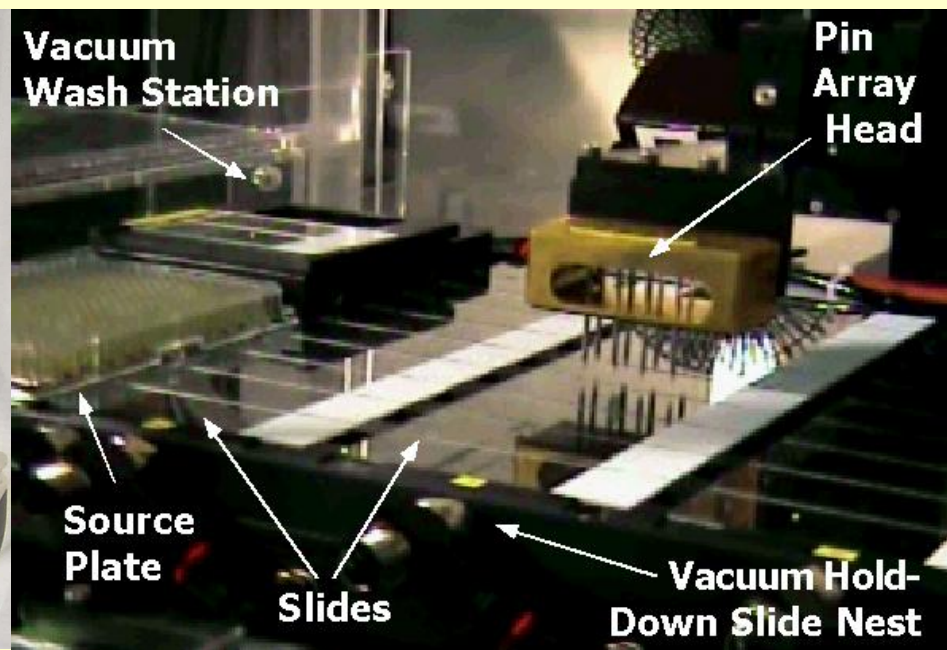
- **环境基因组学 (Enviromental Genomics)** 鉴定机体暴露在特定环境下的那些显示易感性或抵抗性基因的**DNA**多态性。如**DNA**修复基因、细胞周期相关基因、激素代谢基因、受体基因、参与免疫和感染反应的基因和信号转导基因等。
- **蛋白质组学 (Proteomics)** 利用双相电泳技术研究细胞或组织的基因组表达的全部蛋白质
- **生物信息学 (Bioinformatics)** 现代生物技术与计算机科学的结合，收集、加工和分析生物资料和信息。

☆ 又称为后基因组学，它利用结构基因组所提供的信息和产物，发展和应用新的实验手段，通过在基因组或系统水平上全面分析基因的功能，使得生物学研究从对单一基因或蛋白质得研究转向多个基因或蛋白质同时进行系统的研究。

☆ 研究内容：包括基因功能发现、基因表达分析及突变检测。基因的功能包括生物学功能；细胞学功能；发育上功能。

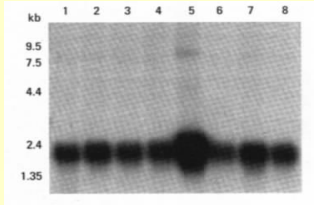
☆ 功能基因组学研究涉及众多的新技术，包括生物信息学技术、生物芯片技术、单核苷酸多态性、转基因和基因敲除技术、酵母双杂交技术、基因表达谱系分析、蛋白质组学技术、高通量细胞筛选技术等。

DNA微列阵(DNA microarrays)是利用**DNA芯片技术**，同时进行大量分子杂交，以分析比较不同组织或器官的基因表达水平，筛选突变基因，从核酸水平分析基因表达模式。是后基因组学研究中的重要方法之一。

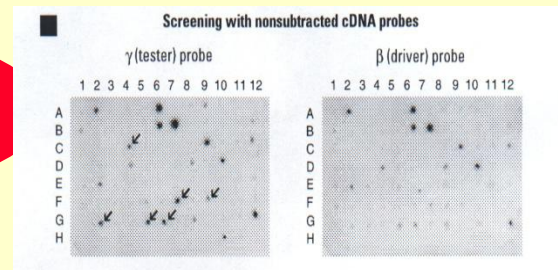


基因芯片发展历史

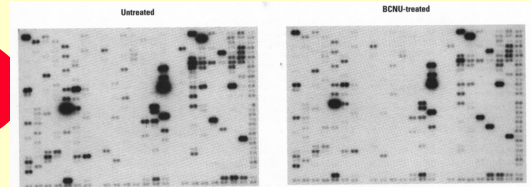
Southern & Northern Blot



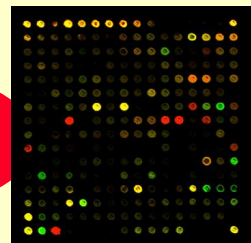
Dot Blot

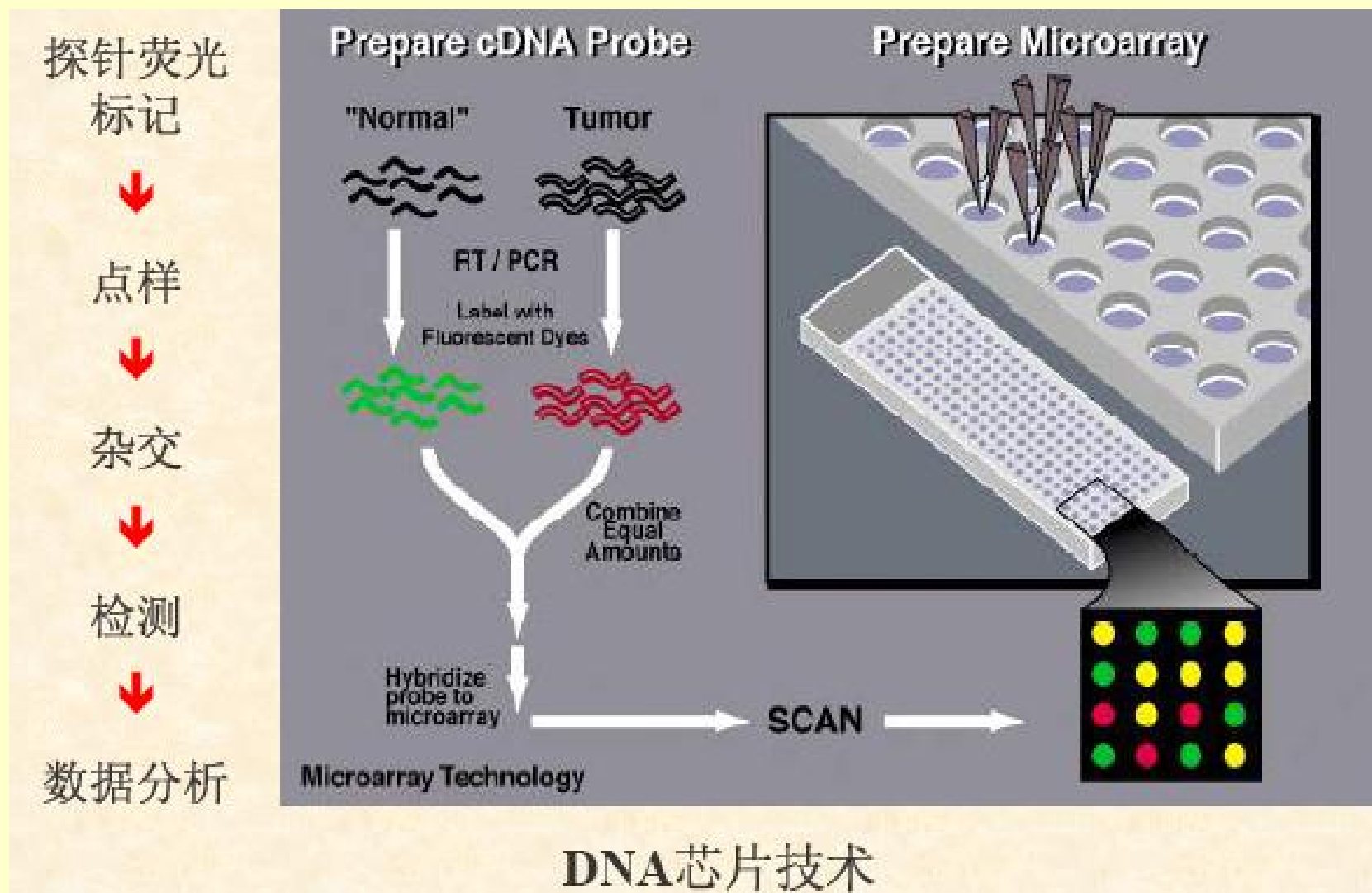


Macroarray



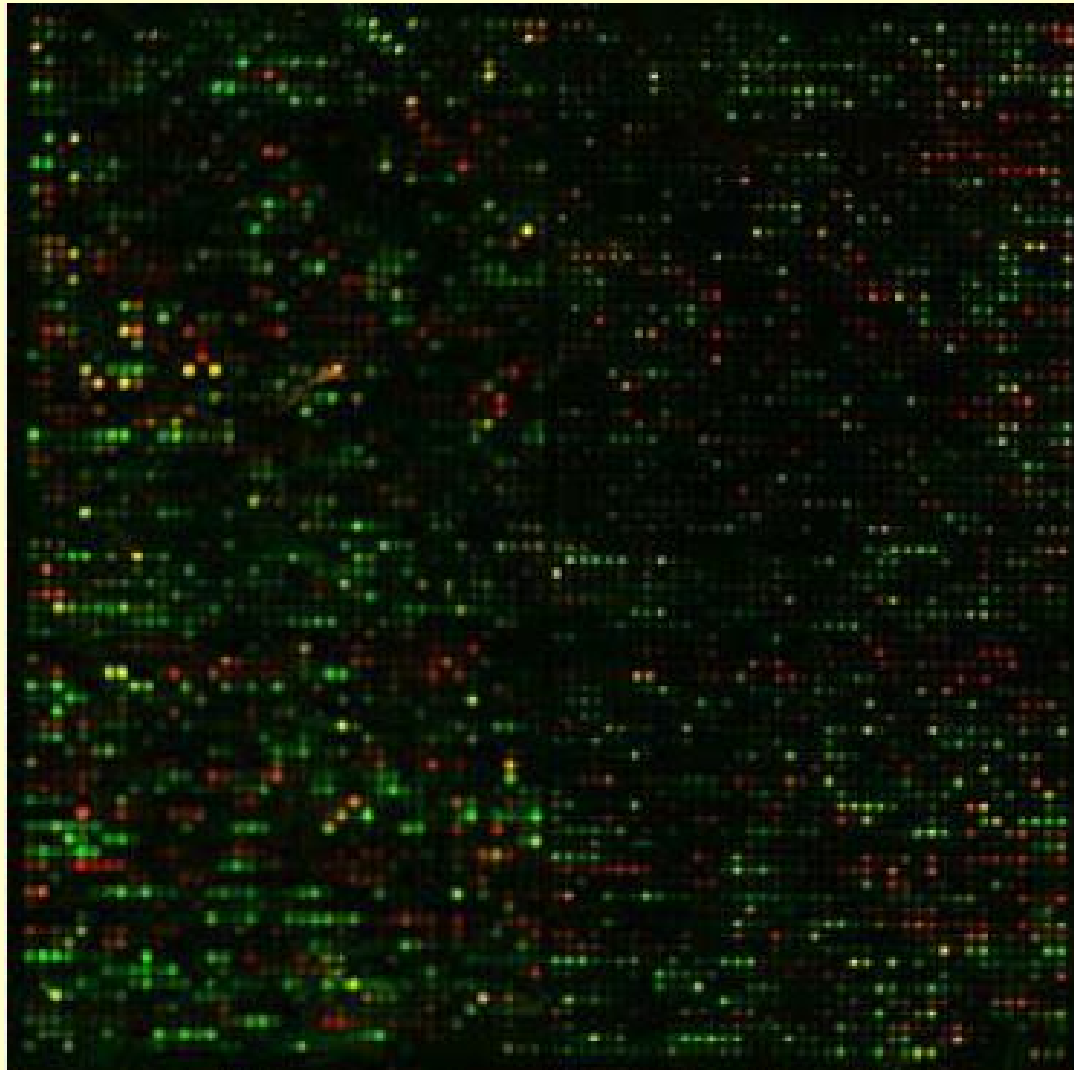
Microarray





高参考价值的真题、答案、学长笔记、辅导班课程，访问：www.kaoyancas.net

基因芯片展示酵母菌在发酵作用和呼吸作用状态下的基因表达

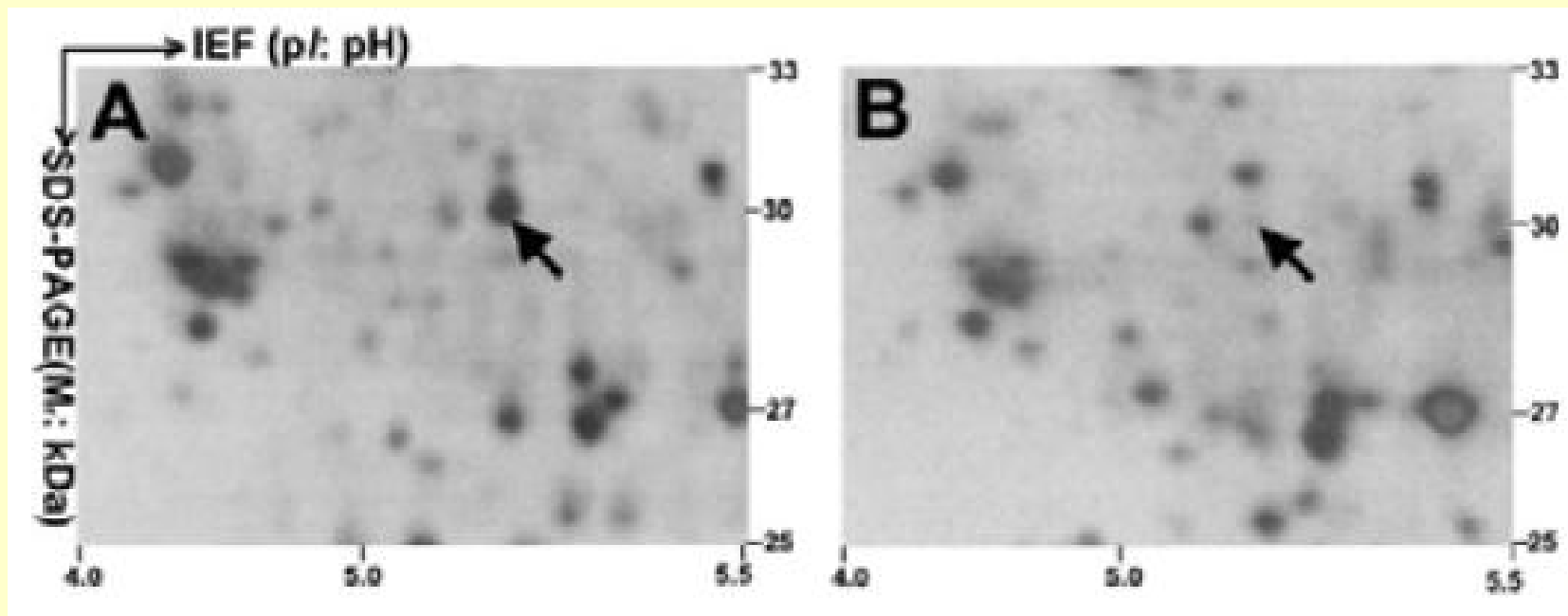


完整版，请访问www.kaoyancas.net 科大科院考研网，专注于中科大、中科院考研

☆在蛋白质水平研究基因组的基因表达。分析基因组的蛋白质类型、数量、空间结构变异以及相互作用的机制。

☆蛋白质组学比基因组学更为复杂：**DNA**线状结构与二级结构的功能差异不大，但多肽链需折叠成一定的三维空间结构才形成有功能的蛋白质；同一种蛋白质经不同的加工修饰可形成不同的功能，因此蛋白质的多样性远复杂于基因本身。

蛋白质组学 (Proteomics)



蛋白质2D电泳分析

A对照；**B**为人的垂体瘤蛋白质2D电泳箭头示有明显差异的蛋白质。

答案、学长笔记、辅导班课程，访问：www.kaoyancas.net

☆生物信息学是现代生物技术与计算机科学的结合，收集、加工和分析生物资料和信息学科。

☆应用生物信息学可以将来自不同的基因组理论和应用综合并标准化，利用大量的生物信息资料了解遗传网络系统、信号传递及相互关系，计算机还可进行一些生物模拟研究。

☆利用生物信息学能够分析从微生物、动物、植物以及人类基因组序列测定产生的大量资料，阐明遗传信息。

研究内容两大类：**DNA** 数据分析；
蛋白质数据分析。

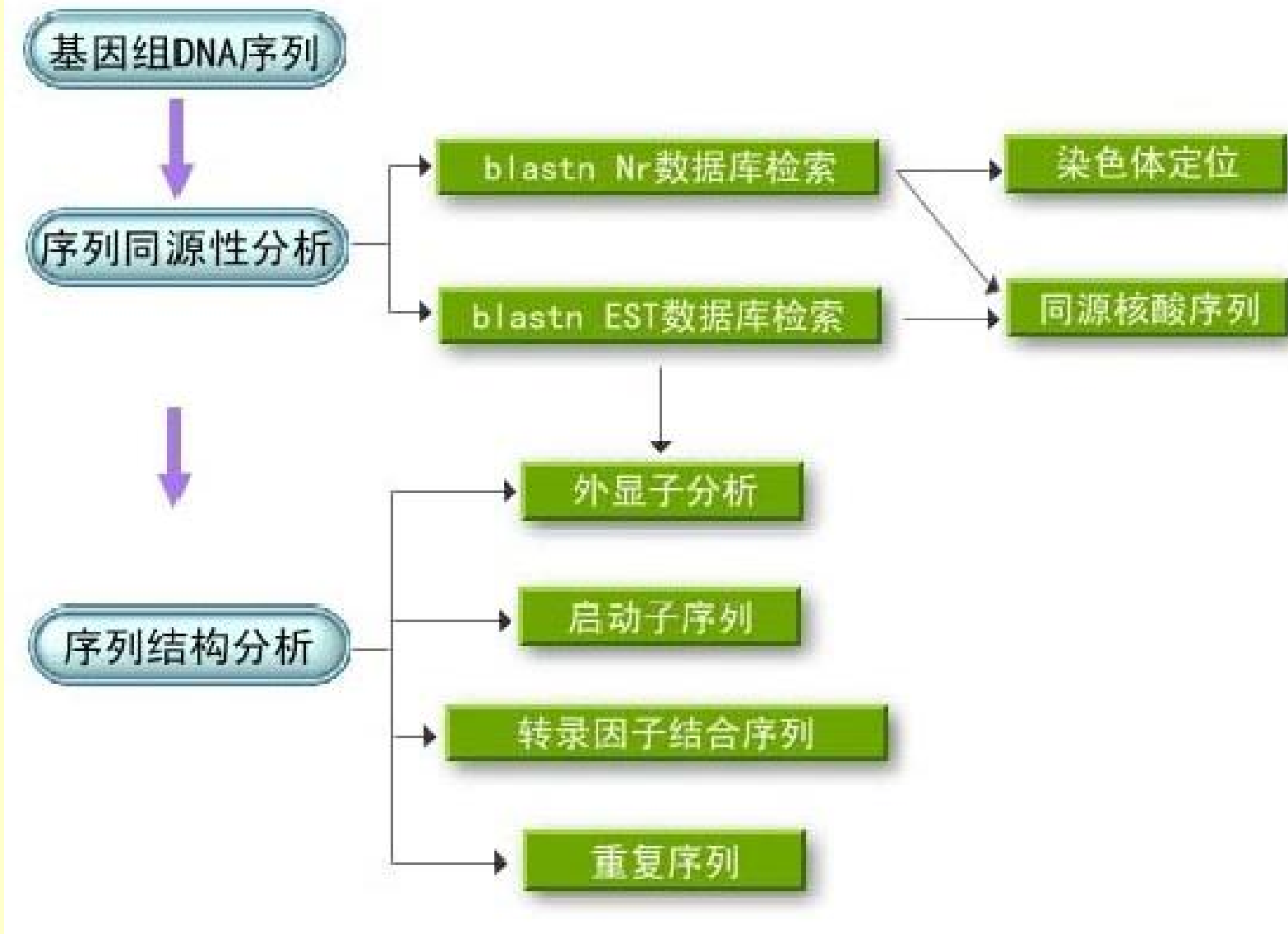
1.DNA分析

高参考价值的真题、答案、学长笔记、辅导班课程，访问：www.kaoyancas.net

☆基因结构域分析，包括启动子、转录因子结合序列、内含子、外显子、重复序列、开放读码框架等。

☆同源分析和检索，包括**DNA**数据库、**EST**数据库、**STS**数据库、**Unigene**数据库、**Swissprot**数据库等。

☆计算机基因克隆化：利用**EST**数据库的重叠序列克隆新基因。





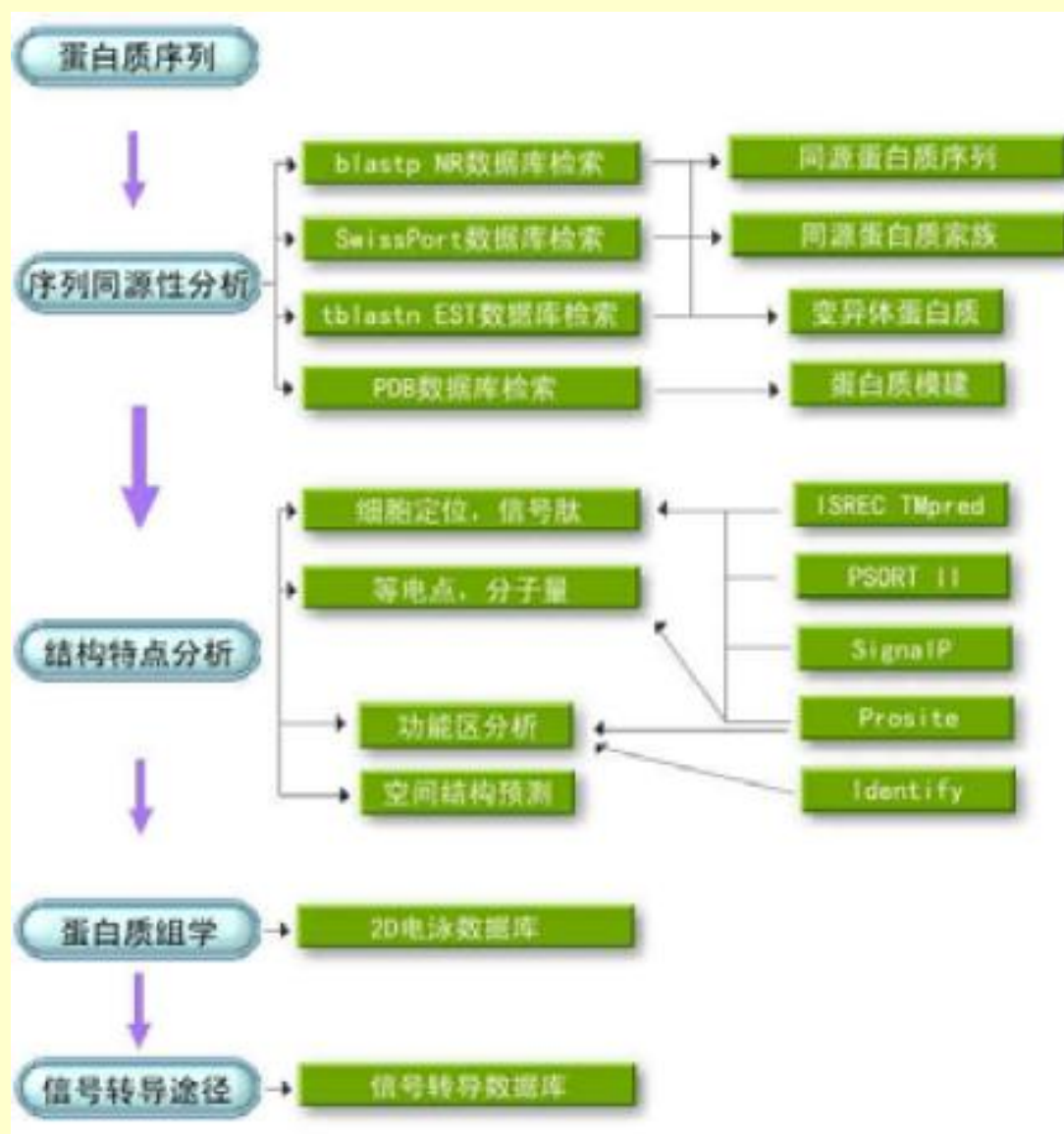
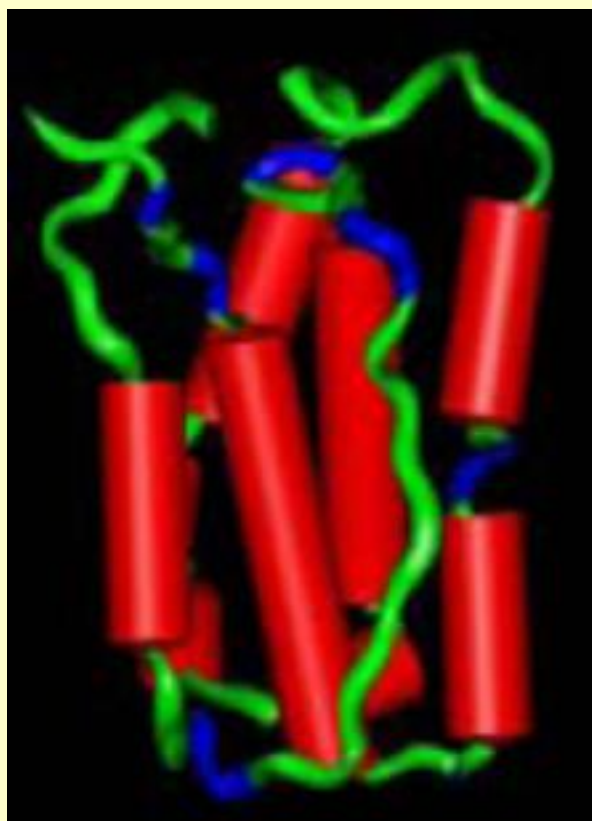
☆蛋白质一级结构分析：

结构特点分析，包括等电点、信号肽、穿膜区、**DNA**结合序列等同源分析和检索，包括**Nr**数据库、**Swissprot** 数据库等功能区分析，包括**Prosite**、**Emotif**、**Identify**分析等。

☆蛋白质空间结构分析：

蛋白质晶体结构数据库检索，如**PDB**数据库。
蛋白质空间结构预测，如**Homology**等软件分析。

干扰素 $\alpha 1b$ 分子模建图





高参考价值的真题、答案、学长笔记、辅导班课程，访问：www.kaoyancas.net

National Center for Biotechnology Information

National Library of Medicine

National Institutes of Health

[PubMed](#)

[Entrez](#)

[BLAST](#)

[OMIM](#)

[Books](#)

[TaxBrowser](#)

[Structure](#)

[Site Map](#)

[Guide to NCBI resources](#)

[About NCBI](#)

[An introduction for researchers,
educators and the public](#)

[GenBank](#)

[Sequence submission support and
software](#)

[Literature databases](#)

[PubMed, OMIM, Books, and
PubMed Central](#)

[Molecular databases](#)

[Sequences, structures, and
taxonomy](#)

[Genomic biology](#)

[The human genome, whole genomes, and related
resources](#)

[Tools](#)

[Data mining](#)

[Research at NCBI](#)

[People, projects, and seminars](#)

[Software engineering](#)

[Tools, R&D, and databases](#)

[Education](#)

[Teaching resources and on-line tutorials](#)

[FTP site](#)

[Download data and software](#)

[Contact information](#)

[How to reach us](#)

<http://www.ncbi.nlm.nih.gov/>

完整版，请访问www.kaoyancas.net 科大科院考研网，专注于中科大、中科院考研

本章重点

- 基本概念
- 基因组图谱绘制（包括分子标记）
- 基因组测序原理和序列组装策略

作业