**基因组学 2015**

**第一章 基因组学研究进展**

**基因组：**生物所具有的携带遗传信息的遗传物质总和称为基因组

**基因组学：**基因组学是研究生命体全部遗传信息的一门科学

**模式生物：**相对比较容易研究的生物，可以从中获得重要的生物学信息用于其它复杂生物的研究参考。

**为什么说基因组学是生命科学的前沿学科？**

因为在科学的发展下，特别是在人类基因组计划的影响下，分子生物学的主要目标已经从传统的单个基因的研究转向对生物整个基因组结构和功能的研究，生命科学正从全新的视角研究与探讨各项生命活动。基因组学研究的对象涉及原核生物和真核生物不同的种属，其所研究的内容触及到生命学科的各个领域，对生命科学的未来发展将产生重大影响

**基因组学形成和发展的科学技术基础。**

DNA重组技术、DNA测序技术、PCR技术、RNAi技术、基因敲除技术、高性能计算机、数据库技术。

**基因组学有哪些特点？**

整体性，大科学性（复杂性），原创性，前沿性，交叉性，可比性，竞争性，自动化，程序化（标准化），规模化，快速化，产业化

**为什么基因组DNA测序能发现许多新基因？**

因为测序之后我们知道了DNA序列的详细信息，就可以通过已知的顺序来判断或寻找与基因有关的序列，比如用ORF筛选，同源查询，EST筛选全长cDNA。或者进行实验研究，比如用基因沉默，来观察其能否表达基因产物以及对表型的影响。

**基因组研究已取得哪些重要进展？**

大肠杆菌测序

果蝇测序

拟南芥测序

2000人类基因组草图顺序的完成

2002人类基因组“精细图”

2003人类基因组序列图绘制完成

狗，鸡以及很多很多各种生物的测序完成

**第二章 基因组结构**

**基因组等容线：**大部分真核基因组表现出一种称为等值区(isochore)的组织形式。等值区定义为“具有一致碱基组成的长区域”或“连续分布的具有相似碱基组成的DNA区段”，它们在基因组中成片相嵌排列

**CpG岛：**基因组中富含GC碱基（60-70%）的DNA区段，一般长度为1-2 kb。

**染色体组：**不同真核生物核基因组均由一定数目的染色体组成，单倍体细胞所含有的全套染色体。

**序列复杂性：**基因组中单拷贝的DNA序列称为单一序列，多拷贝的DNA序列称为重复序列，不同序列的DNA总长称为复杂性(complexity)。

**C值：**一个物种单倍体基因组的DNA含量是相对恒定的，它通常称为该物种DNA的C值。

**支架附着区（SAR）：**与染色体骨架附着区结合的DNA顺序称为SAR。

1. **为什么要在基因组水平上研究生命现象？**

因为基因组是作为一个整体的，有着自己完整的结构和功能，很多调控机制，分子机理遗传变异的原理等等都是发生在基因组水平上的。加上现在DNA测序等技术的蓬勃发展，人类基因组计划的完成，我们有能力在基因组水平上进行研究。所以在基因组水平上研究生命现象是势在必行的。

1. **为什么植物线粒体基因组大小在不同生物中变化很大,而叶绿体基因组大小相对稳定?**

线粒体多变：

1)结构多变性: 重组与重排事件频繁发生，导致基因组结构、不同DNA分子拷贝数比例变化，出现无功能的假基因等

2）高比例的非编码序列，如重复序列、基因间隔序列、内含子

3）等价基因比其他生物要大

叶绿体变化不大：

叶绿体基因组在物种间变化不大，所有的基因组都有相似的组成。高等植物叶绿体基因组都有两段较大的反向重复区段，反向重复结构可阻止叶绿体基因组的分子内重组，使其保持较为稳定的组成。

1. **有什么RNA没能成为细胞生物的遗传信息的载体？**

1.2′-OH非常靠近连接两个核苷酸的磷酸二脂键位置，使RNA的磷酸二酯键对 碱性环境变得十分敏感，即使在正常细胞内微偏碱的PH条件下也容易降解。

2. 2′-OH的化学作用使RNA构型的选择范围受到很大限制，通常RNA双螺旋区段均在数十碱基对以下

3.长链RNA增加了2′-OH与磷酸二酯键相互作用引起分子断裂的可能，限制了RNA长度的增加，由于不能形成很长的多聚分子，RNA只能储存有限的遗传信息。4.DNA中胸腺嘧啶的位置在RNA由尿嘧啶取代，当RNA分子中的胞嘧啶甲基化脱氨基形成尿嘧啶后，细胞无法区分突变的尿嘧啶与正常的尿嘧啶，无疑会增加RNA的突变概率。天然的DNA中存在胸腺嘧啶但缺少尿嘧啶时，课立即由细胞的修复系统识别并予以修复，维持了遗传的稳定性。

**4.真核生物与原核生物基因组的结构有何差异？**

**真核生物基因组**

1.分为核基因组和细胞器基因组，

2.基因组结构复杂,数目庞大，多个复制起始点

3.mRNA为单顺反子，

4.含大量重复序列，

5.非编码序列占90%以上，

6.基因间有间隔区(spacer DNA),具有内含子,外显子，

7.功能相关的基因串联在一起形成基因家族，

8、存在可移动成分.

9、密码子使用偏性

10.CpG岛

11、等容线或称等值区(isochore)

**原核生物基因组**：

基因组为环状双链DNA分子，只有一个复制起始点，具有操纵子结构，一般无重叠基

因，基因是连续的,无内含子，编码区在基因组中的比例大于真核基因组，小于病毒基因组

**5. 什么是DNA序列复杂性？哪些因素影响序列复杂性？**

基因组中单拷贝的DNA序列称为单一序列，多拷贝的DNA序列称为重复序列，不同序列DNA总长称为复杂性(complexity)

特定DNA的序列复杂性用C0t1/2＝1/k，如果基因组中每一种基因只有一个，即都是单拷贝序列，那么基因组愈大则基因组的复杂性愈大，还取决于每个基因的核苷酸序列的重复次数，重复次数愈少则复性愈慢，C0t1/2的位置愈后，重复次数愈多，C0t1/2位置愈前

**第三章 遗传图绘制**

**遗传图谱：**是以遗传距离表示基因组内基因座位相对位置的图谱

**遗传作图：**采用遗传学分析方法将基因或其它DNA顺序标定在染色体上构建连锁图。这一方法包括杂交实验，家系分析。

**DNA标记：**一段DNA顺序，具有2个或多个不同的可以区分的版本

**重组热点:** 染色体的某些位点之间比其它位点之间有更高的交换频率

**共分离:** 在有性繁殖的后代，这种分子标记与连锁的基因有最大的可能同时出现在同一个体中，这一现象被称为共分离

**人类基因组计划中为什么要构建遗传图？**

因为人类基因组有30亿个碱基对，其中还包含有大量的重复序列。要在这样大的序列中确定某一基因或特定DNA序列的位置，如同大海中捞针，因此有必要对整个人类基因组予以划分。而遗传图谱是根据可观察形状实验得出的重组率绘制的，就像给整张图谱画出了大概的框架，是基因组测序、组装的支撑点。

**性状形态标记进行遗传学分析有哪些局限？DNA分子标记的优点是什么？**

**性状形态标记的局限**：可利用的形态标记数量少，多态性差，会在非编码序列段留下空白，多基因影响的同一性状遗传分析困难，易受环境条件影响等问题

**DNA分子标记的优点**：中性、共显性、多态性高、数量多、分布均匀、不受环境影响等优点

多基因影响的同一性状遗传分析可以将一个数量性状分解为多个QTLs（Quantitative traits locus），通过分析数量性状的基因位点，估算每个QTL的遗传效应等

**有哪些标记可供遗传分析，构建遗传图谱？各自的特点是什么？**

**形态标记**：可利用的形态标记数量少，多态性差，多基因影响的同一性状遗传分析困难，易受环境条件影响等问题

**细胞标记**：染色体核型分析，三体、单体、缺体等

**生化标记**：营养型突变体、同功酶、血型等

**DNA分子标记**：不同的生物，同一种生物不同的个体，DNA核苷酸排列顺序存在差异，这样DNA本身可以成为遗传标记。中性、共显性、多态性高、数量多、分布均匀、不受环境影响等优点

**DNA分子标记**：

**第一代标记: 限制性片段长度多态性（RFLP)**，限制酶处理，会产生不同的限制性片段

**第二代标记:简单序列长度多态性（SSLP）：**生于重复顺序的可变排列，同一位点重复顺序的重复次数不同，表现出DNA序列的长度变化。与RFLP不同的是，SSLP具多等位性（multiallelic），每个SSLP都有多个长度不一的变异体。

小卫星序列（大量可变串联重读VNTR）

微卫星序列（简单串联重复STR）：比小卫星应用广泛，因为分步均匀，且便于PCR

**第三代标记:单核苷酸多态性(SNP)：**位点极其丰富。分布密集

**为什么说基因标记不是理想的标记？**

1）高等生物如脊椎动物和显花植物，可用作标记的基因十分有限，许多性状都涉及多基因

2）高等生物基因组中存在大量的基因间隔区，纯粹用基因作为标记将在遗传图中留下大片的无标记区段

3）只有部分基因其等位基因成员可以通过常规实验予以区分，因而产生的遗传图是不完整的

4）重要性状由多基因控制

5）性状=遗传+环境

**水稻细胞遗传学相对落后于多倍体植物，水稻经典遗传图远远落后于玉米、小麦，但为什么它能成为基因组研究的模式植物呢？**

水稻重要粮食作物

基因组小、重复顺序又较少

转基因技术成熟

禾谷类作物基因组共线性

**如何设计实验将野生稻有利基因（如控制产量的基因）转入栽培稻，选育出更加优良水稻品种？**

用来自于马来西亚的普通野生稻(O. rufipogon)为供体, 与三系杂交稻不育系系V20A,恢复系明恢63受体配制杂交组合能结实, 得到F1, 然后分别以Ce64和明恢63为轮回亲本进行连续回交获得BC1F1、BC2F1和BC3F1代的个体单株材料

**第四章 物理图谱**

**DNA指纹**：指确定DNA样品所具有的特定DNA片段组成。

**辐射杂种群**：杂种细胞的集合体称为辐射杂种群

**锚定标记**：用于衔接克隆重叠物理图与遗传图单一序列的分子标记

**限制性作图**：它是将限制性酶切位点标定在DNA分子的相对位置

荧**光原位杂交**：将荧光标记的探针与染色体杂交确定分子标记的所在位置。

**序列标记位点**: 是一小段长度在100到500 bp的DNA顺序，每个基因组仅一份拷贝，很易分辨。

**序列标记位点作图**: 通过PCR或分子杂交将小段DNA顺序定位在基因组的DNA区段中。

**克隆文库**：包括全基因组文库DNA和指定单一染色体文库。

**物理图谱:** 物理图指表示DNA序列上DNA标记之间实际距离的图。通常由DNA的限制酶片段或克隆的DNA片段有序排列而成。标记之间的物理距离以DNA上核苷酸数目的多少来表示

**物理作图:** 采用分子生物学技术直接将DNA分子标记、基因或克隆标定在基因组

实际位置

**重叠群:** 相互重叠的DNA片段组成的物理图称为重叠群

**稀有切点限制酶:** 所谓稀有切点限制酶系指该酶识别的碱基顺序在基因组中只有很少数量，可产生较大的DNA片段。

**染色体步移:** **克**隆重叠群的组建方法，以起始克隆A1为探针筛选基因组文库，两个与A1杂交的序列分别为B1和B2，通过A1两端的序列可以确定B1和B2的相对位置。然后再以B1和B2为探针筛选基因组文库，以此类推。

**辐射杂种：**是含有另一种生物染色体片段的啮齿类细胞

**厘镭：**辐射杂种的做图单位，是DNA分子暴露在N拉德X射线剂量下两个分子标记之间发生1％断裂的频率

**作图试剂：**覆盖整条染色体或整个基因组的DNA片段群体，用于STS作图

**物理图深度:** 指物理图的任何一点上DNA片段的重复次数。深度与物理图的精确性直接相关，深度越大，物理图的精确性就越高

**1.为何要绘制物理图？遗传分析绘制的基因组图为何不能指导基因组计划的测序？**

（1）遗传图的分辨率有限

人类及大多数高等真核生物由于不可能获得大量的子代，只有少数的减数分裂事件可供研究，连锁分析的分辨力受到很大限制。人类遗传图其标记密度平均为599 kb，离每100 kb一个标记的要求仍差距甚远，后者是进入基因组全面测序的前提。这种高密度基因组图仅仅采用遗传作图技术是无法完成的，必须借助于其他非遗传分析的方法

（2）遗传图的精确性较低

连锁分析/重组热点的存在使染色体某一区段的交换频率高于其他区段。特别是倒位区段，由于受到交换限制，无法绘制精细遗传图

（3）遗传图

与物理图并不完全一致，主要表现在2个方面：1.基因座位之间的相对距离不同。遗传作图受重组频率的影响，热点重组区比其它区段图距更长2.相对位置可能有点偏差，例如一对基因座位chal和glk1的相对位置在遗传图与物理图中正好相反

**2.基因组的物理图有什么作用？**

根据遗传研究所得的有关信息，人们可指望利用物理图获得预期有用的基因

物理图为测定全基因组DNA全序列提供了必要的“骨架”。而基因组DNA全序列的测定，就能使人们最终可以在分子（核苷酸）以知水平上解开生命的遗传奥秘

**3.克隆载体有哪几类？**

质粒；

λ克隆载体；

粘粒40kb~；

YAC（酵母人工染色体）；

噬菌体P1载体；

BAC（细菌人工染色体）；

PAC（P1人工染色体）。

4**.YAC克隆系统和BAC克隆系统各自的特点？**

YAC克隆系统由着丝粒、端粒和自主复制起始点构成，酵母染色体的可载容量为230 kb至1700 kb之间，相当于自然的酵母染色体。

标准的YAC操作程序克隆的DNA片段平均约600 kb，改进的程序可获得1400 kb插入子。YAC克隆存在插入子不稳定，同一酵母细胞多个YAC引起交换产生嵌合顺序

组建BAC克隆的质粒可通过电激转染到大肠杆菌内，转化效率比酵母原生质的转化效率高10-100倍

克隆的DNA片段不会因重组发生嵌合问题

较易完整分离带有较少切点的大片段DNA

能使外源DNA保留在BAC系统内

可采取类似制备质粒的方法直接提取克隆的DNA，在大规模测序时便于机械化操作

**5.为什么STS绘图能成为大基因组物理图的主流技术？**

限制性作图虽然快速，并可提供详细的信息，但不适合大基因组。重叠群构建程序复杂，费时费力。指纹作图对大基因组仍存在不少问题，如选用的限制酶不当或重复顺序干扰，会出现误排。原位杂交因其操作困难，资料积累慢，一次实验定位的分子标记不超过3-4个

顺序标签位点或STSSTS（（sequence tagged sitesite））是一小段长度在100100到500 bpbp的DNADNA顺序，每个基因组仅一份拷贝，很易分辨。当两个片段含有同一STS顺序时，可以确认这两个片段彼此重叠。2个不同的STS出现在同一片段的机会取决于它们在基因组中的位置靠得多近。如果它们彼此邻接，这2个STS总会同时出现在相同片段上

**6.为什么说辐射杂种作图对于一些缺少遗传作图资源的摸式生物基因组图绘制具有重要意义？**

辐射杂种作图对于一些缺少遗传作图资源的模式生物基因组图绘制具有重要意义，特别是牛和马这类妊娠期较长的大生物，它们的单仔生殖只能提供有限的作图数据。其他如河豚鱼，因其繁殖困难不适宜经典遗传学分析，辐射杂种作图是一种理想的方法。许多哺乳动物，如猪、狒狒、斑马鱼、鸡都采用全基因组辐射杂种作图

**7.为什么说荧光原位杂交是一项重要的细胞分子生物学技术？它在哪些方面具有重要应用价值？**

重要性：能在成分复杂的组织中进行单一细胞的研究不需从组织或细胞中提取核酸，对含量极低的靶序列灵敏度高能准确反映组织细胞的相互关系及功能状态

方面：完整的染色体，荧光原位杂交方法是一种物理图谱绘制方法，使用荧光素标记探针，以检测探针和分裂中期的染色体或分裂间期的染色质的杂交。该技术不但可用于已知基因或序列的染色体定位，而且也可用于未克隆基因或遗传标记及染色体畸变的研究。在基因定性、定量、整合、表达等方面的研究中颇具优势。

**8.为什么克隆库的构建，指纹分析和分子标记的定位是构建物理图的关键环节？**

BAC库为物理图提供高质量的DNA原料，指纹分析克服了水稻基因组重复顺序带来的误差，分子标记使重叠群排成物理图

**9.哪些因素影响稀有酶切限制图绘制？**

一般而言，识别顺序越长产生的片段越。识别位点的碱基组成影响限制性片段的大小。有些限制酶识别位点较长，可采用变通的方法，将这类识别位点引入待测的基因组中。

**10.什么是DNA指纹？有哪些DNA指纹？**

指纹指确定DNA样品所具有的特定DNA片段组成。一个克隆的指纹表示了该克隆所具有的限定的物理特征，可以同其它克隆产生的同类指纹相比较。如果指纹重叠，表明这二个克隆具有共同的区段

DNA指纹包括：

限制性带型指纹

重复顺序 DNA 指纹

**11. 如何将重叠群锚定到遗传图上？**

锚标法：指用分子标记与全部22000个DNA片段杂交以获得阳性片段，再将全部阳性片段定位在重叠群上。每一个分子标记在整个基因组中只出现一次，也就是说，用某一个分子标记与整个水稻基因组DNA杂交，这个分子标记只能定位于染色体的某一个特定的位置上。

**第五章 基因组测序**

**1.DNA测序是结构基因组的基础，DNA测序有哪几种方法，它们各自的特点？**

1）链终止法：通过合成与单链DNA互补的多核苷酸链来读取待测DNA分子的顺序，合成的互补单链可在不同位置随机终止反应。链终止测序第一步是制备相同的单链模板DNA，然后将其与一小段称为引物（Primer）的寡聚核苷酸退火（anneal），形成双链后起始新链合成。反应由DNA多聚酶催化，底物是4种脱氧核苷酸。DNA多聚酶的要求：高酶活性；无5’→3’外切核酸酶活性；无3'→5'外切核酸酶活性

2）化学降解法：基本原理:在选定的核苷酸碱基中引入化学基团再经化合物处理使DNA分子在被修饰的核苷酸位置降解。4种核苷酸A，C，G和T均可分别修饰，分别降解，形成只差一个核苷酸的降解DNA群体。起始DNA样品为双链DNA，在测序前要将其转变为单链。每个单链的同一方向末端都结合了放射性同位素标记，可显示DNA条带的位置。可采用不同的方法获得不同的末端处理单链，由此读取所有碱基顺序

3）自动化测序：标准的链终止测序法依赖一系列的技术创新与集成/ 荧光标记物、毛细管电泳、计算机荧光标记物兼具灵敏度与分辨力的优点，便于仪器阅读，已成为测序的主流方法。其一，以荧光颜色为标记信号，每种ddNTP各有一种代表颜色。其二，整个反应在一个试管中进行。其三，当新合成的终止单链通过荧光监测仪时，可由光信号读出末端核苷酸并由电脑记录。1）免除了同位素标记必需同时进行四组反应的麻烦，简化为由一个泳道同时判读四种碱基；2）自动化荧光测序系统极大地提高了测序的效率，为基因组大规模测序提供了可能。3）避免肉眼分辨减少了差错，阅读信号与计算机相连后可直接对数据进行电脑处理，加快了基因组测序的进程。

4）非常规测序：、

i.光点测序：无须电泳或其它任何的分离程序，也不必加入ddNTP来终止反应，可以一直合成，因此比链终止法更加迅速。原理：当每一个配对的核苷酸加入到正在生长的新链末端时，根据发出的信号直接判读。脱氧三磷酸核苷酸连接到DNA 3DNA 3’’--末端时会释放一个焦磷酸（（PPiPPi），），焦磷酸在磷酸化酶的作用下转换为化学能，并发出光亮

ii.DNA芯片测序：将各种排列顺序的寡聚核苷酸点播在面积很小的芯片上，每个点播的寡聚核苷酸在排列的方阵中都有指定的位置。待测的DNA分子与芯片温浴，凡是能杂交的寡聚核苷酸都会在确定位置发出信号，然后根据获取的信息将寡聚核苷酸的顺序进行对比组装，拼接成完全的DNA顺序

**2.为什么对真核生物的基因组进行随机定向测序时要建立多个文库？**

采用多个文库是因为任何一种载体都会因某些插入片段与宿主菌的不兼容而不能扩增，使一些DNA片段丢失。不同类型的载体遭遇的不兼容是不相同的，因而第二种文库可以保留在第一种文库中可能失去的克隆片段。另外采用二种文库也增加了克隆片段的总长，扩大了覆盖面。第三，10 kb文库的构建有助于在序列组装时校正重复顺序产生的差错。

1. **随机测序与序列组装（鸟枪法）战略的优点和局限是什么？**

鸟枪法的主要优点在于它的速度以及无需提供相关的遗传图与物理图。

鸟枪法也有它的局限。如果基因组大太，结构过于复杂，序列组装的起始阶段工作量非常大。必须对所有已知小段顺序进行分析，找出共有序列组装成很小的重叠群。随着重叠群的扩大，比较的顺序会逐渐减少，但总的来说，需要分析的小片段数量仍达到现有计算机能力的极限

**4.人类基因组计划对了解人类自身和重要的生命现象具有极重要的价值，但也引发了大量伦理学争论，为什么会出现这种现象？你认为这种现象正常吗？**

**5.在人类基因组测序计划的竞争中，为什么Celera Genomics能后来者居上？**

Venter等人（1998）提出，一种建立在基因组图谱基础上的所谓“指导鸟枪法”（the directed shotgun approach）在人类基因组计划进入测序组装阶段是可以实行的。

Venter 设想在三个层次上进行:

1.建立插入片段长度约 150kb 的BAC 库，此库由300,000个克隆组成，对每一个BAC克隆进行两端测序，各端测出500bp。这样可以得到约10%左右的人类基因组序列，即平均每5kb就有一个500bp的序列标记。

2.用插入长度为 10kb 的低 拷贝质粒克隆，建立一个有 500 万

个克隆组成的库 。对每一个克隆进行两端各测出500bp，共得到50亿碱基序列。

3.用多拷贝质粒作载体建库 ，插入长度为2kb，库容量是30,000,000个。同样对此库进行两端测出500bp，共得到300亿碱基序列。所有的序列加起来，超过人类整个基因组DNA序列的10倍。每个层次是独立进行的，最后对所得到的DNA序列加以总装。

**第六章 基因组诠释**

**开放阅读框:** 它们由一系列指令氨基酸的密码子组成。开放读框有一个起点，又称起译密码，一般为ATG，还有一个终点，又称终止密码，分别为TAA，TAG和TGA，三者含义相同

**孤独基因：**当某一顺序从数据库中无法找到同源序列，又无法排除其不是基因的可能性时，必需依靠实验来进一步确认。在基因分类时这些缺少同源顺序的ORF被称为孤独基因（orphan gene）

**直系基因：**这是指不同物种之间的同源基因，它们来自物种分隔之前的同一祖先

**平行基因：**同一种生物内部的同源基因，它们常常是多基因家族的不同成员, 其共同的祖先基因可能存在于物种形成之后，也可能出现于物种形成之前

**动物园杂交：**一种确定人细胞DNA片段是否在亲缘物种中存在同源顺序的分析方法。对那些Northern杂交不易检测到的基因还可用另一种途径验证。一些亲缘关系相近的物种，其基因的编码区相似性较高，而非编码区的同源性很低。如果某一物种的DNA顺序与来自另一亲缘种的DNA片段杂交产生阳性信号，该区段可能含有一个或多个基因。

**一致性(identity):**指同源DNA序列的同异碱基位置上相同的碱基成员，或者蛋白质中同一氨基酸位置相同的氨基酸成员的比例，可用百分比表示

**相似性(similarity)：**指同源蛋白质的氨基酸序列中一致性和可取代氨基酸所占的比例。

**1.为什么说完成基因组测序仅仅是基因组计划第一步？**

完成基因组测序仅仅是基因组计划的第一步，更大的挑战在于弄清：

1）基因组顺序中所包涵的全部遗传信息

2）基因组作为一个整体如何行使其功能

**2.如何根据ORF 结构特点搜寻基因？**

所有编码蛋白质的基因都含有开放读框（open reading frames, ORFs），它们由一系列指令氨基酸的密码子组成。开放读框有一个起点，又称起译密码，一般为ATG，还有一个终点，又称终止密码，分别为TAA，TAG和TGA，三者含义相同。从DNA顺序中搜寻基因总是从第一个ATG开始，然后向下游寻找终止密码。在开始这项工作之前，我们并不知道DNA双链中哪一条单链是编码链，或称正（+）链，也不知道准确的转译起始点在何处。因此，每条链都有3种可能的读框，二条链共计6种读框ORF扫描的关键是终止密码在6种读框中出现的频率。如果DNA的碱基排列是随机的，并且GC含量为50%，则3个终止密码子——TAA，TAG和TGA—出现的平均机率为每43=64 bp一次。假如GC比大于50%，因终止子中AT比例高，则每隔100-200 bp才会出现一个终止密码。大多数基因的ORF均多于50个密码子，因此最可能的选择应该是ORF不少于100个密码子的读框。

**3.为什么说利用生物信息学的方法延伸EST序列是获得全长cDNA序列的重要方法之一？**

快速，无需实验操作

**4.为什么特定种属有特征性的密码子偏爱？**

避免密码子的相似性

对组织内有着大量tRNA的那一个密码子有明显的偏爱

某些特定基因的遗传密码的使用会影响到其表达

**5.通过同源性分析给出整个基因或其中某一区段功能的信息时，为什么以氨基酸顺序进行同源性比较其结果更为准确？**

由于组成蛋白质的氨基酸有20种，而DNA核苷酸只有4种，因此氨基酸顺序的差异要比核苷酸的差异大得多。以氨基酸顺序进行同源性比较其结果更为准确

**6.为什么剔除老鼠能成为分析人类基因的功能模式系统？**

老鼠是一种模式动物，因其基因组及其基因组成与人类相似，常被用于检测人体基因的功能

老鼠是多细胞生物，基因的功能常常同生长发育有关，具组织专一性。因此要观察某一基因的表型，需要获得完整个体。为了解决这一问题，可利用一类特殊的老鼠细胞，即胚胎干细胞（ES细胞）。将缺失代换获得的ES细胞注射到老鼠胚细胞中使之混合成为嵌合体，胚细胞继续发育长成的老鼠也是嵌合体。嵌合体老鼠相互交配，从其子代中可获得基因型纯合的缺陷个体。这些个体的每个细胞都携带失活的基因，称之为剔除老鼠（knockout mice）。观测剔除老鼠的表型即可获知失活基因相关功能的信息。该项技术适用许多基因的功能检测

**7.为什么说蛋白质组分析是一个远比与基因组测序和转录物组研究困难得多的任务？**

1.蛋白质有许多的加工方式；

2.由于mRNA的可变剪接、程序性移码和可控突变，一个基因可编码许多不同的蛋白质，常常表现为组织特异性；

3.蛋白质之间存在大量的相互作用；

4.一种蛋白质可参与多种反应，或多种蛋白质参与一种反应

8.人类基因组顺列已经完成，但编码蛋白质基因的准确数仍存在不同的看法，为什么？

同7

**第七章 染色体结构与基因表达**

**核小体：**核小体是染色质的基本结构单位，其核心8聚体与DNA结合可阻止转录因子的接触

**核小体相位：**指一段顺序确定的DNA与核小体核心8聚体的可变结合方式。也就是缠绕在核心组蛋白8聚体上的DNA顺序可向前或朝后移动若干碱基对，从而改变双螺旋大小沟的朝向

**位置效应：**因染色体不同区段的结构而影响基因表达的现象称为位置效应

**座位控制区：**位于一组基因的上游，可以控制染色质的结构变化的一段特定DNA顺序。座位控制区与其下游一个或一组基因共同组成独立的功能域

**绝缘子：**这种可以阻止邻近位置激活或失活效应的顺序现在称之为绝缘子

**后生遗传（表观遗传EPIGENETICS）：**影响表型并能通过体细胞遗传，但不改变细胞的基因型

**1.核小体通过哪些结构的变化来调节基因的表达？**

（1）改变核小体相位：平移定位，旋转定位。缠绕核小体的序列将会改变位置，从而使核小体之间连接的顺序发生移位。平移定位使核小体排除在某些顺序之外，使转录复合物能与启动子区接触并进行组装。转录因子为了启动基因表达，必须附着到DNA双螺旋的表面。旋转定位可将蛋白质结合的DNA基序暴露在向外的一侧，便于转录因子识别。核小体的定位方式在基因表达的微调控方面也扮演了关键角色，通过不断地调控多聚酶的接触来控制转录速率

（2）组蛋白通过酰基化、甲基化、磷酸化改变染色质的状态

组蛋白酰基化伴随着染色质活性开放

组蛋白去酰基化使染色质由开放状态转变为紧凑的状态而失活变为异染色质

**2.基因组采取何种方法来防止某些增强子不加区别地作用于其它的基因？**

这种可以阻止邻近位置激活或失活效应的顺序现在称之为绝缘子

绝缘子有增强子阻隔效应，将调控序列与启动子隔离开来。

以绝缘子界定增强子的作用范围是真核基因组一种普遍的控制基因活性的机制

**3.什么叫后生遗传？后生遗传形成的机理？**

影响表型并能通过体细胞遗传，但不改变细胞的基因型

机理：决定细胞类型的不是基因本身，而是基因表达模式，通过细胞分裂来传递和稳定地维持具有组织和细胞特异性的基因表达模式对于整个机体的结构和功能协调是至关重要的

DNA修饰 蛋白修饰 非编码RNA修饰

**4.高等生物基因组甲基化有哪些显著特点？**

1）范围广，人类基因组中绝大多数的CpG位点均被甲基化；

2）可逆性，许多甲基化位点可以根据细胞活性的要求重新甲基化或去甲基化,；

3）组织特异性，不同的组织细胞具有不同的甲基化模式，为基因表达设定程序**。**

**5.体细胞克隆中存在胎儿发育及成年个体形态或生理的异常，你如何解释这种现象？**

后天环境的不同，造成了基因的表达不同。体内环境等因素影响了基因的表达以及修饰，从而造成了个体形态以及生理上的异常。

基因组甲基化可能影响体细胞克隆个体的基因表达模式的异常。体细胞克隆中移植的体细胞核在胚胎发育时必须经历去甲基化及重新甲基化的过程，可能导致某些基因甲基化模式的改变，产生基因表达的不平衡。

**第八章 比较基因组**

**比较基因组学：**比较基因组学是基因组学与生物信息学的一个重要分支。通过模式生物基因组或模式生物基因组与人类组之间的比较与鉴别，为分离重要的候选基因，预测新的基因功能，研究生物进化提供依据

**相似性：**不同物种基因组的相似性是它们共同的祖先在渡过漫长的进化历程后留下的遗迹, 由此可以追踪物种的起源和分支路径。不同物种基因组的相似性首先来自于基因组成的相似性，这种相似性宏观上表现为近缘物种。

**基因共线性：**除了基因组成的相似性, 在不同基因组中基因排列顺序的一致性更能够体现基因组的共同起源, 这种基因排列顺序的一致性称为共线性(synteny, colinearity) 。

**微观共线性：**指物理图上基因顺序的一致排列。

**基因岛：**植物中某些区段基因密度比全基因组的平均密度高得多，形成“基因岛”

**直系同源集簇：**即指由一个共同的祖先基因衍生的一组基因，包括不同基因组中执行同一生物学功能的种间同源物(ortholog)，也包括同一基因组中因基因加倍产生的种内同源物(paralog)，或平行基因。

**基因协同进化：**首先是在进化中一部分基因缺失，且缺失的基因在功能上相关。与此同时，为了补偿丢失基因所执行的功能，导致其它具有类似功能的基因群高度分化。

**1.根据最少基因数目（核心基因）的推测，可以将核心基因联结起来而制造出人造生物吗？**

**2.为什么说比较基因组学是基因组学和生物信息学重要的分支?**

通过模式生物基因组或模式生物基因组与人类组之间的比较与鉴别，为分离重要的候选基因，预测新的基因功能，研究生物进化提供依据

**3.什么是基因共线性?哪些因素影响基因的共线性?**

除了基因组成的相似性, 在不同基因组中基因排列顺序的一致性更能够体现基因组的共同起源, 这种基因排列顺序的一致性称为共线性。包括基因的宏观共线性和微观共线性。破坏基因组共线性的因素很多, 包括转座、插入、染色体重排、区段加倍和缺失。

**4.为什么直系同源集簇在预测新基因的功能具有重要价值？**

直系同源集簇综合了比较基因组学, 种系发生学和蛋白质分类学的信息, 对预测新基因的功能具有重要意义。因为直系同源簇可进一步分成不同的亚家族，同一亚家族的直系同源集簇具有共同的功能域。而找到了新基因所属的直系同源簇，就可判断出其主要功能。

比较基因组学研究已经产生了两个基本的信息：(1) 大多数基因是进化上保守的，因而基因组比较可以提供很丰富的信息；(2) 基因组之间的关系和它们的进化轨迹非常复杂。基因组比较的结果显示出保守性和多样性之间的复杂相互作用。

**第九章 RNA组学**

**转录物组：**基因组在整个生命过程中所表达的全部转录产物的总和称为转录物组

**RNA组学：**对细胞中全部RNA分子的结构域功能进行系统的研究，从整体水平阐明RNA的生物学意义

**RNA编辑：**通过化学修饰改变转录物的编码性质，使其指令的蛋白质发生氨基酸顺序的变化，mRNA这类修饰称为RNA编辑

**超编辑：**涉及靶分子广泛的脱氨基，使RNA的50%腺嘌呤核苷酸转变为次黄嘌呤的一种修饰mRNA加工方式

**miRNA:** miRNA是一类约22nt的具有调控功能的小片段单链RNA。前体具有70-100 nt发夹结构

miRNA具有较高的保守性，存在于亲缘关系很近的物种一般miRNA 5’端带有单磷酸集团，3’端有羟基，是典型的Dicer剪切的产物。

**siRNA:** iRNA是一类约20~30nt的互补双链RNA，其5’端为单磷酸基团，3’末端是羟基基团，且3’端有2个突出的不配对的碱基。

**ncRNA：**DNA序列中专门转录成非编码RNA的部分称为RNA基因或是非编码RNA基因

**细胞中RNA的组成特点**

编码RNA约占总RNA的百分之四，非编码RNA占总RNA的百分之九十六，其中非编码RNA又包括rRNA前体，tRNA前体，snRNA，snoRNA，scRNA，tmRNA以及其他类型RNA。

**mRNA的加工事件及可能的功能**

mRNA的末端修饰：

加帽：使得mRNA从细胞核转移到细胞质，转译起始；

多聚腺苷酸化：RNA多聚酶转录终止，转译起始，mRNA周转

剪接：

前体mRNA/tRNA/rRNA剪接：出去内含子

切除事件：

前体rRNA/tRNA加工：从前体分子释放成熟rRNA/tRNA

化学修饰：

rRNA修饰：增加rRNA催化反应的范围

tRNA修饰：使氨基乙酰-tRNA合成酶识别tRNA，由一个rRNA识别2个或多个密码子

RNA编辑：使一个mRNA分子编码两个蛋白质，讲所写的转录物转变为功能RNA，使某些线粒体mRNA中产生终止密码子

**人类如何利用有限的基因成员来构建复杂的机体？**

mRNA前体的可变剪接可能是造成这一现象的主要原因。mRNA可变剪接极大地丰富了蛋白质组的多样性，而生物的表型最终是由蛋白质决定的。可变剪接增加了遗传信息的内涵。

**RNA酶（核酶）是什么？为什么RNA酶的发现者能获得诺贝尔奖？**

具有催化活性的RNA分子称为核酶（ribozyme）

核酶催化的生化反应有：

自我剪接:在内含子I型，II型和III型中已谈到这种加工能力；

催化切断其它RNA:在mRNA和rRNA的加工中切除内含子

合成多肽键:这是rRNA分子的重要功能之一

催化核苷酸的合成:在试管中合成的RNA分子已证明可以完成合成核糖核苷酸、RNA催化活性的发现解决了以往关于先有多聚核苷酸还是先有多肽链的两难困境，表明最初的生化系统整个地集中在RNA

**为什么转录组学需要独立出来作为一个研究对象？**

由于每个细胞都有全套遗传信息，但不同生长发育阶段、不同组织或不同环境胁迫或生理病理情况下的蛋白质谱是不同的。

脱氧核糖核酸与蛋白质的不完全对应性，一个基因并不对应一种蛋白质

细胞中转录产物的组成是随时间和空间的变化在不断地改变，包括数量的变化。

还需了解基因组作为一个整体如何工作，如何指令与协调细胞中各种不同的生化活性。人们已试图着手探明基因在不同组织中表达的模式，以及不同发育阶段基因表达的状态，特别是有关人类疾病基因的调控方式.

转录组学是研究蛋白基因转录的时空关系及其生物学意义

**可变剪接的生物学意义是什么？**

可变剪接增加了遗传信息的内涵：剪接位点的选择涉及复杂的机制，许多可变剪接的方式与组织特异性有关。人类遗传疾病中约有15%与mRNA的异常可变剪接有关。可变剪接反映了遗传信息流的动态变化与遗传信息在更高层次的重新组合，极大地扩充了遗传信息的内涵

**为什么ncRNA研究能成为生命科学研究的热点？**

ncRNA即mRNA之外的所有RNA，包括miRNA，siRNA等。这些RNA是非编码RNA，具有非常重要的调控作用。ncRNA可以调节和关闭基因的表达，进而调控细胞的各种高级生命活动。这不仅提升了人们对RNA分子的认识，还会大大推进基因功能的研究，更为各种传染病的防治、肿瘤等医学顽症的根治，开辟了一条新的有效的途径。

**第十章 蛋白质组学**

**1.为什么说蛋白质组比基因组更复杂？**

蛋白质组学比基因组学要复杂得多。基因组是相当稳定的实体，而蛋白质组通过与基因组的相互作用而不断发生着改变。一个生命体在其机体的不同部分以及生命周期的不同阶段，其蛋白表达可能存在巨大的差异

1）蛋白质有许多的加工方式，如磷酸化、糖基化、乙酰基化、泛素化和二硫键等，不仅形式多样，而且种类繁多，它们均与蛋白质的功能密切相关；

2）由于mRNA的可变剪接、程序性移码和可控突变，一个基因可编码许多不同的蛋白质，常常表现为组织特异性；

3）蛋白质之间存在大量的相互作用，如形成同源或异源二聚体、三聚体或多聚体，不同的结合状态有不同的活性；

4）一种蛋白质可参与多种反应，或多种蛋白质参与一种反应。蛋白质组学专家认为，由于转录、转译调控和转译后加工，蛋白质组的复杂性比基因组要扩大一个数量级

**2.蛋白质翻译后有哪些加工方式？**

磷酸化、糖基化、乙酰基化、泛素化和二硫键

**3.为什么说蛋白质水平的调控是生物活动重要的调节方式之一？**

通过对特殊蛋白修饰或改变蛋白的构象实现对基因表达的调控

**4.为什么要进行蛋白质组学的研究？**

􀂋 1.由于mRNA的可变剪接、程序性移码和可控突变，一个基因可编码许多不同的蛋白质，常常表现为组织特异性

􀂋 2.蛋白质有许多的加工方式，如磷酸化、糖基化、乙酰基化、泛素化和二硫键等，不仅形式多样，而且种类繁多，它们均与蛋白质的功能密切相关

􀂋 3.蛋白质之间存在大量的相互作用，如形成同源或异源二聚体、三聚体或多聚体，不同的结合状态有不同的活性

􀂋 4.一种蛋白质可参与多种反应，或多种蛋白质参与一种反应

**5.蛋白质组学应用了哪些新的技术？**

双相电泳技术（two dimensional electrophoresis gel 2D）􀂙

“双向”高效柱析法分子筛柱层析反向柱层析􀂙

质谱分析( MS):

辅助激光解析/离子化(matrix-assisted laser disorption/ ionization)电喷雾离子化（electrospray ionization ESI）􀂙

生物信息：数据库、检索系统、数据分析和图象分析软件

**第十一章 基因组进化模式**

RNA世界：生命进化的初期，所有生物学反应均以RNA为中心

核酶：具有催化活性的RNA分子称为核酶，它能催化的生化反应有自我剪接、催化切断其他DNA分子、合成多肽键、催化核苷酸的形成。

基因重排：

逆转录转座因子：分散在基因组中的重复序列，结构类似整合的逆转录病毒基因组，有些仍具有转座子活性。

假基因：由于核苷酸顺序变化而失活的基因。

不等交换：位于同源染色体上不同位置的相似核苷酸顺序之间发生的重组事件，其结果是在重组的区段产生一段DNA的重复。

多倍体：具有三个或三个以上染色体组的整倍体，即：三倍体及以上均称为多倍体。

有哪些证据支持RNA世界假说?

RNA世界是指生命进化的初期，所有生物学反应均以RNA为中心。

1. RNA功能很多；
2. RNA不仅可以是信息的携带者，而且还可以是功能的执行者；
3. 在人工模拟的原始地球的条件下，核糖核苷酸或多聚核糖核苷酸要比脱氧核糖核苷酸或多聚脱氧核糖核苷酸相对容易形成一些；
4. RNA能贮存遗传信息这一点也是无疑的。在现代生物中，仍然有少数病毒的基因组完全由RNA组成；
5. RNA一般以单链的形式存在，而单链的RNA可以折叠成多种多样的结构，这就为RNA可具有多种功能提供了结构上的基础；
6. 在原始环境下形成具某种功能的简单蛋白质较为困难。在这种情况下，把RNA看作是先于蛋白质的生物催化剂是可行的；
7. 单独由RNA组成的原始生命系统有进化潜力。

为何将现存生物分为三大类群?

比较所有生物的核糖体RNA序列所得结果，现存的细菌不像通常认为的那样属于同一家族，而是分为两个群体。将所有的生物分成古细菌界、真细菌界和真核生物界

基因组测序结果支持关于生命三界的假说

基因组主要通过什么方式增加基因的数目?

不等交换：位于同源染色体上不同位置的相似核苷酸顺序之间发生的重组事件，其结果

是在重组的区段产生一段DNA的重复。

姐妹染色体之间的不等交换：姐妹染色单体之间的不等交换涉及2个同源染色体的染色单

体之间不等交换

DNA放大：复制泡内二条子链DNA之间发生不等交换使一条链的某一段顺序加倍。

新基因的产生有哪些主要方式?

部分或全部现有的基因加倍:

1）整个基因组加倍；2）单条或部分染色体加倍；3）单个或成群基因加倍。

基因重排

从其它物种获取基因基因突变

**什么是外显子洗牌? 它在基因组进化中起什么作用?**

由不同基因中编码不同结构域的片段彼此连接形成的全新编码顺序称为功能域或外显子洗牌。它们有一个全新的结构组合，可为细胞提供完全不同的生物学功能。

可以产生创新功能的蛋白质。

**有什么方法可以判断基因组曾经发生过整体加倍?**

基因组的重复有些特别的标志。例如可以找到成片的重复基因，并且其中基因排列的顺序相同。对每个基因与其他基因的同源性进行比较，以鉴定基因的来源，是简单的基因重复还是基因组的加倍。要成为一对可能的重复基因，其所编码的氨基酸顺序至少要有25%的一致性。

**简述转座子对基因组进化的影响。**

转座因子对基因组的整体进化有多方面的影响，其中最重要的是转座子可引起基因组重排，如缺失、重复、倒位与移位。转座子引发重组的原因与其自身的功能毫无关联，只是因为同一转座因子的不同拷贝有相同的顺序，当它们位于同一染色体或不同染色体的不同区段时，可促使同源顺序之间发生重组。

在非编码顺序的扩张中逆转录转座因子扮演了重要的角色。逆转录转座因子与DNA转座子不同之处在于，前者的转以RNA为中介。通过转录产生的大量RNA可为逆转录转座提供合成cDNA的模板，由此发生爆发式的转座事件，使基因组急剧扩张。

从进化的角度观察，转座子对生物有潜在的价值，但从短期效果分析，转座子常常产生不利的后果，因为转座子插入到基因编码区导致基因的失活。正是基于这一原因，大多数细胞中转座子的活性因甲基化而受到抑制。

**基因组非编码DNA序列是基因组“垃圾”吗？**

**真核生物为什么能忍受如此之多的非编码DNA？**

非编码DNA可能具有某种尚未识别的功能，如果没有这些DNA细胞将不能存活。

基因的调控顺序虽然没有编码功能，但仍有重要的控制基因表达的作用。

大多数转座因子特别是逆转录转座因子含有控制基因表达的顺序，当它们插入到基因附近时有可能成为增强子的组成元件，为基因表达多样性提供分子基础。