

10 DNA 技术及生物信息学分析简介

王强

December 10, 2024

南京大学生命科学学院

Outline

10.1 DNA 技术

10.2 生物信息学分析

10.3 基因克隆技术

10.4 遗传工程的应用

10.5 遗传工程的风险和伦理学问题

生物技术 (biotechnology) 是在生物学理论的基础上建立的多学科交叉的应用技术科学

10.1 DNA 技术

DNA 的变性与复性

- DNA 变性: 较高温度解链成单链.
- DNA 复性: 变性的 DNA 逐渐冷却, 分离的两条单链 → 双链的 DNA.
 - ▶ 短链容易精确复性, 长链复性较难.
- 杂交分子: 碱基序列大部分互补, 可以复性.
 - ▶ DNA 与 RNA 之间, 也一样.

DNA 凝胶电泳

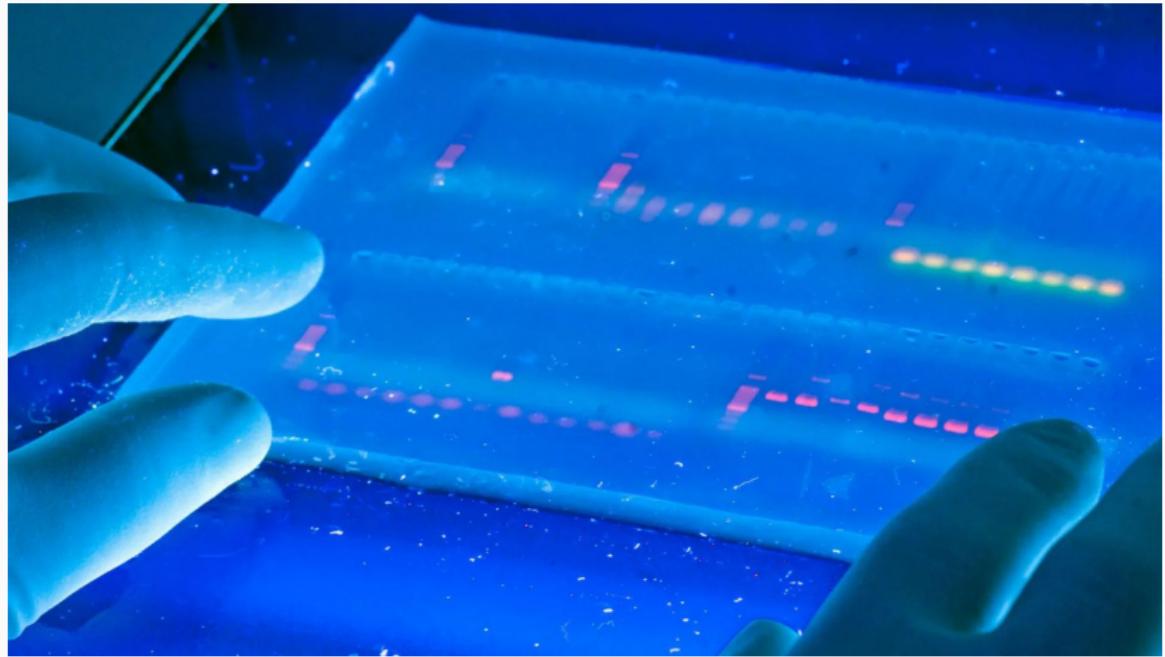


Figure 1. DNA 琼脂糖凝胶电泳

分子探针

- Southern 印迹和 Northern 印迹
 - ▶ Southern 印迹可以检测特定的 DNA 序列.
 - ▶ Northern 印迹可以检测特定基因的表达情况.
- 原位杂交

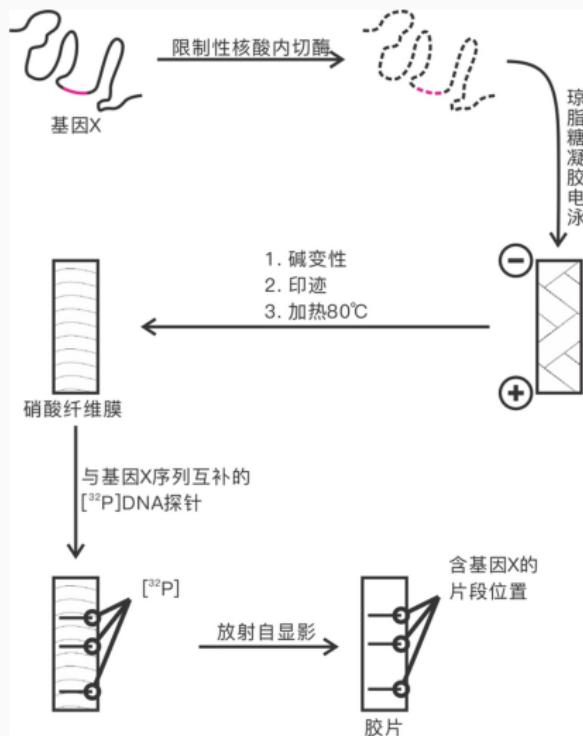
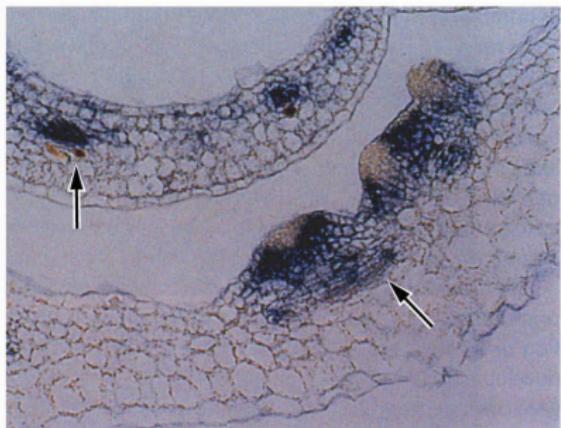


Figure 2. Southern 印迹

(A)



(B)



Figure 3. *KNAT1* 的表达 (原位杂交)

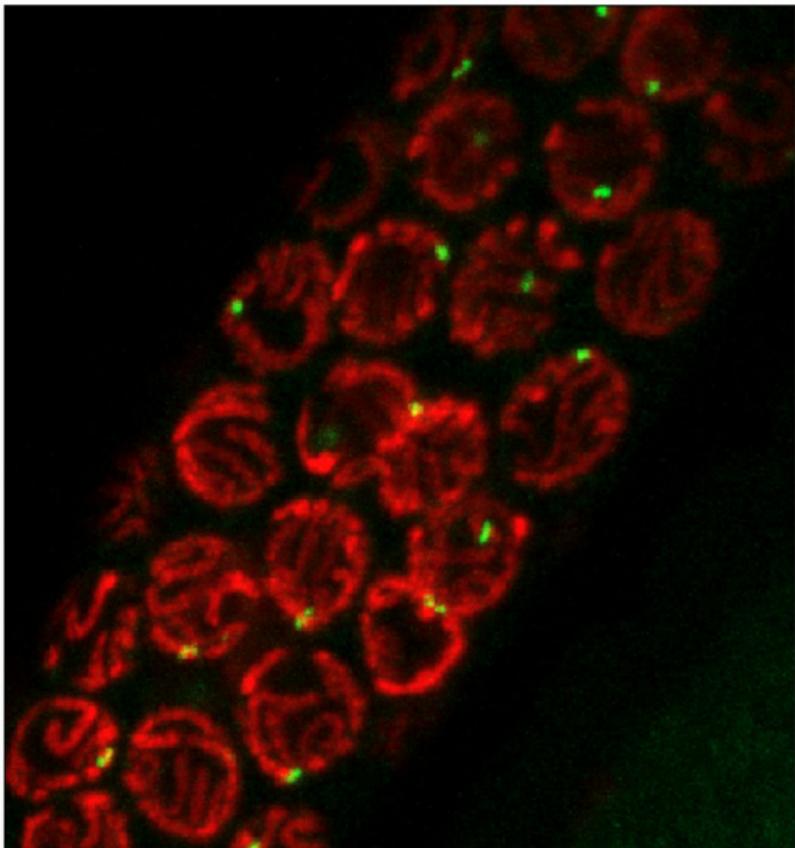


Figure 4. 荧光原位杂交

10.1.1 聚合酶链式反应



Figure 5. 凯利·穆利斯, 1993 年诺贝尔化学奖

1. PCR 技术的基本原理. PCR 反应过程:

- ▶ 双链 DNA 变性 (90–95 °C) 成为单链 DNA;
- ▶ 引物复性 (37–60 °C) 同单链 DNA 互补序列结合;
- ▶ DNA 聚合酶催化 (70–75 °C) 使引物延伸.

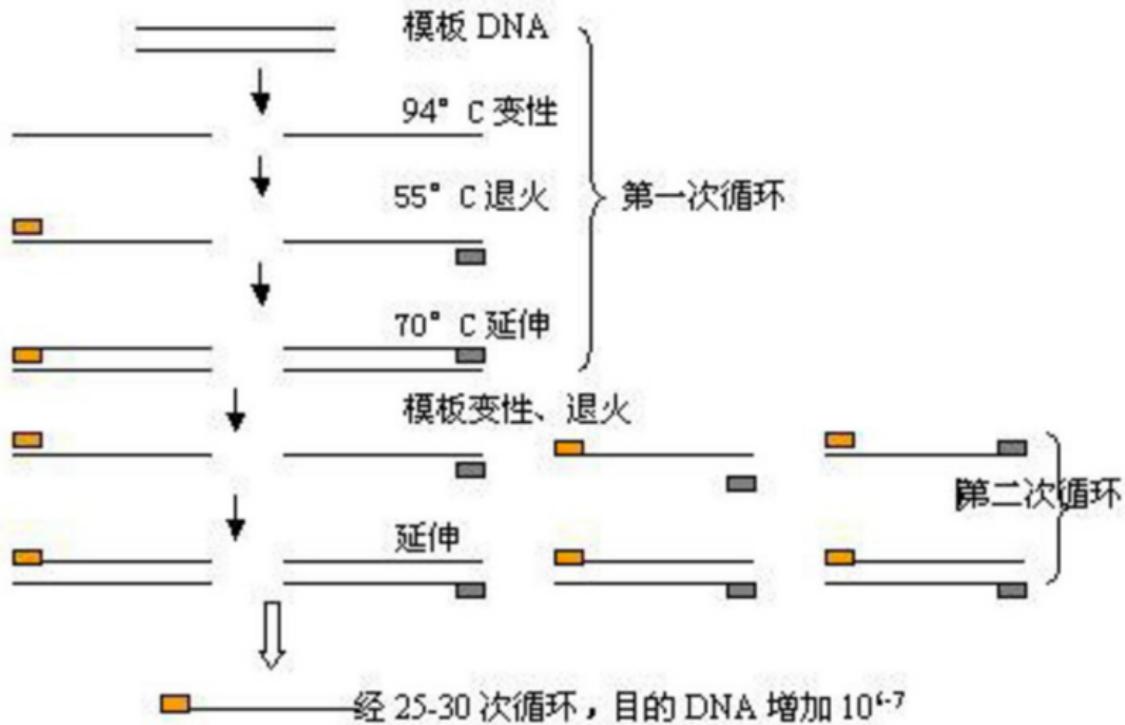


Figure 6. PCR 原理

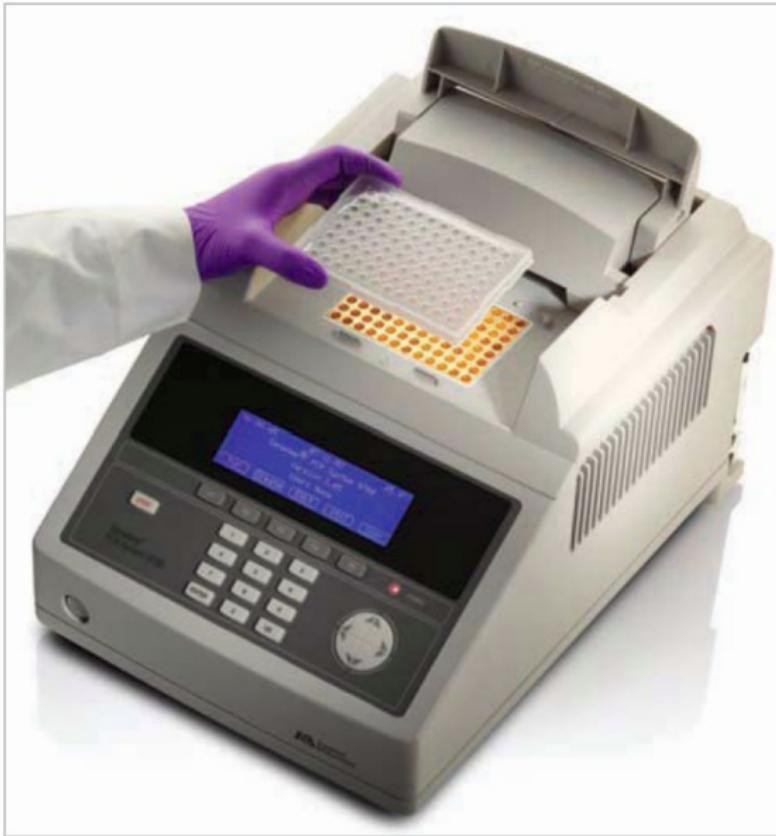


Figure 7. PCR 仪

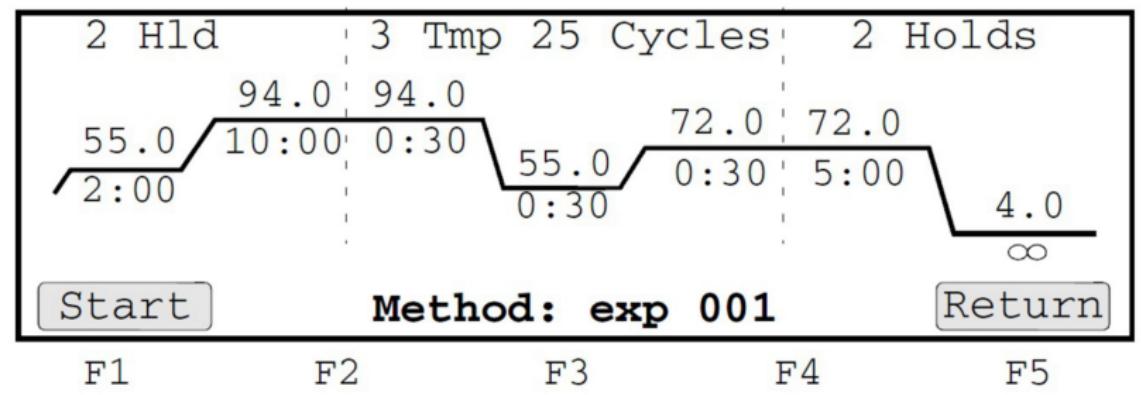


Figure 8. PCR 程序设定

2. Taq DNA 聚合酶
3. 寡核苷酸引物
4. 反应体系: 缓冲液, 离子浓度
5. PCR 技术的应用

10.1.2 DNA 测序技术

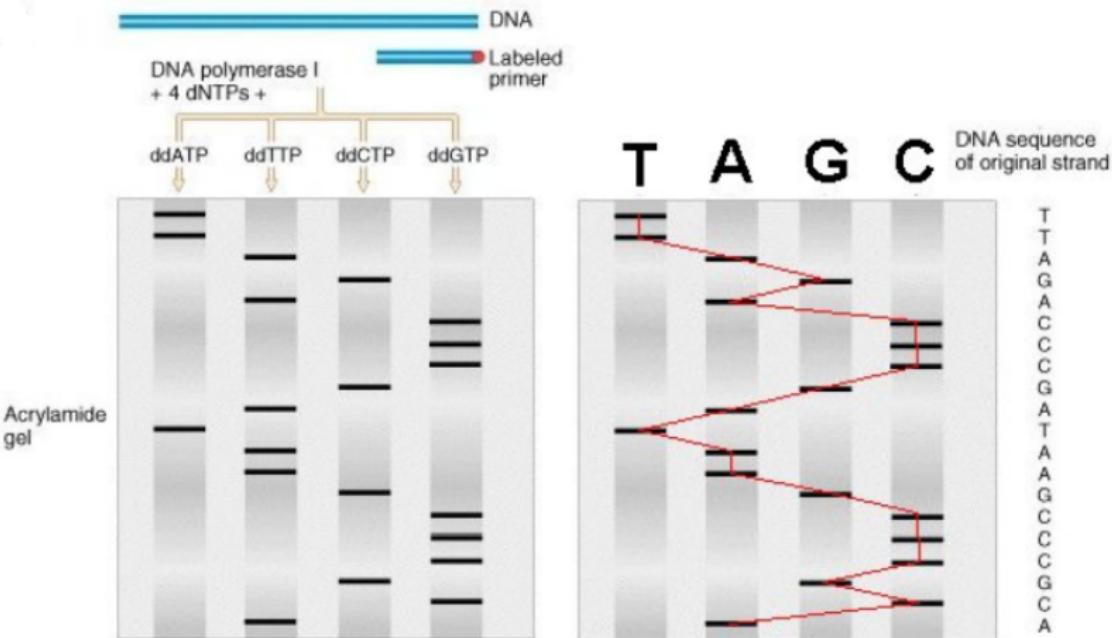


Figure 9. DNA sequencing ladder

10.2 生物信息学分析

10.2.1 序列分析

1. 序列比较

- ▶ 双序列比对
- ▶ 多序列比对
- ▶ 构建系统发育树

2. 序列识别与基因组注释

10.2.2 蛋白质序列和结构分析

- 基本性质
- 二级结构与结构域
- 三级结构

10.2.3 基因表达分析

- RNA-seq
- scRNA-seq
- 蛋白组

10.2.4 在其他领域的应用

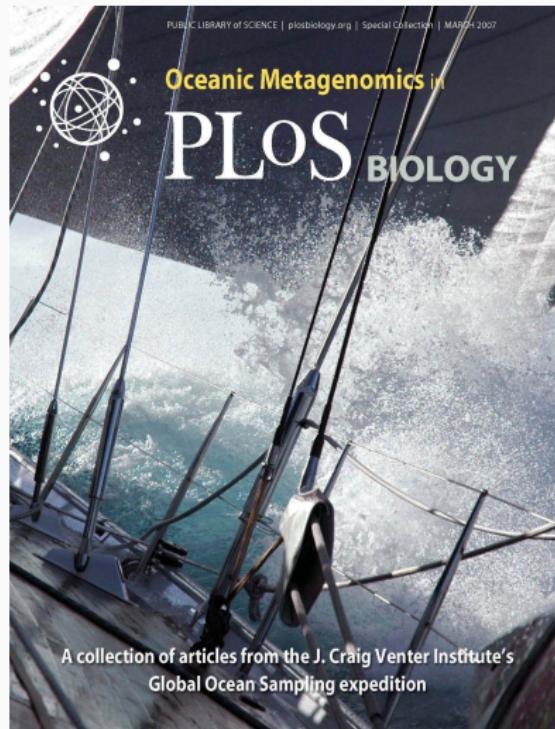


Figure 10. 宏基因组

10.3 基因克隆技术

- 重组 DNA 技术, 也称基因工程或遗传工程.
- 基因工程是指将特定的基因 (即外源基因), 通过载体或其它手段送入受体细胞, 使它们在受体细胞中与受体细胞的基因进行重组, 并能增殖表达, 这样的一种遗传学操作.
- 基因工程的目的: 通过与优良性状相关的基因的重组, 获得具有高度应用价值的新物种或新产品.

基因工程的优点

- 克服物种间的屏障;
- 有目的, 有计划, 有选择地加工制造各种生物制品;
- 遗传育种, 医学等研究和开发.

10.3.1 基因工程主要的工具酶

- 内切酶
- 连接酶
- 逆转录酶

限制性内切核酸酶

1. 限制性内切酶的作用

- ▶ 识别 DNA 中特定核苷酸序列, 使每条链的一个磷酸二酯键断开.

2. 限制性内切酶的类型

- ▶ I 型
- ▶ II 型 → 基因工程
- ▶ III 型

3. 限制性内切酶的命名

- ▶ 根据来源命名.
- ▶ 如 EcoRI → 大肠杆菌 (*E. coli*), R 株系, 第一种.

4. 限制性内切酶的识别序列

- ▶ 能识别的特定核苷酸序列
- ▶ 4–8 个碱基对组成, 且碱基互补对称
- ▶ 只写单链的核苷酸序列

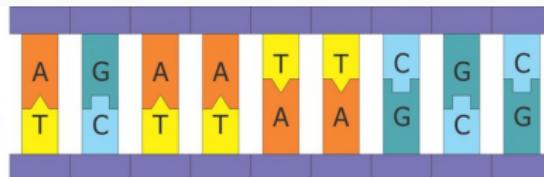
5. 限制性内切酶的切割位点

- ▶ II 型酶切割位点在识别序列区内

6. 切割片段的末端

- ▶ 粘性末端: 两条链末端交错对称.
 - 5' 粘性末端
 - 3' 粘性末端
- ▶ 平头末端: 两条链末端平齐.

DNA双链



限制性内切酶



Figure 11. 粘性末端

DNA 连接酶

催化 $-PO_4$ 和 $-OH$ 形成磷酸二酯键

1. *E. coli* DNA连接酶

- ▶ 大肠杆菌基因组编码;
- ▶ 连接具互补粘性末端的 DNA 片段.

2. T4 DNA 连接酶

- ▶ T4 噬菌体 DNA 编码;
- ▶ 既连接具互补粘性末端的 DNA 片段, 也能连接平头末端.

逆转录酶

- 从反转录病毒中制备得到的.
- 该酶能以具有 3'-OH 的 DNA 或 RNA 为引物, 以 mRNA 为模板从 5' → 3' 聚合生成 cDNA.

10.3.2 基因克隆的载体

基因载体, 运送外源 DNA 片段进入受体细胞.

三个条件:

- 有插入位点
- 能在受体细胞内复制
- 有筛选标记基因

质粒

1. 存在于细菌, 蓝藻, 绿藻, 真菌等.
2. 染色体外裸露环状双链 DNA 分子, 小的不足 1500bp,
大的 100kb 以上.
3. 宿主细胞内能自主复制. 松弛型和严紧型复制质粒. 选
用分子小和松弛型复制的质粒.

质粒载体 pBR322

1. 有复制起始点, 能在受体细胞内复制;
2. 有2种筛选标记基因;
3. 有允许外源 DNA 插入的位点;
4. 有高的拷贝数.

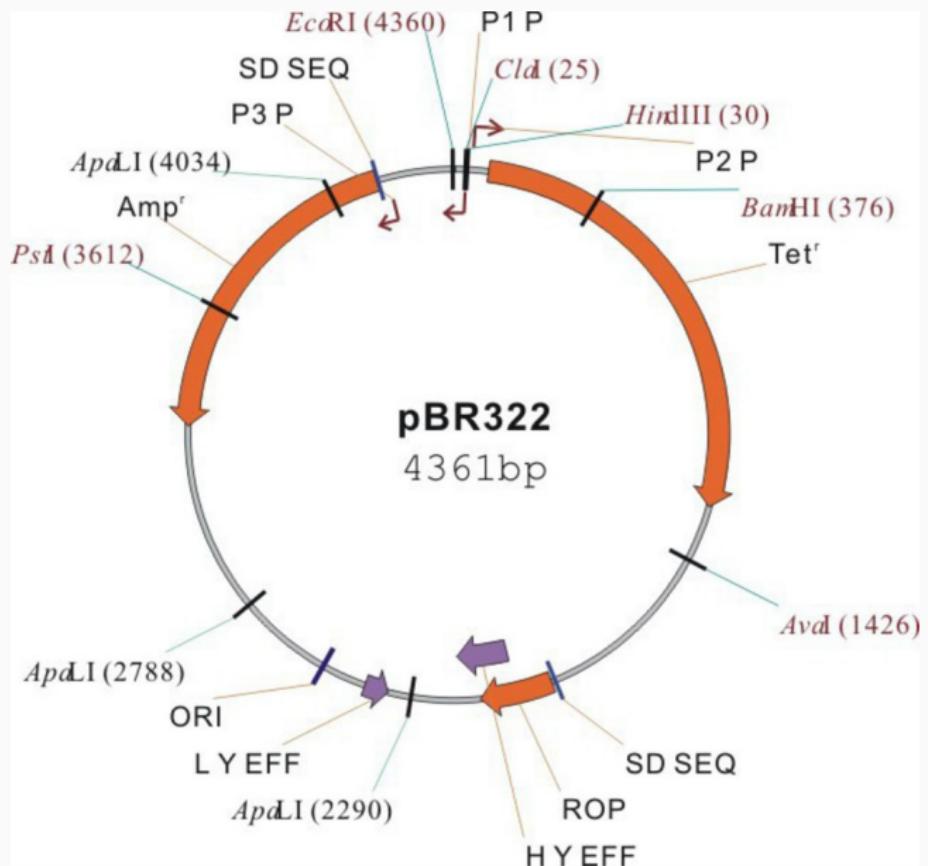


Figure 12. pBR322 图谱

10.3.3 重组 DNA 的基本步骤

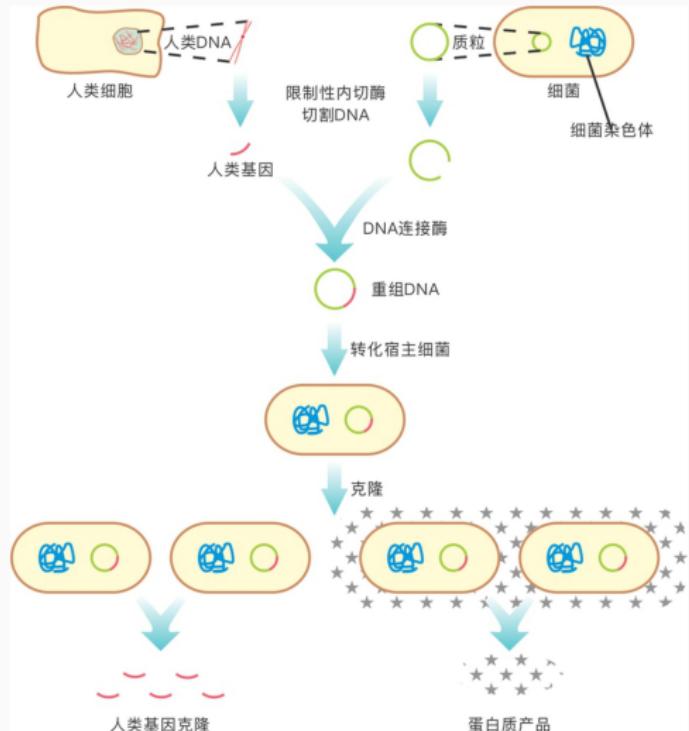


Figure 13. 重组 DNA 技术

获得目的 DNA 片段

1. 限制性内切酶酶切产生待克隆的 DNA 片段
2. 人工合成 DNA
3. 反转录酶酶促合成法
 - ▶ cDNA
4. PCR 扩增特定的基因片段

DNA 分子的体外重组

酶切和连接.

引入宿主细胞和筛选鉴定

1. 重组 DNA 引入宿主细胞

- ▶ 原核生物细胞是很好的受体细胞
 - 容易摄取外界的 DNA
 - 增殖快
 - 基因组简单
 - 便于培养和基因操作
- ▶ 大肠杆菌, 蓝藻, 农杆菌等

2. 重组体克隆的筛选与鉴定



Figure 14. 蓝白斑筛选

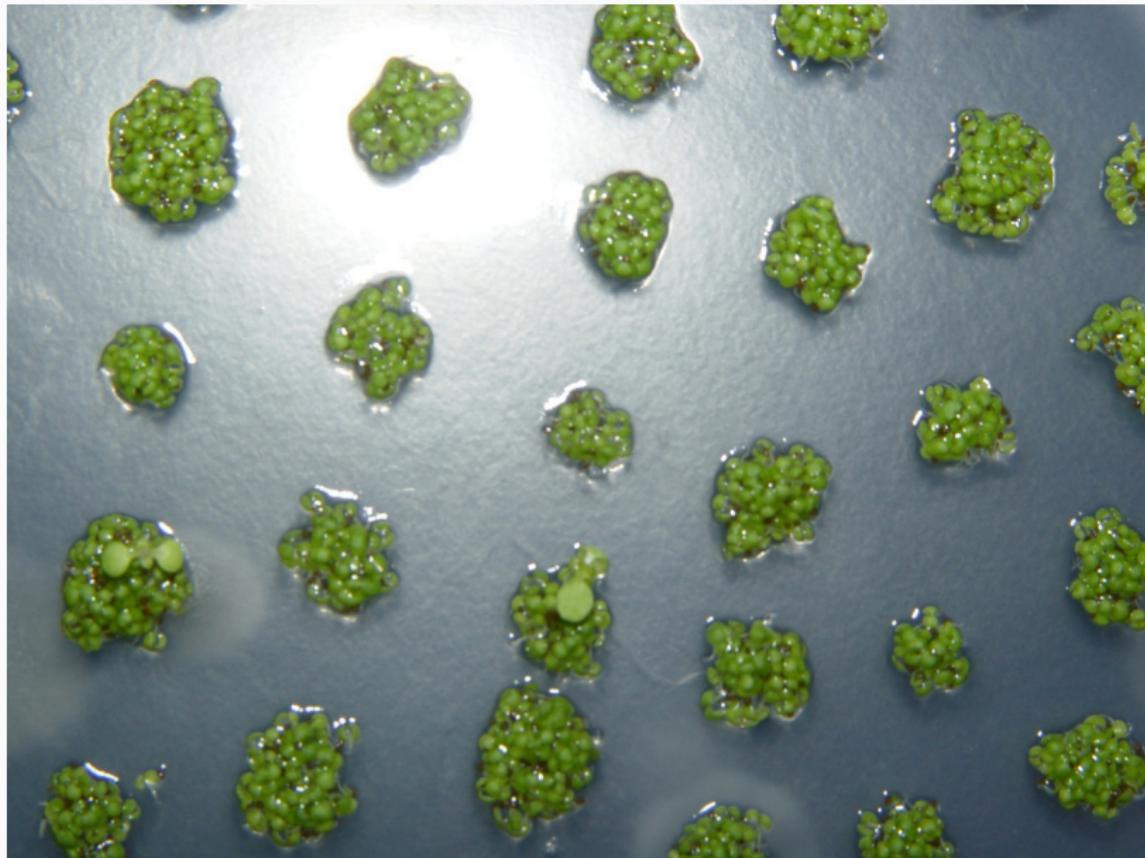


Figure 15. 转基因植物筛选

10.4 遗传工程的应用

10.4.1 健康医疗

- 胰岛素
- 生产新型疫苗
 - ▶ 蛋白
 - ▶ mRNA
- 生长激素

基因诊断和基因治疗

- 1990 年, 首例应用基因治疗, 治愈一名四岁女孩的腺苷脱氨酶缺乏症. 采用逆转录病毒转移腺苷脱氨酶基因.
- 基因治疗基因: 单基因疾病基因
 - ▶ 导致腺苷脱氨酶缺乏症,镰刀形贫血病等; 肿瘤抑制物基因和肿瘤形成基因等.
- 腺病毒和单疱疹病毒等
 - ▶ 囊性纤维化病, 帕金森氏病, 艾滋病和癌症.
- 基因诊断
 - ▶ 精准医疗

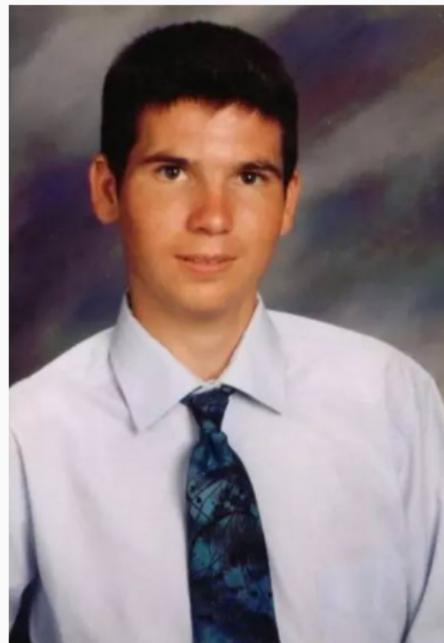


Figure 16. Jesse Gelsinger, 1981–1999

10.4.2 动植物基因工程

动物基因工程

■ 模式动物

- ▶ 模式动物可用来揭示生物学困难领域中的许多奥妙, 像人脑, 免疫系统和胚胎发育等.
- ▶ 在试验遗传病的新疗法中, 模式动物也很有用.
- ▶ 癌鼠, 转基因猴.

■ 生物反应器动物

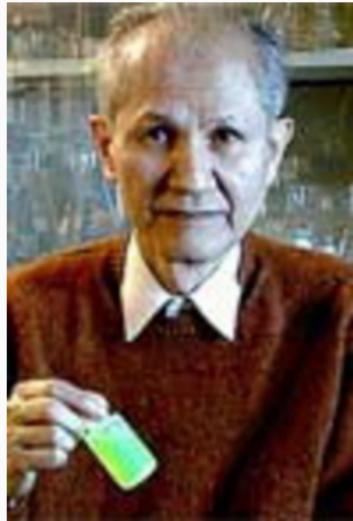
- ▶ 生物反应器动物: 利用其乳腺分泌药用蛋白质来制药的转基因动物.
- ▶ 羊 β -乳球蛋白启动子, α -抗胰蛋白酶.

■ 供体动物

- ▶ 英国科学家在 1992 年成功培育出转基因猪, 其心脏带有人类的成分; 猪心脏来代替人心脏用于移植手术.



Figure 17. 转生长素小鼠



(a) 下村修



(b) 马丁·沙尔菲



(c) 钱永健

Figure 18. 2008 年诺贝尔化学奖, 绿色荧光蛋白 (GFP)

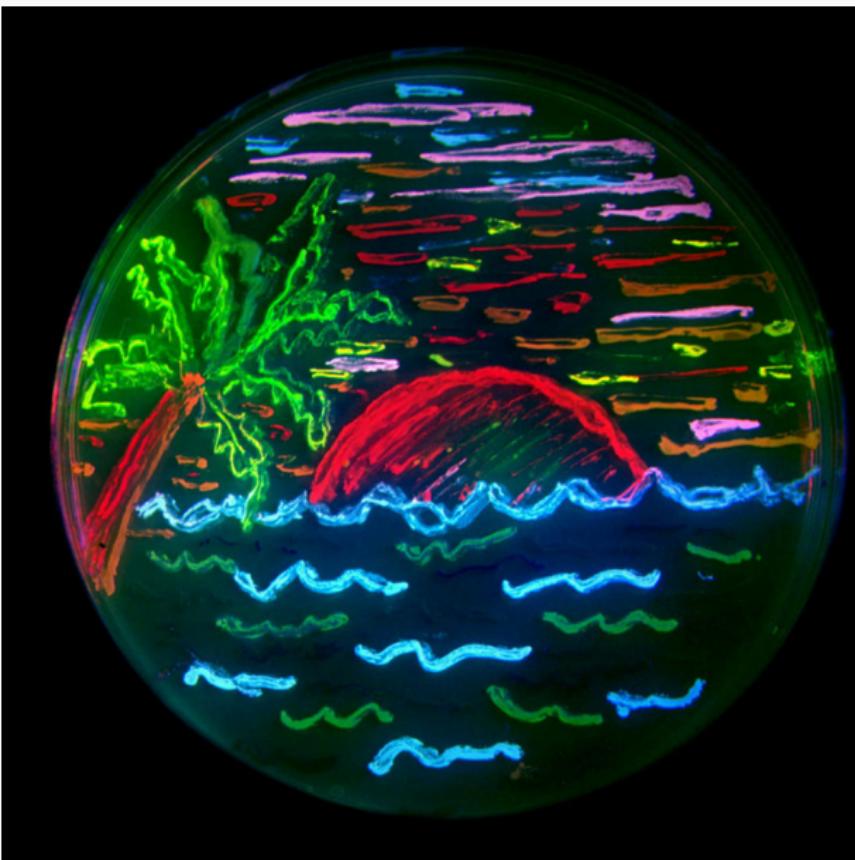


Figure 19. GFP and RFP



Figure 20. YFP

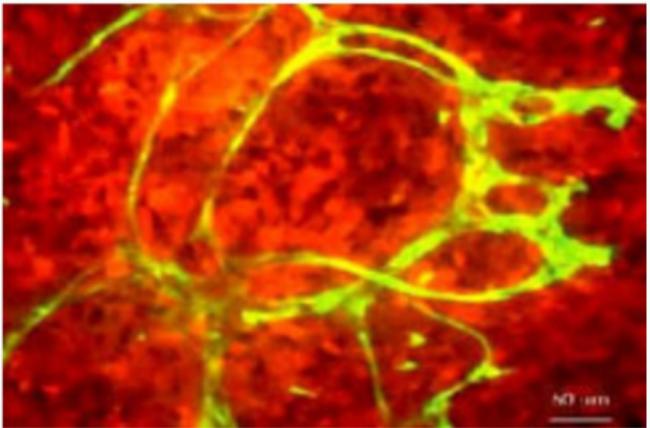
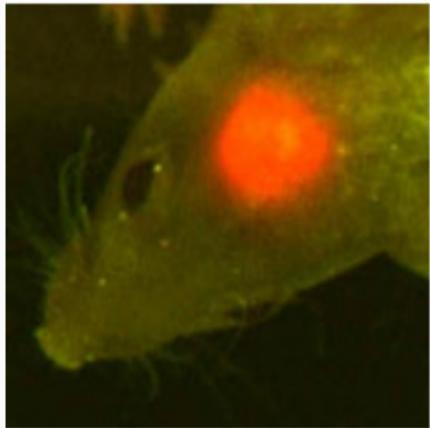


Figure 21. 转 GFP 裸鼠

转基因植物

■ 抗虫植物

- ▶ 苏云金杆菌, 毒蛋白.

■ 抗除草剂植物

- ▶ 草甘磷是一种广谱除草剂. 它的靶位点在叶绿体中的 EPSP 合成酶. 由于阻断芳香族氨基酸的合成, 植物最终会死亡.

■ 改良药用植物

- ▶ 日本科学家用重组 DNA 技术提高了镇静药茛菪碱合成的效率.

■ 生产疫苗的植物

- ▶ 土豆生产疫苗.

Chemical Synthesis of Poliovirus cDNA: Generation of Infectious Virus in the Absence of Natural Template

Jeronimo Cello, Aniko

Full-length poliovirus complementary oligonucleotides of plus and minus strand cDNA was transcribed by RNA polymerase and replicated in a cell-free extract infectious poliovirus. Experiments in transgenic mice confirmed that the genic characteristics of poliovirus synthesize an infectious agent by in by following instructions from a written

Generating a synthetic genome by whole genome assembly: ϕ X174 bacteriophage from synthetic oligonucleotides

Hamilton O. Smith, Clyde A. Hutchison III, Cynthia Pfankoch, and J. Craig Venter*

Institute for Biological Energy Alternatives, 1901 Ross

Contributed by J. Craig Venter, November 3, 2003

We have improved upon the methodology and time required for accurate assembly of DNA from synthetic oligonucleotide methodology. We have established conditions for assembly of three synthetic oligonucleotide genomes— ϕ X174 (5,386 bp) from a single pool of chemically synthesized oligonucleotides. The procedure involves the purification of pooled oligonucleotides to remove monomers and dimer contaminants in the pool under stringent annealing conditions against annealing of molecules with incorrect assembly of ligative products into full-length genome cycling assembly. A measurement of reannealing frequency can be extended to the full-length molecule. We observed a discrete assembly upon gel analysis of the polymerase product, without any PCR amplification. PCR thermal cyclization of the genome formed full-circularization and infectivity measurements had a lower infectivity than natural DNA, indica one lethal error per 500 bp. However, fully infected cells were recovered and incorporated into E . coli for analysis of several infectious properties of these synthetic genomes. One such intended sequence. We propose to assemble joining separately assembled 5'- to 4-kb segments would be required for a minimal cellu-

Complete Chemical Synthesis, Assembly, and Cloning of a *Mycoplasma genitalium* Genome

Daniel G. Gibson, Gwynedd A. Benders, Cynthia Andrews-Pfankoch, Evgeniya A. Denisova, Holly Baden-Tillson, Jayshree Zaveri, Timothy B. Stockwell, Anushka Brownley, David W. Thomas, Mikkel A. Alzaga, Chuck Merryman, Lei Young, Vladimir N. Noskov, John L. Glass, J. Craig Venter, Clyde A. Hutchison III, Hamilton O. Smith*

We have synthesized a 582,970-base pair *Mycoplasma genitalium* genome. This synthetic genome, named *M. genitalium* JCv-1.0, contains all the genes of wild-type *M. genitalium* G37 except MG408, which was deleted by an integrase marker insertion. The genome was also cloned into *Escherichia coli*. To identify the genome as synthetic, we inserted "watermarks" at intergenic sites known to tolerate transposon insertions. Overlapping "cassettes" of 5 to 7 kilobases (kb), assembled from chemically synthesized oligonucleotides, were joined by *in vitro* recombination to produce intermediate assemblies of approximately 24 kb, 72 kb ("1/8 genome"), and 144 kb ("1/4 genome"), which were all cloned as bacterial artificial chromosomes in *Escherichia coli*. Most of these intermediate clones were sequenced, and clones of all four 1/4 genomes with the correct sequence were identified. The complete synthetic genome was assembled by transformation-associated recombination cloning in the yeast *Saccharomyces cerevisiae*, then isolated and sequenced. A clone with the correct sequence was identified. The methods described here will be generally useful for constructing large DNA molecules from chemically synthesized pieces and also from combinations of natural and synthetic DNA segments.

RESEARCH ARTICLE

genome, we needed to establish convenient and reliable methods for the assembly and cloning of much larger synthetic DNA molecules.

Strategy for synthesis and assembly. The native 580,070-base pair *M. genitalium* genome (1) (*Mycoplasma genitalium* G37 ATCC 33530 genome ID: NC_002567; 2) (L5987; 3) was partitioned into 101 cassettes of approximately 5 to 7 kb in length (Fig. 1) that were individually synthesized, verified by sequencing, and then joined together in stages. In general, cassette boundaries were placed between genes so that each cassette contained one or several complete genes. This will simplify the future deletion or manipulation of the genes in individual cassettes. Most cassettes overlapped their adjacent neighbor by 80 bp; however, some segments overlapped by as much as 360 bp. Cassette 101 overlapped cassette 1, thus completing the genome.

Short "watermark" sequences were inserted in cassettes 14, 29, 39, 55 and 61. Watermarks are inserted or substituted sequences used to identify or encode information into DNA. This information can be either in noncoding or coding sequences (10–12). Most commonly, watermarking has been used to encrypt information within coding sequences without altering the amino acid sequences (10, 11). We opted to insert watermark sequences at

Figure 22. 全合成基因组

10.5 遗传工程的风险和伦理学问题

1. 对人的影响
 - ▶ 超级细菌
 - ▶ 对宗教, 习俗和生活方式的影响.
2. 对环境的影响
 - ▶ 转基因的逃逸
 - ▶ 超级细菌
 - ▶ 超级杂草
 - ▶ 对生物多样性的影响
3. 严格的释放规定
 - ▶ 生物安全实验室规则

罗马教廷公布 新七宗罪

- 滥用药物
- 有道德争议的科学实验
- 环境污染
- 制造贫困
- 社会不平等和不公义
- 基因改造
- 大肆聚敛财富