



中国科学院水生生物研究所

INSTITUTE OF HYDROBIOLOGY, CHINESE ACADEMY OF SCIENCES

分析测试中心

一代测序技术应用及常见问题解析



尹思



分析测试中心



027-87530080



2020-07-31



报告目录

一、一代测序原理

二、一代测序技术应用

三、常见问题解析



中国科学院水生生物研究所

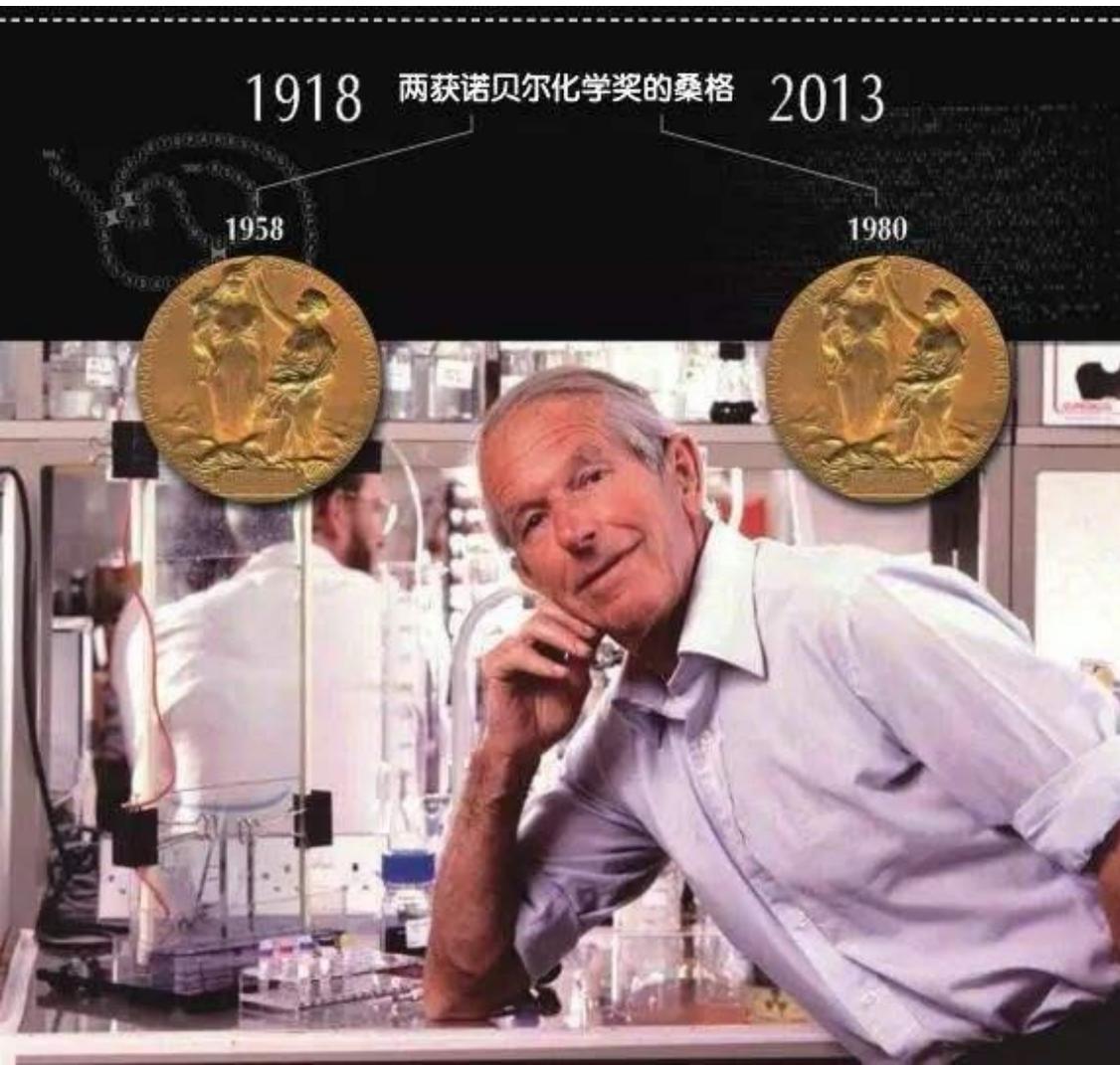
INSTITUTE OF HYDROBIOLOGY, CHINESE ACADEMY OF SCIENCES

分析测试中心

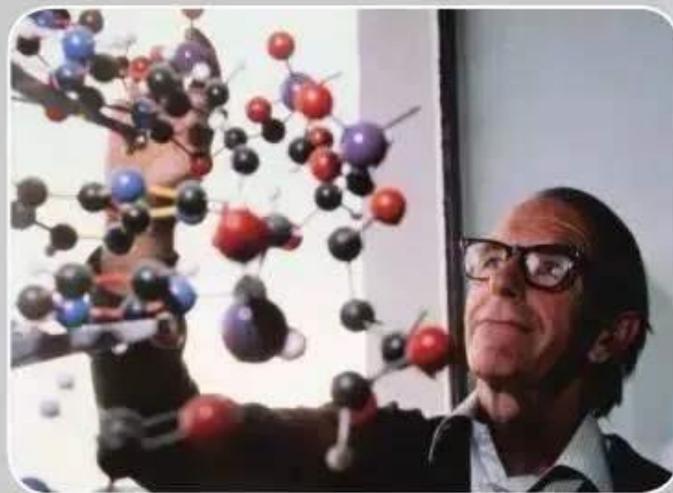
一代测序原理

“现代基因组学之父” - 两次获得诺贝尔奖的科学家

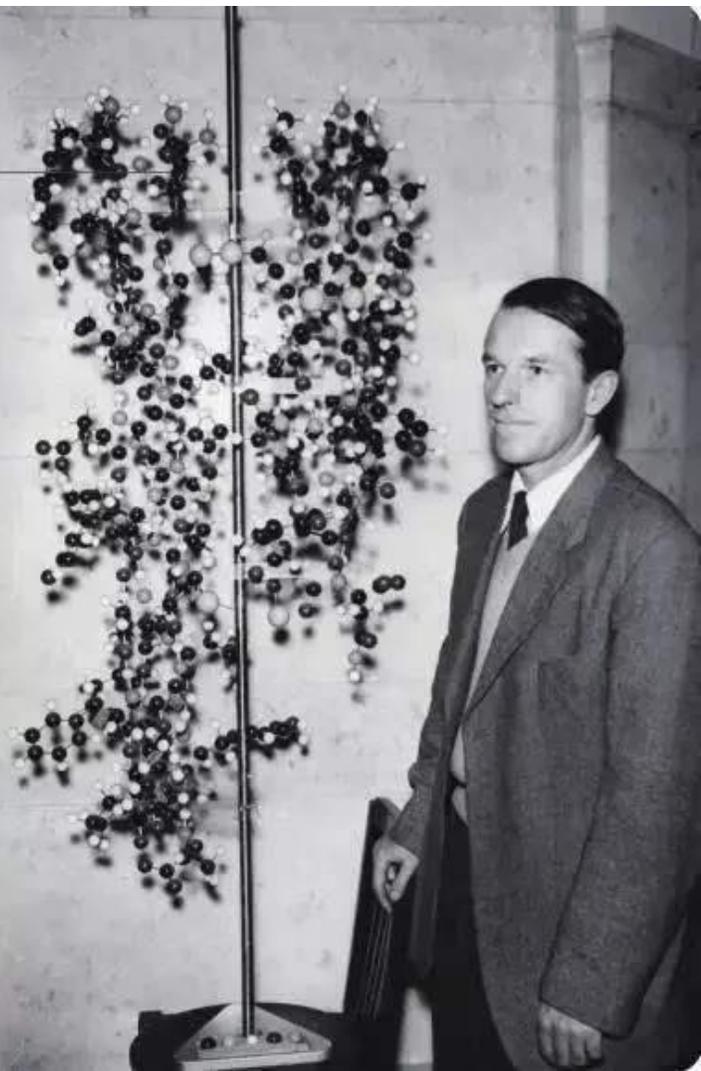
英国生物化学家弗雷德里克·桑格 (Fred Sanger) 分别获得1958年和1980年诺贝尔化学奖。他是同一领域内两次获奖的第二人，更关键的是，两次获奖理由都可归结为：**测序**。1951年，Fred Sanger首次确定了两种蛋白质（牛胰岛素A和B形式）的序列，表明它们是截然不同的分子，为此，他于1958年获得了诺贝尔化学奖。1977年，他引入“双脱氧终止法”，也被称作“桑格法”，因此与另外两名科学家共同获得了1980年诺贝尔化学奖。



1958年，桑格博士站在英格兰剑桥实验室的胰岛素原子模型结构旁边。



桑格在看一个DNA分子模型



Sanger测序原理

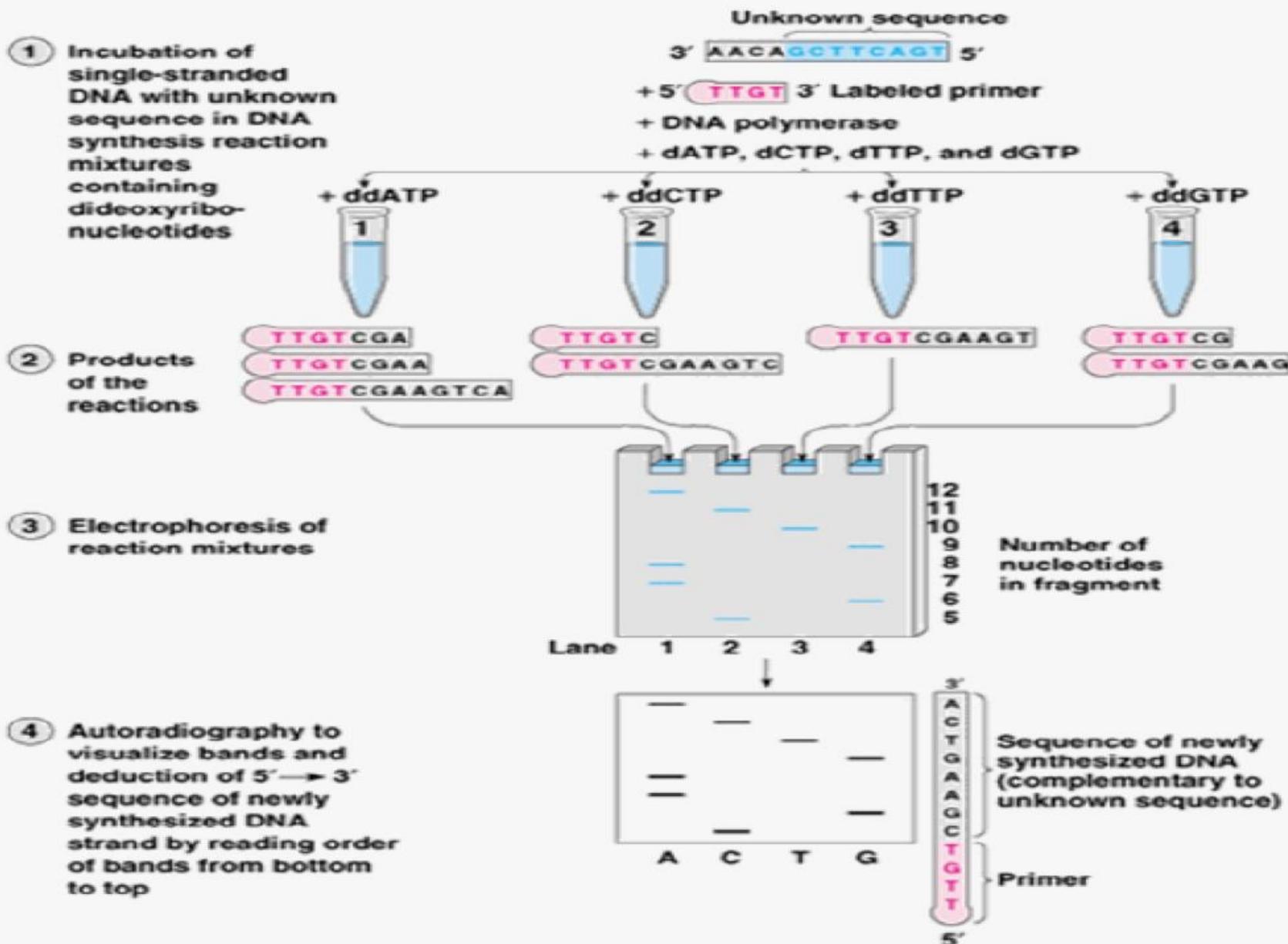


Dr. Fred Sanger

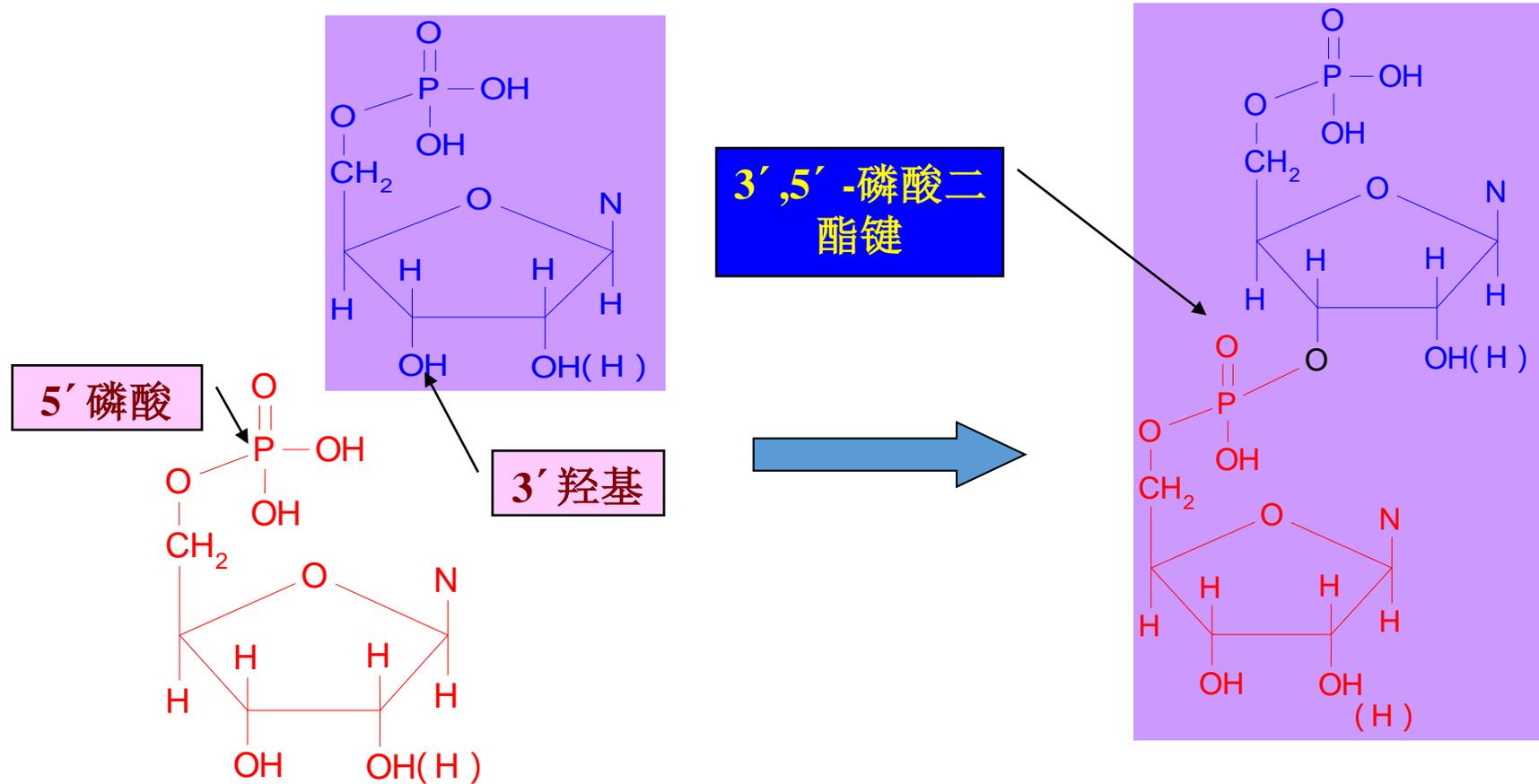
Frederick Sanger was awarded the prize in both 1958 and 1980. He is the fourth person in the world to have been awarded two Nobel Prizes and the only person to receive both in chemistry.

"dideoxy" sequencing technique (Sanger et al., 1977)

DNA双脱氧链终止法测序

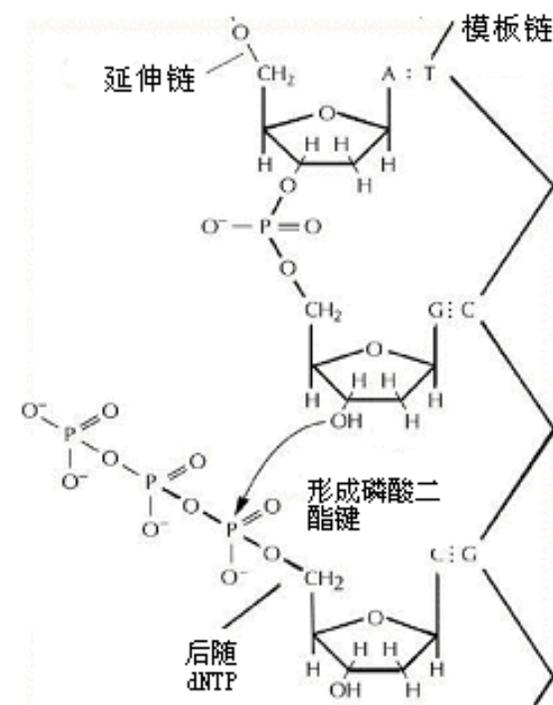
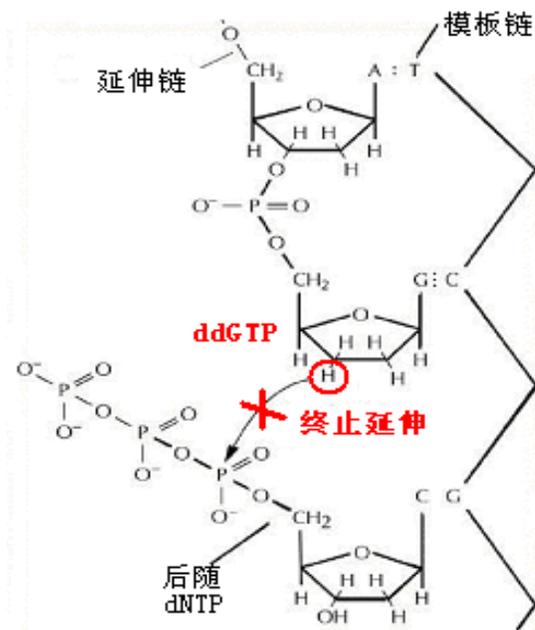
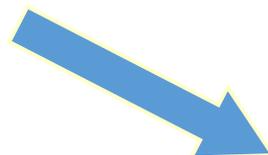
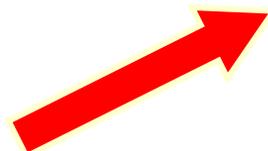
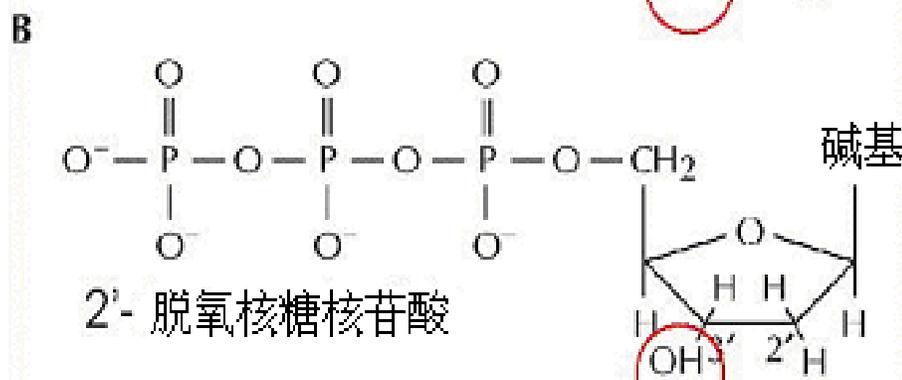
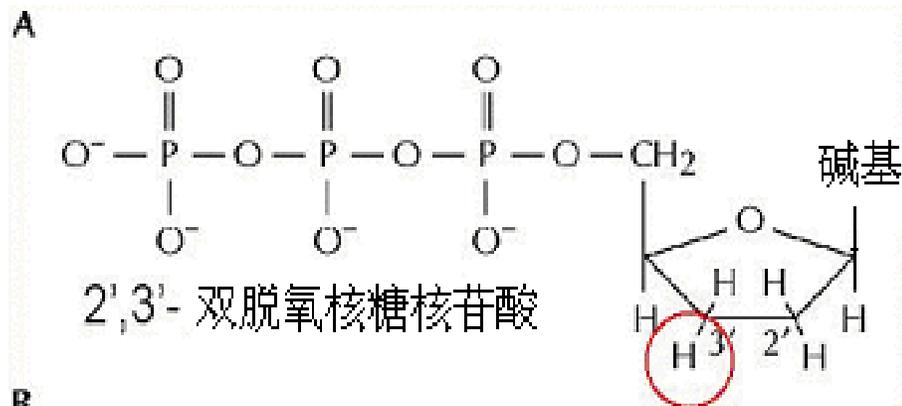


核酸序列形成的基础

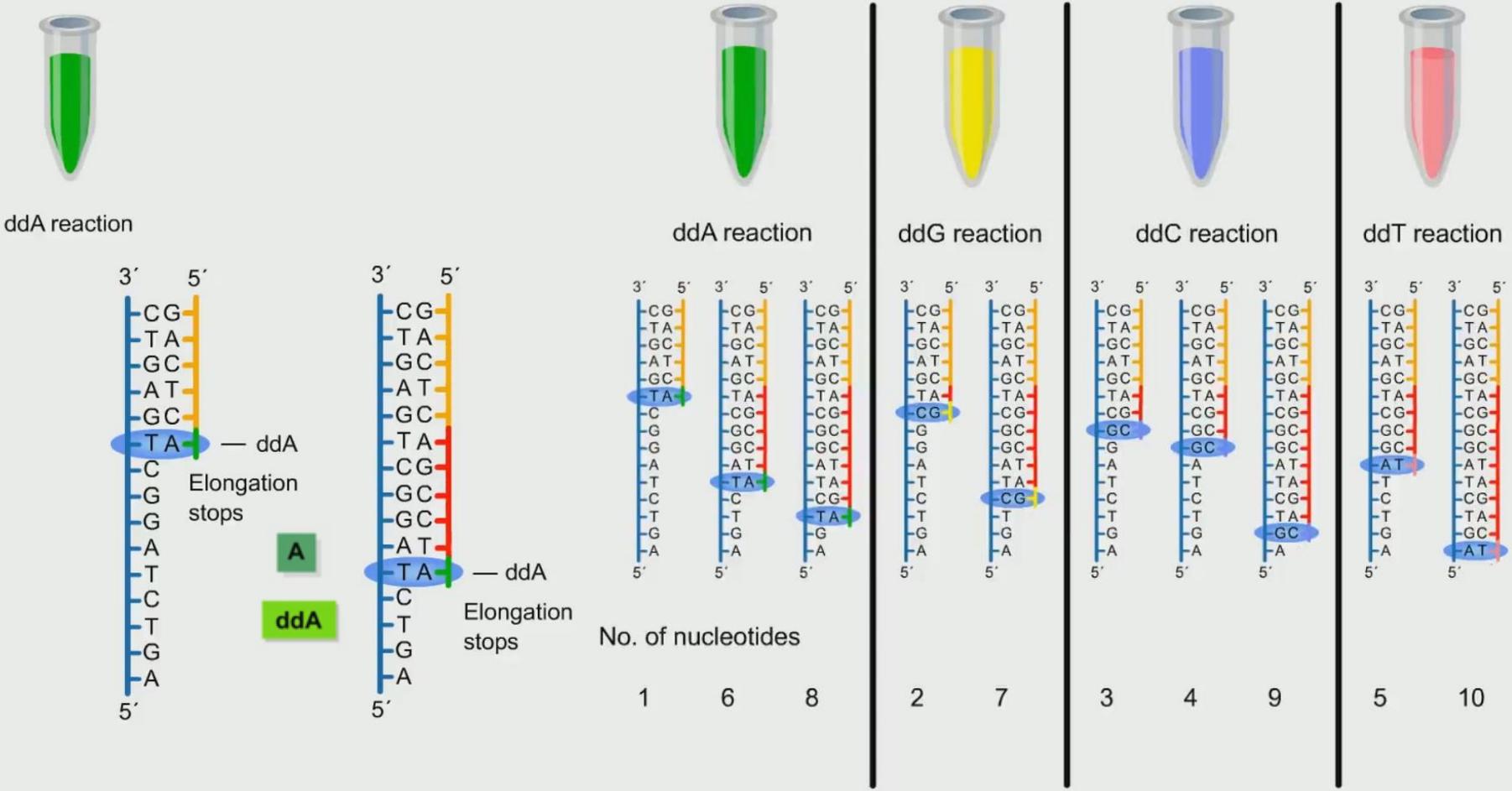
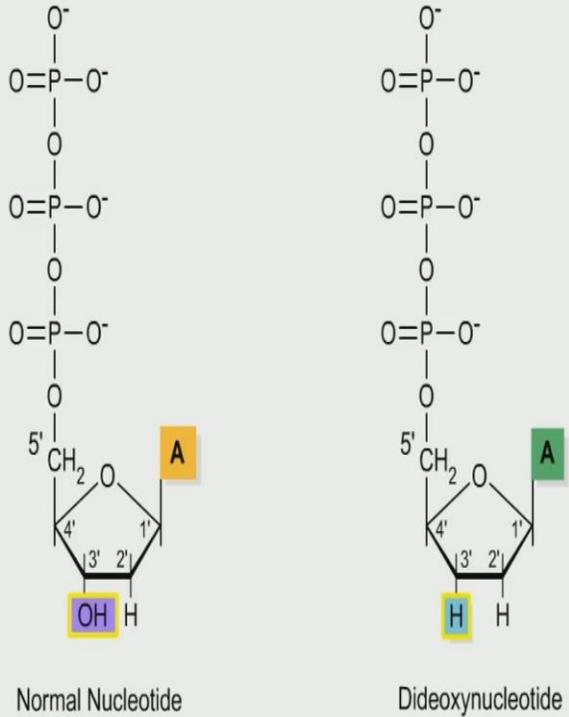


核苷酸形成3',5'-磷酸二酯键示意图

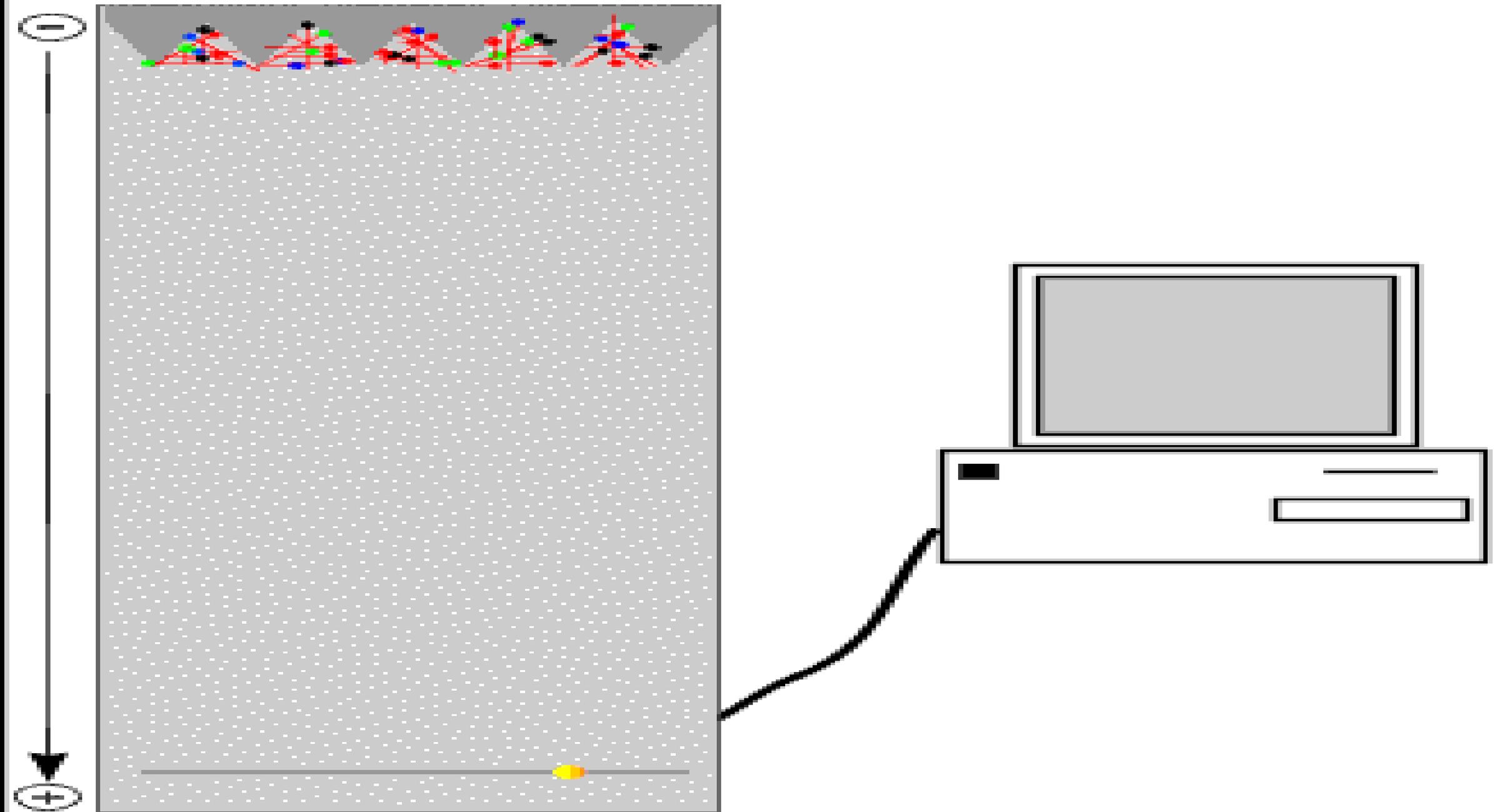
“双脱氧末端终止”的含义



双脱氧末端终止法原理

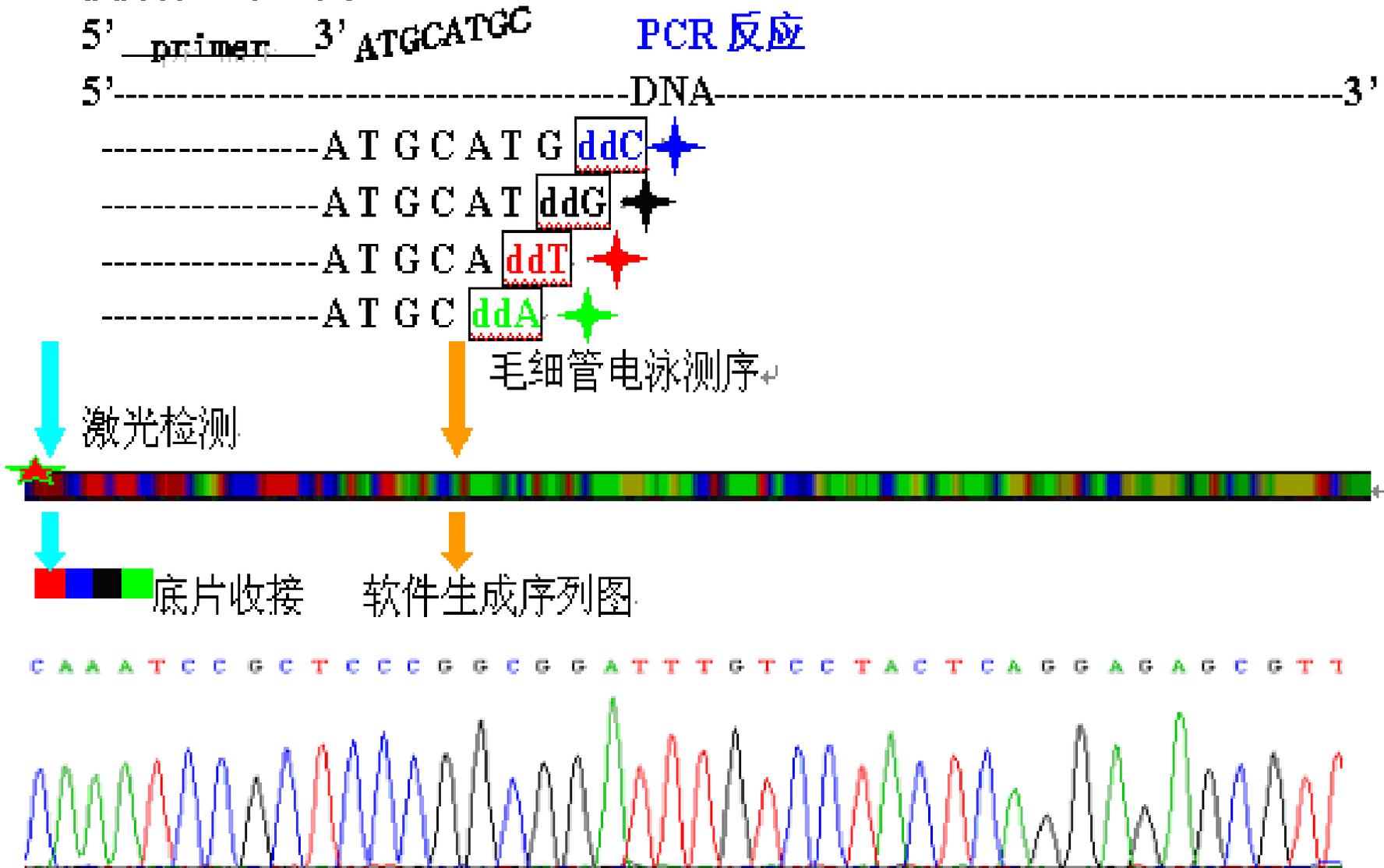


Gel electrophoresis



Sanger测序原理

测序原理示意图:

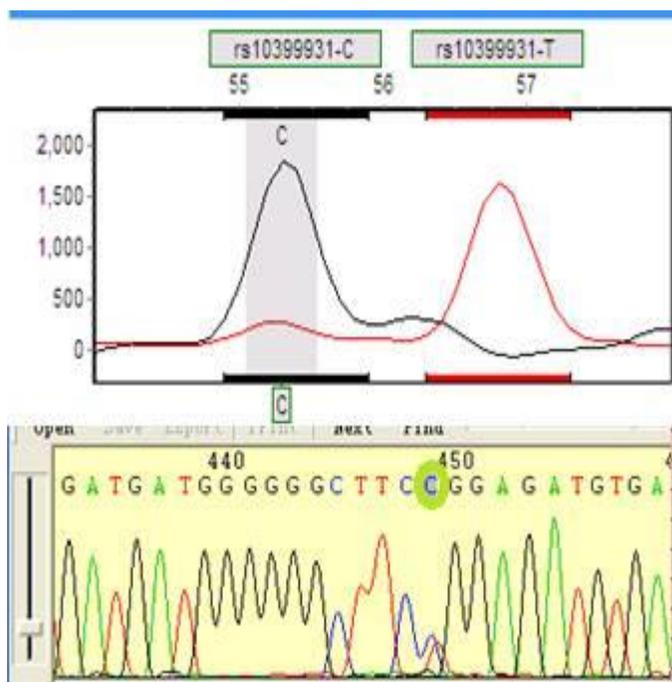


- ddNTP被加到正在合成的链上，由于3'-OH已被氧化，下一个dNTP的5'-磷酸基团不能与之形成磷酸二酯键，使延伸终止
- 在引物延伸反应过程找那个，ddNTP在不同位置加入，产生一系列不同长度的新的DNA片段

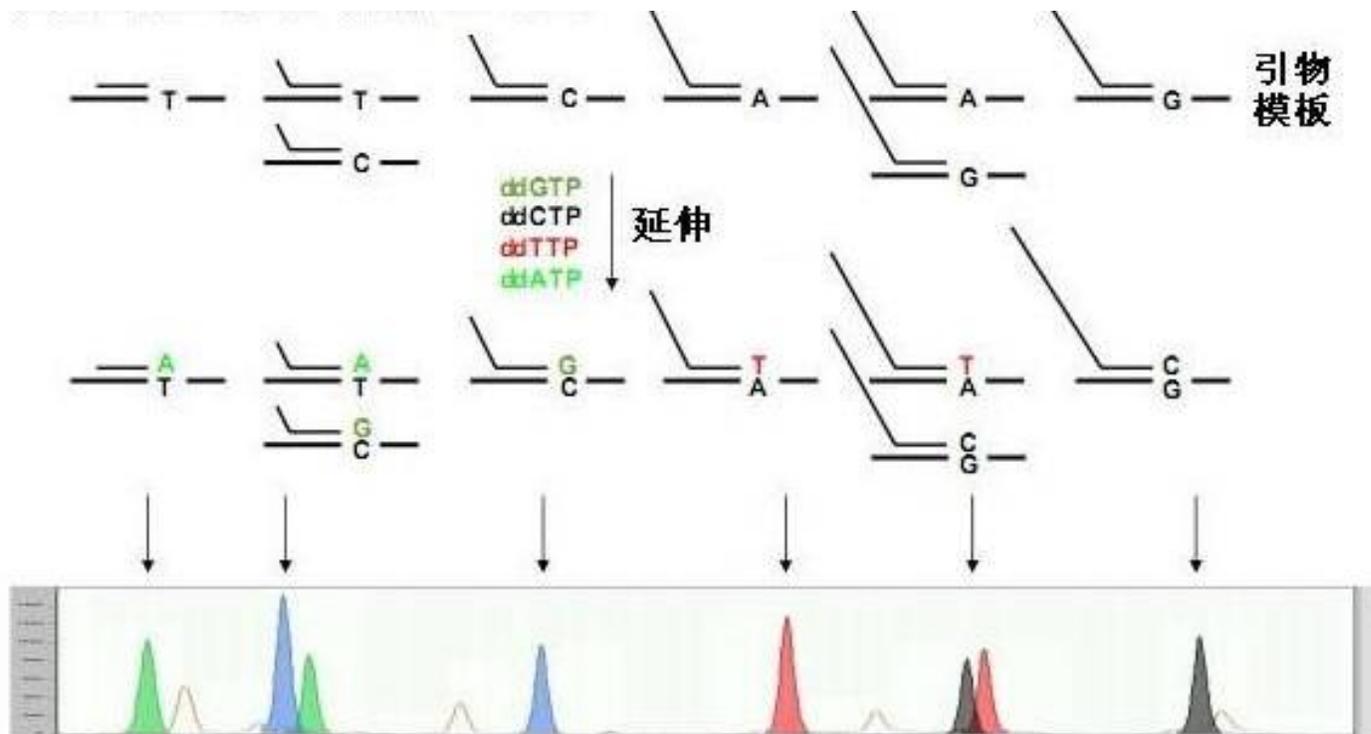
SNaPshot小测序原理

原理：即**单碱基延伸技术**。原理与Sanger测序类似，不同的是，应用SNaPshot进行定点的序列分析时，PCR反应中只有不同荧光标记的ddNTP而无dNTP。每个SNP位点的引物的**3'端都紧靠SNP点**，在聚合酶作用下，根据模板的序列，**只延伸一个核苷酸**。然后用先进的荧光检测系统，检测延伸的那个核苷酸的种类，针对不同的SNP位点设计不同长度的延伸引物，可以做到多个SNP在一个反应体系中进行分型。

SNaPshot分型结果



Sanger法分型结果



毛细管电泳结果

测序分析仪的发展



3730xL

96

仪器



3730

48



3500xL

24



3500

8



3130xL

16



3130

4



310

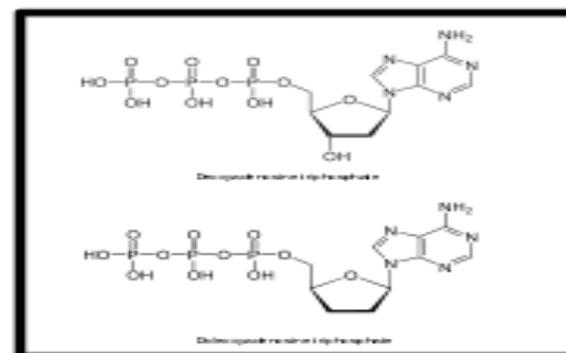
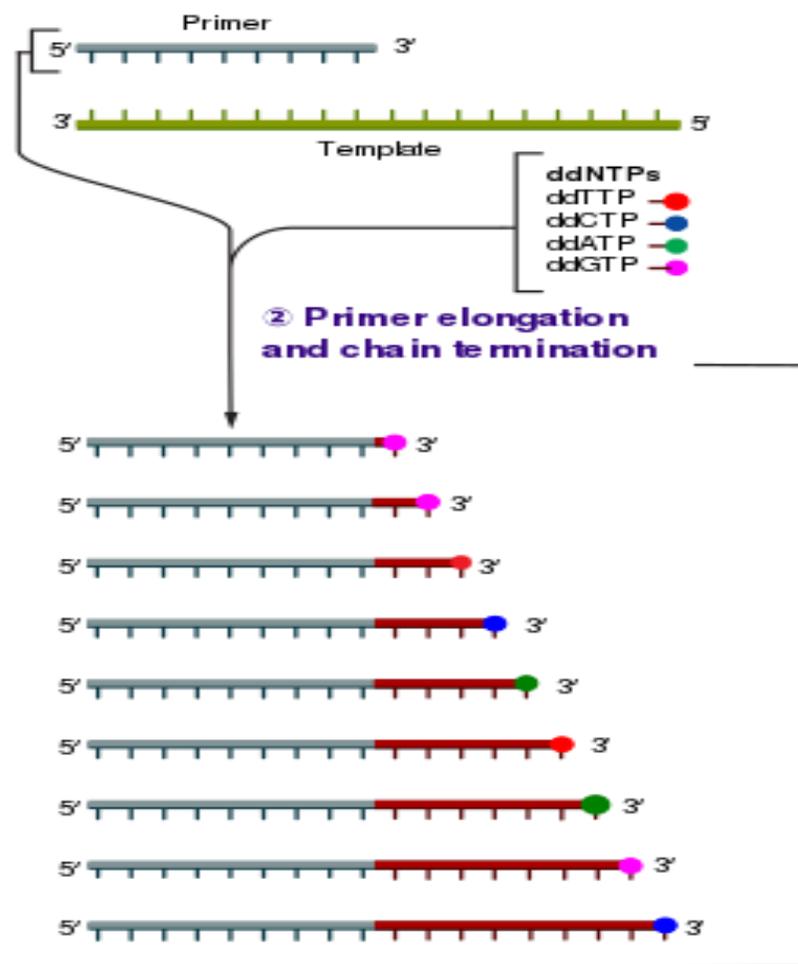
1

Number of
Capillaries

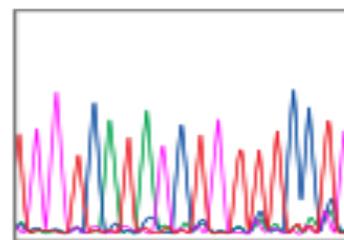
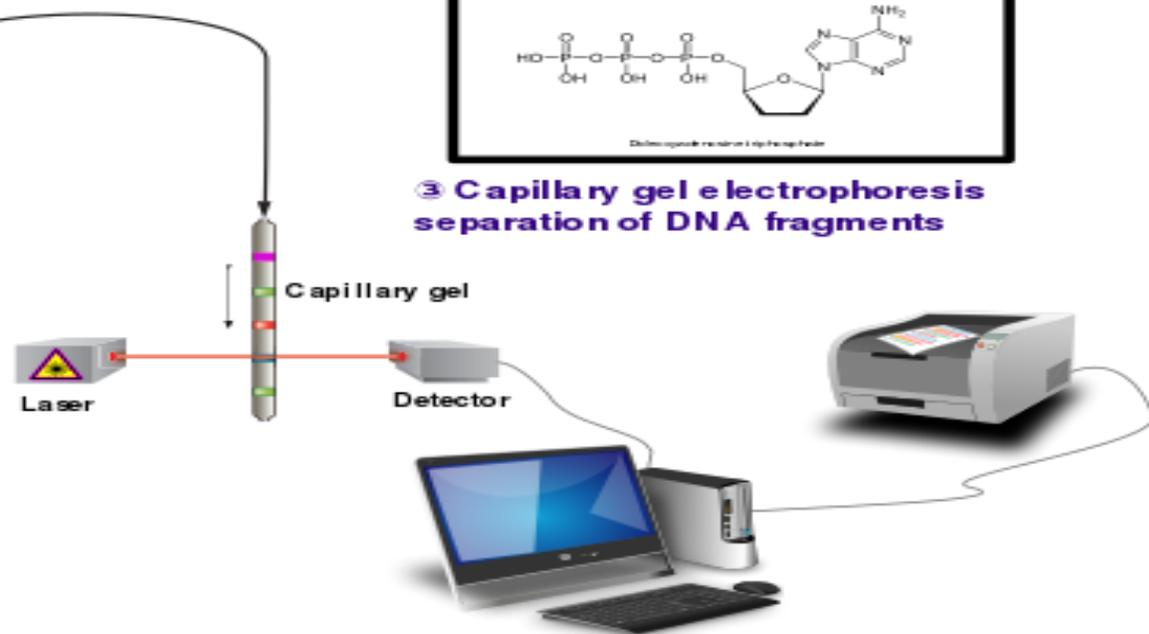
测序分析仪工作原理

① Reaction mixture

- ▶ Primer and DNA template
- ▶ DNA polymerase
- ▶ ddNTPs with flouochromes ▶ dNTPs (dATP, dCTP, dGTP, and dTTP)

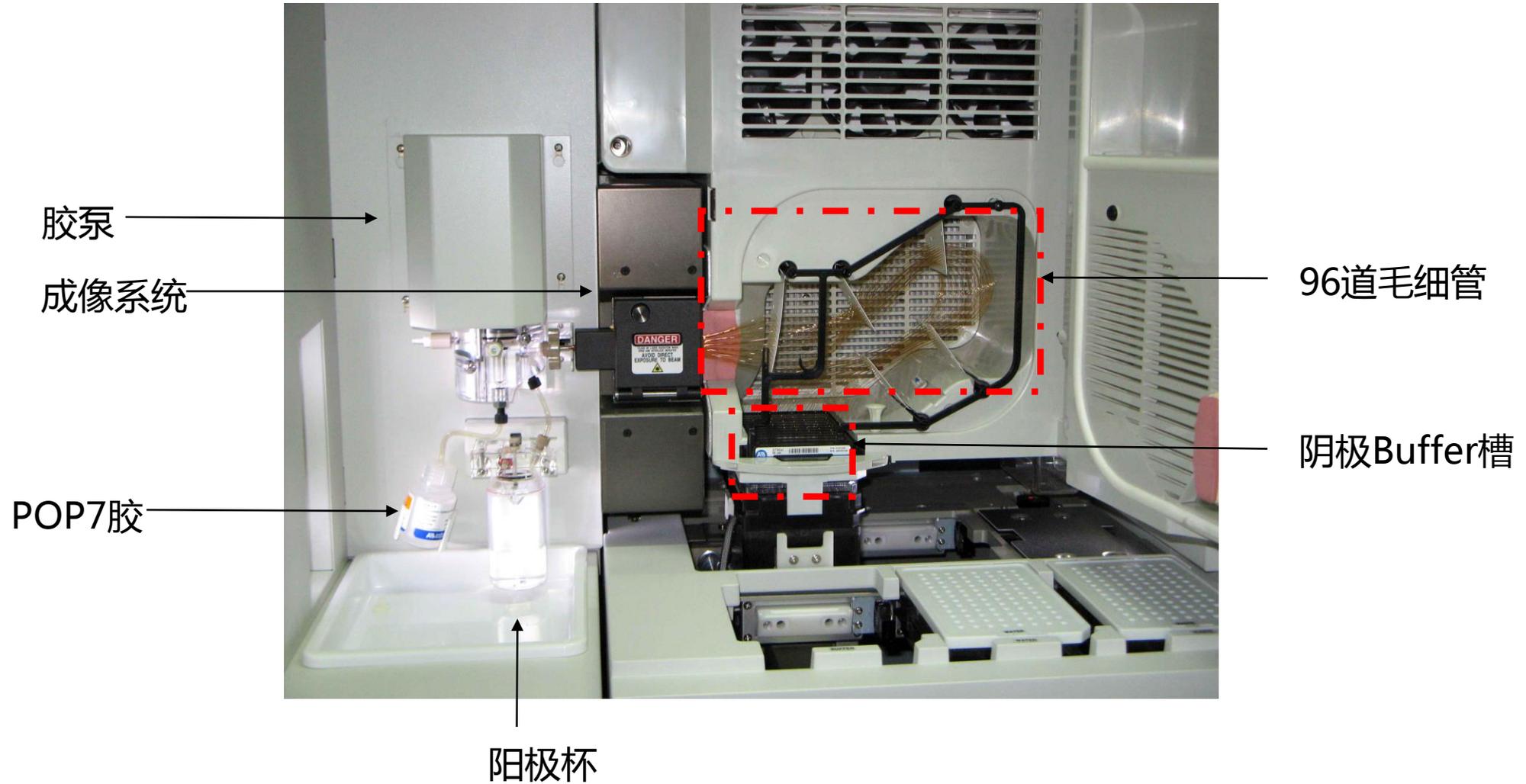


③ Capillary gel electrophoresis separation of DNA fragments

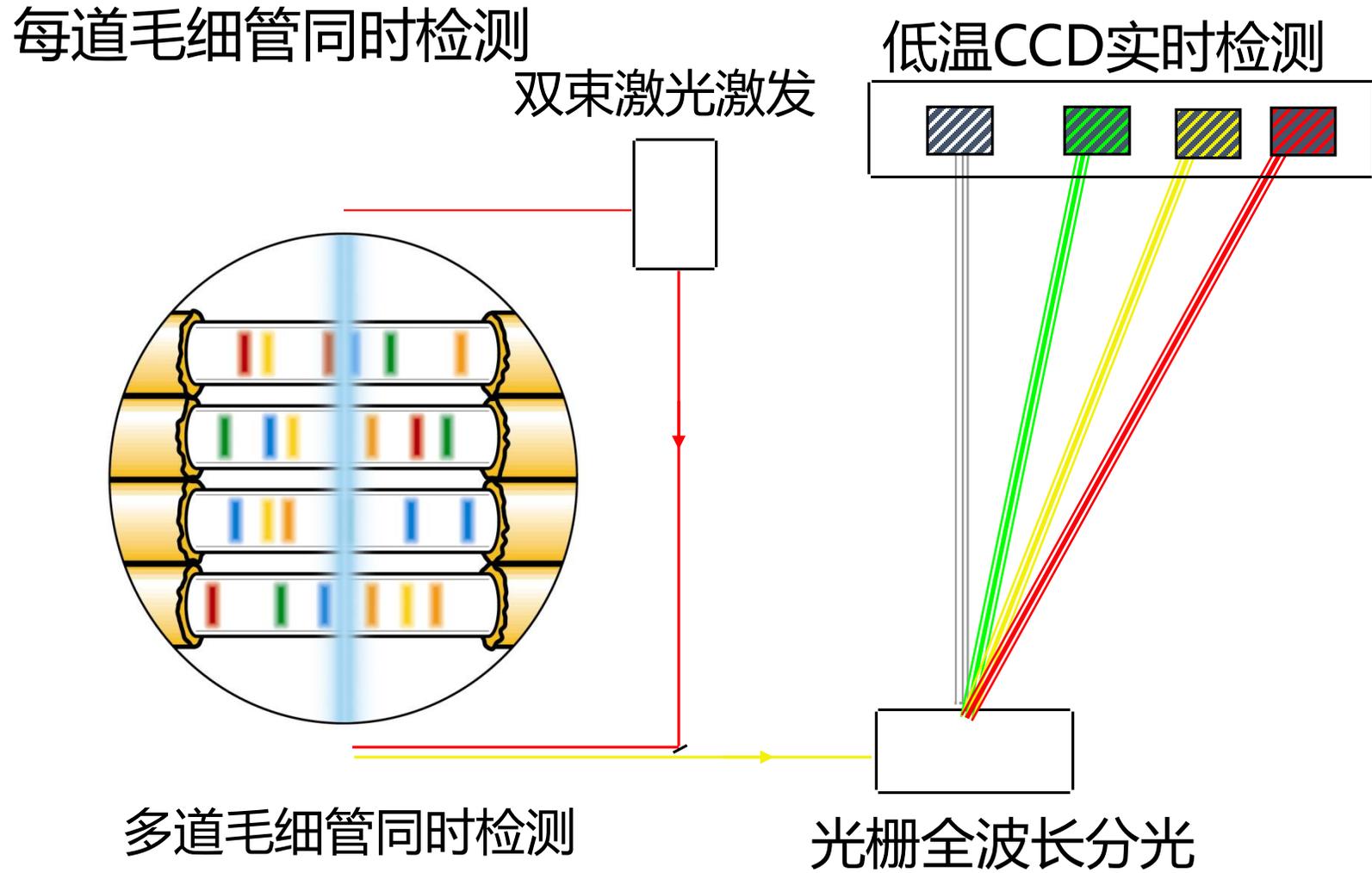


④ Laser detection of flouochromes and computational sequence analysis

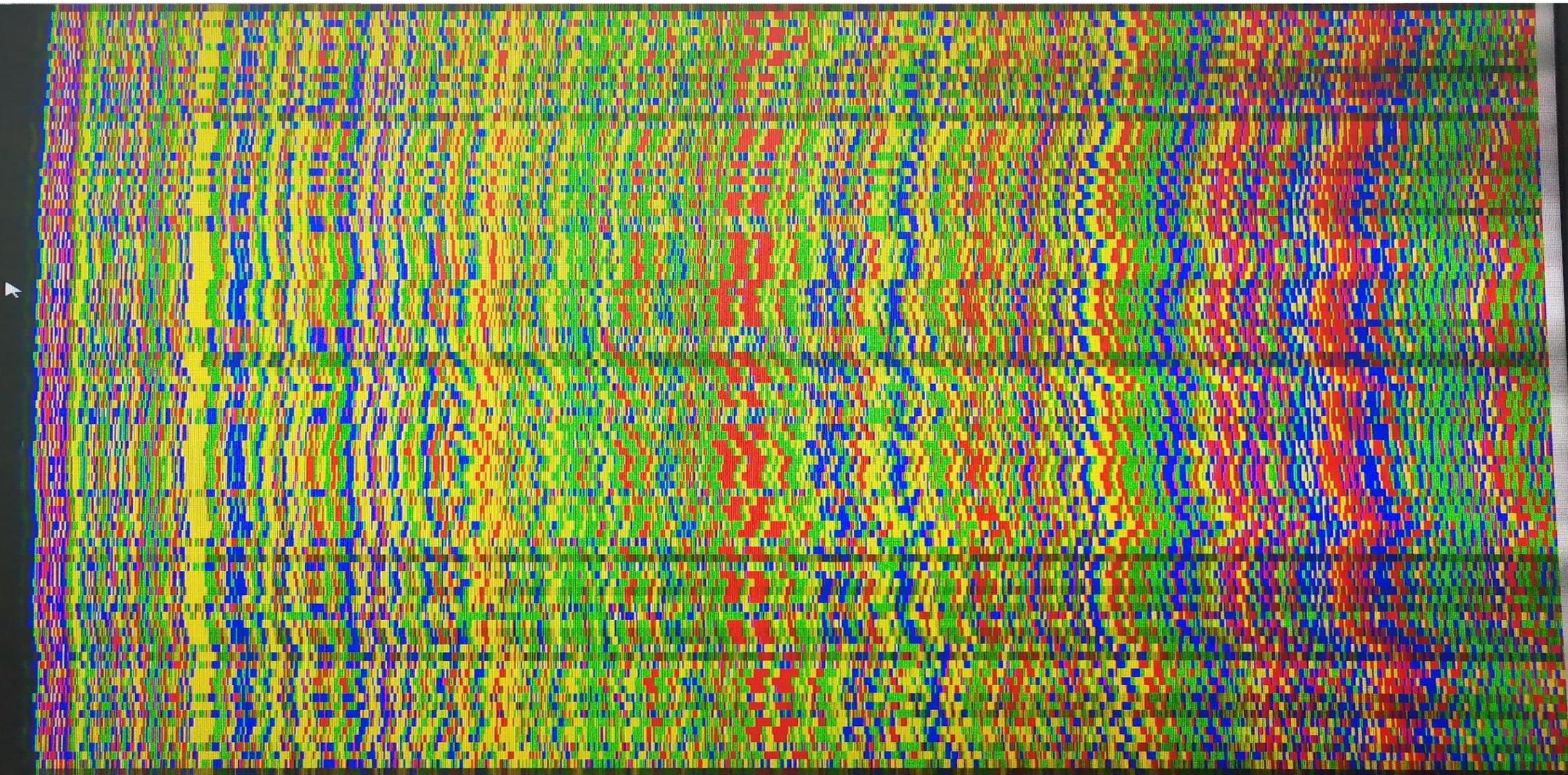
测序分析仪硬件结构



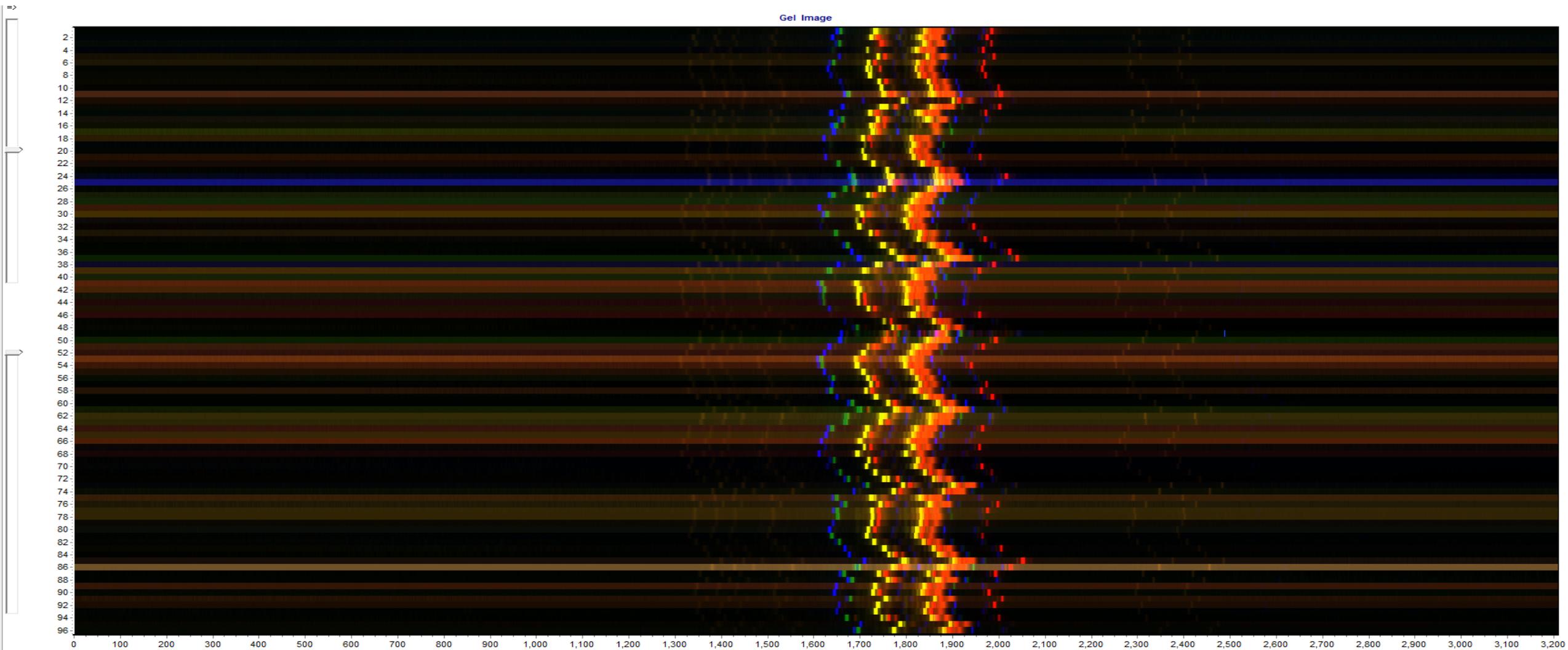
专利的荧光检测技术



Sanger测序毛细管电泳图



DNA片段分析毛细管电泳图





中国科学院水生生物研究所

INSTITUTE OF HYDROBIOLOGY, CHINESE ACADEMY OF SCIENCES

分析测试中心

一代测序技术应用范围

面向研究和应用市场的多种CE和NGS应用方案

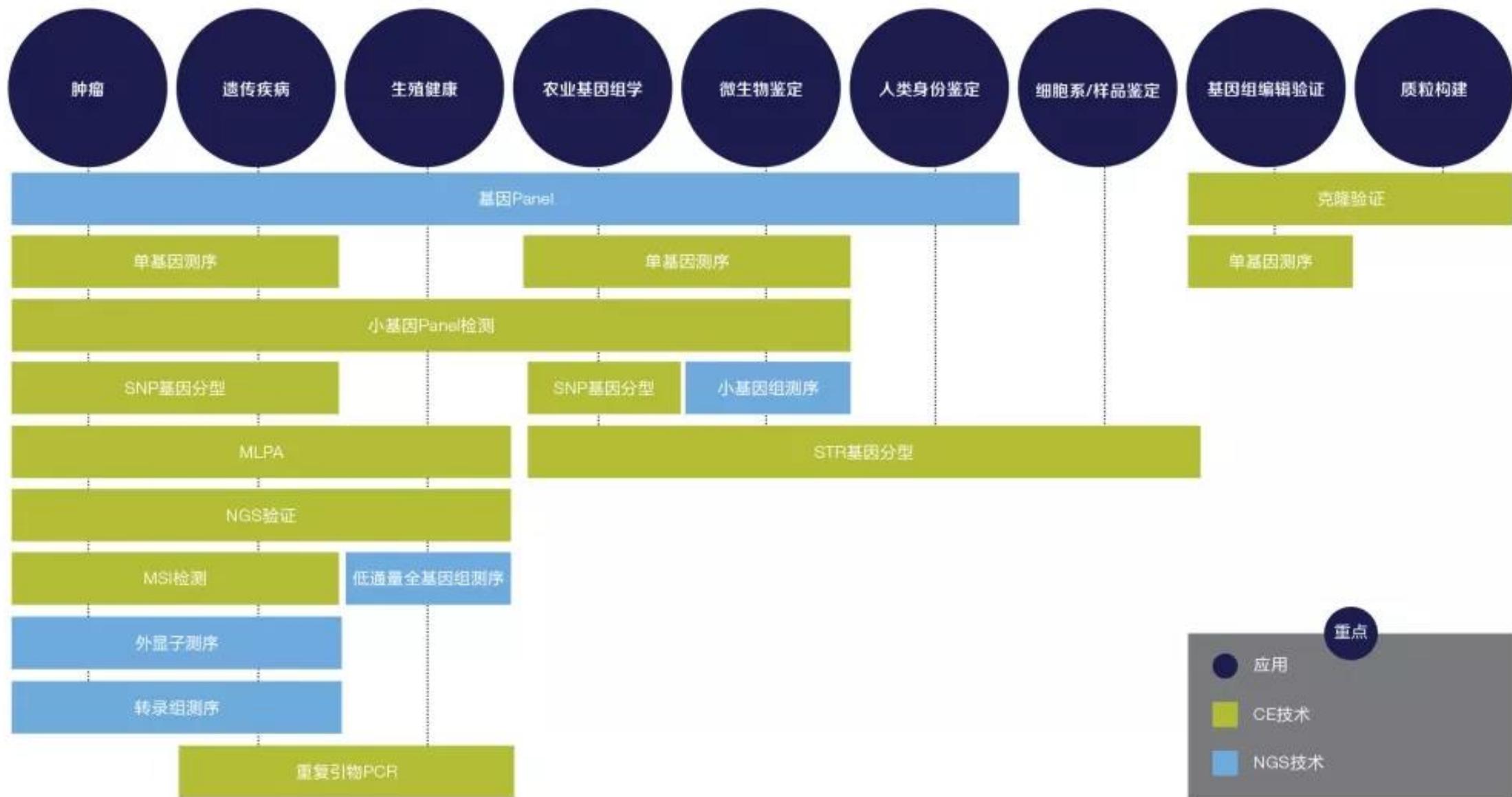
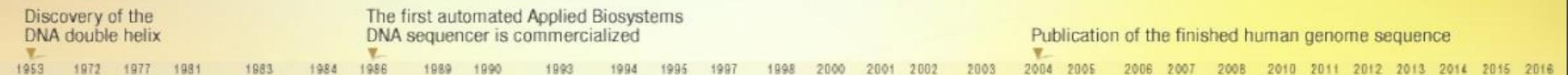
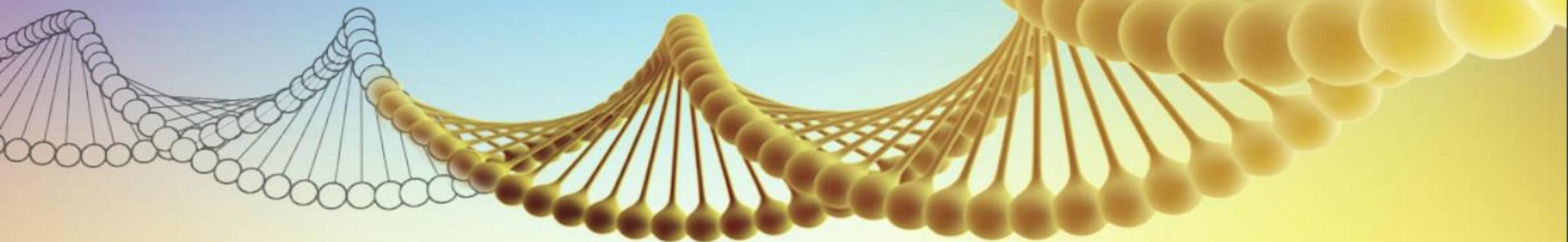


图1. 毛细管电泳(CE)和NGS技术(根据研究领域和应用市场划分)。

为何在二代三代测序技术早已出现的今天

Sanger测序还有其独特的地位

并且是行业内的**金标准**



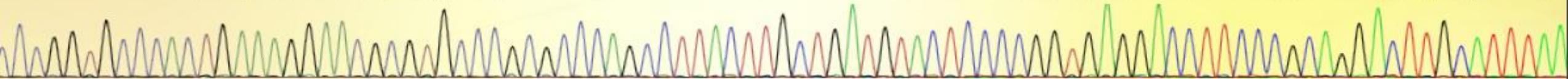
Discovery of the DNA double helix

The first automated Applied Biosystems DNA sequencer is commercialized

Frederick Sanger develops Sanger sequencing

Publication of the finished human genome sequence

Minor Variant Finder enables 5% somatic variant detection using Sanger sequencing



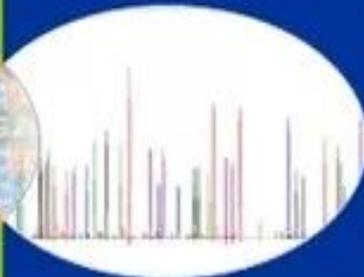
测序金标准

“**金标准**”是什么意思？就是公认最可靠的办法，新的基因测序准确率多高，还需要Sanger测序来验证。

Phred-scale 质量分数起源于人类基因组计划，定义为 $Q = -10 \log_{10} P$ ，其中Q是质量得分，P是概率误差。例如，Phred质量得分为10，误差为10%，准确度为90%。Sanger测序的总体准确度约为99.999%，或总误差为0.001%。由于读取长度在1000bp，Sanger测序的这两个方面(准确性和读取长度)使其成为测序的金标准。

	第一代	第二代	第三代
	Sanger-毛细管电泳测序法	可逆链终止物和合成测序法	实时单分子DNA测序
误差	十万分之一	千分之一	百分之十五
准确率	99.999%	99.9%	PacBio准确率仅为85%
QV值	50	30	8
读长(碱基数)	1000	2x150 2x300	~1000

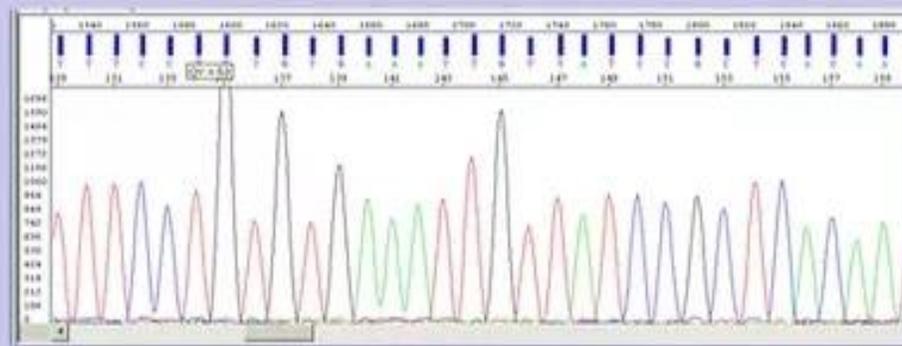
Sanger测序



多重荧光片段分析

- 靶基因测序
- 二代测序验证
- HLA 分型
- 微生物测序
- 线粒体DNA测序

- 微卫星不稳定性 (MSI) 检测
- 短片段重复序列 (STR) 检测
- 多重SNP 分析 (SNapShot技术)
- 多重相对荧光定量分析 (QF- PCR技术)
- 多重连接探针扩增法 (MLPA技术)
- 扩增片段长度多态性 (AFLP)



MSI



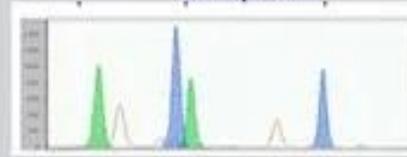
HLPA



MLPA



SNapShot





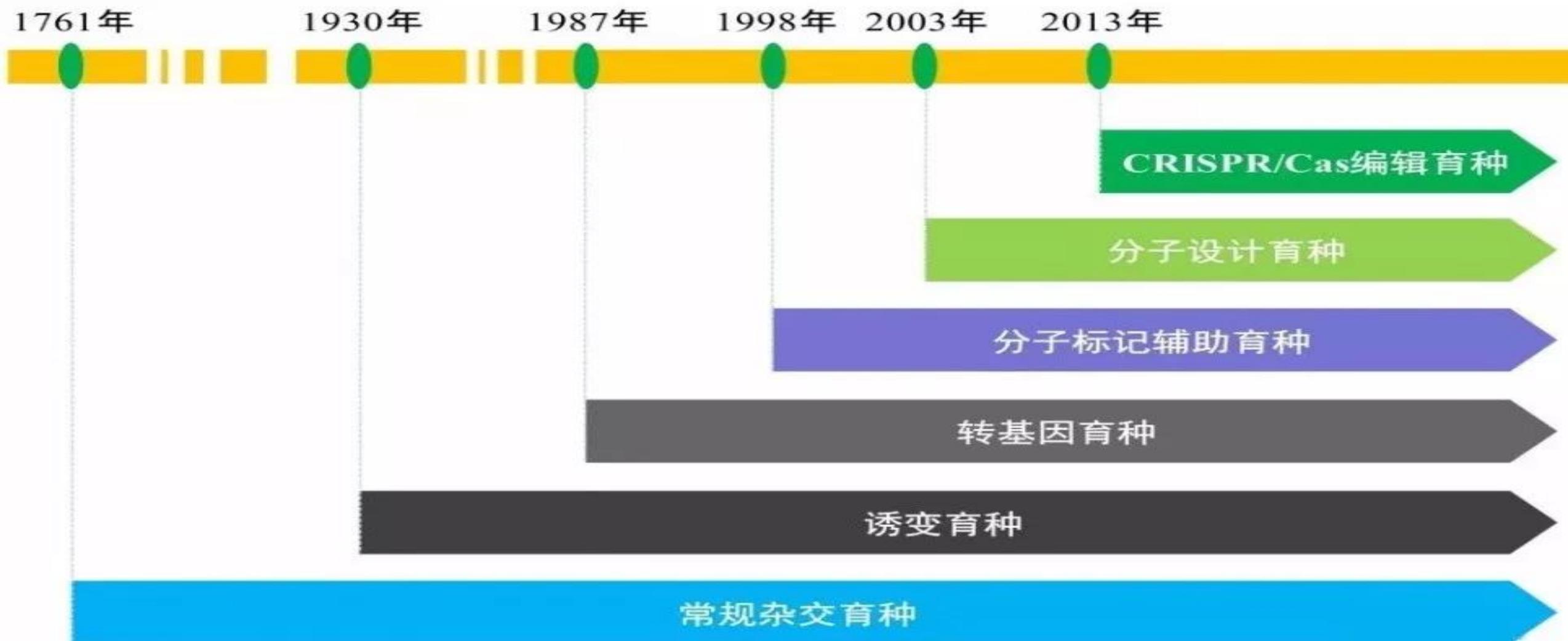
中国科学院水生生物研究所

INSTITUTE OF HYDROBIOLOGY, CHINESE ACADEMY OF SCIENCES

分析测试中心

一代测序在水生生物研究中的应用

基因是性状的底层密码



异育银鲫家族



(异育银鲫家族“1号”：异育银鲫)



(异育银鲫家族“2号”：高体型异育银鲫)



(异育银鲫家族“3号”：异育银鲫“中科3号”)



(异育银鲫家族“4号”：长丰鲫)



(异育银鲫家族“5号”：异育银鲫“中科5号”)

鱼类基因功能的研究



纯合型(-/-)
Homozygous



杂合型(+/-)
Heterozygous



野生型(+/+)
Wildtype



×



基因分离



25%

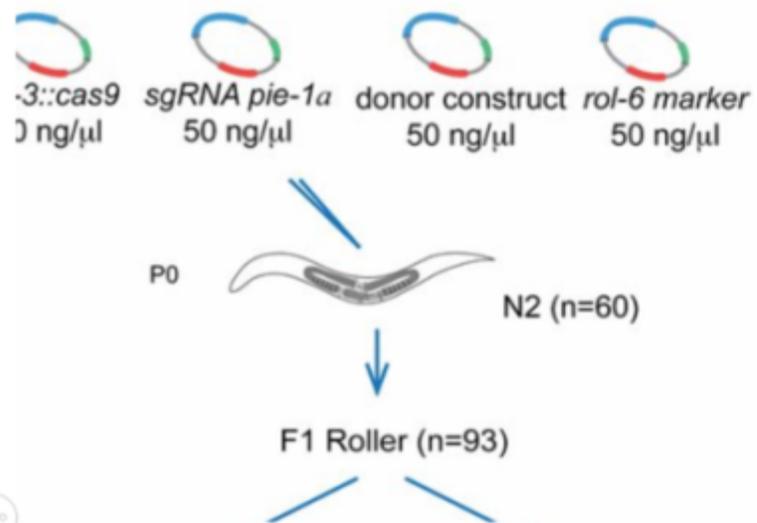
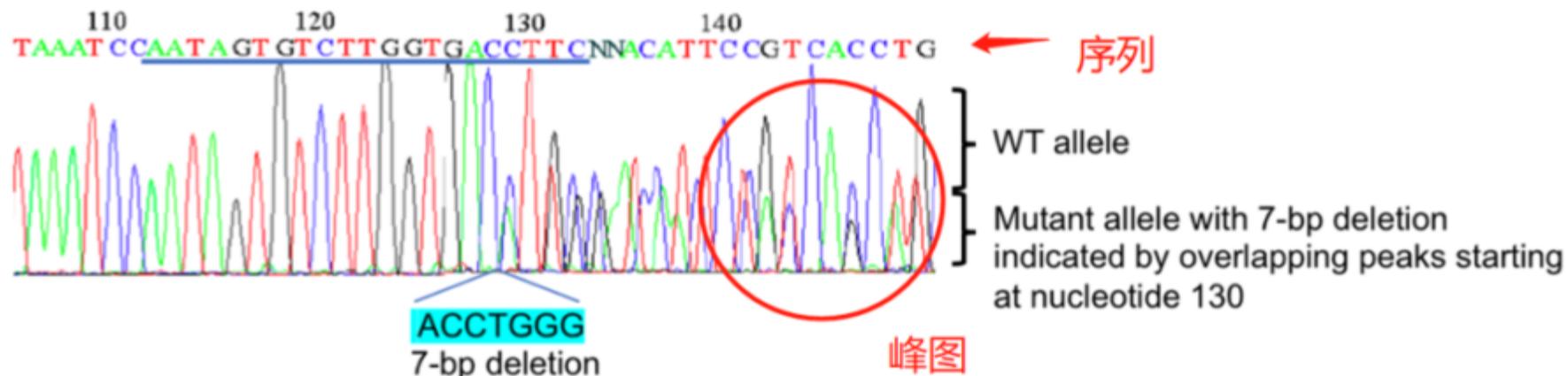


50%



25%

鱼类基因功能的研究



B P L A Q I D E A P A T K R

CCATTGGCTCAGATTGACGAGGCGCCGGCAACTAAAAGA WT

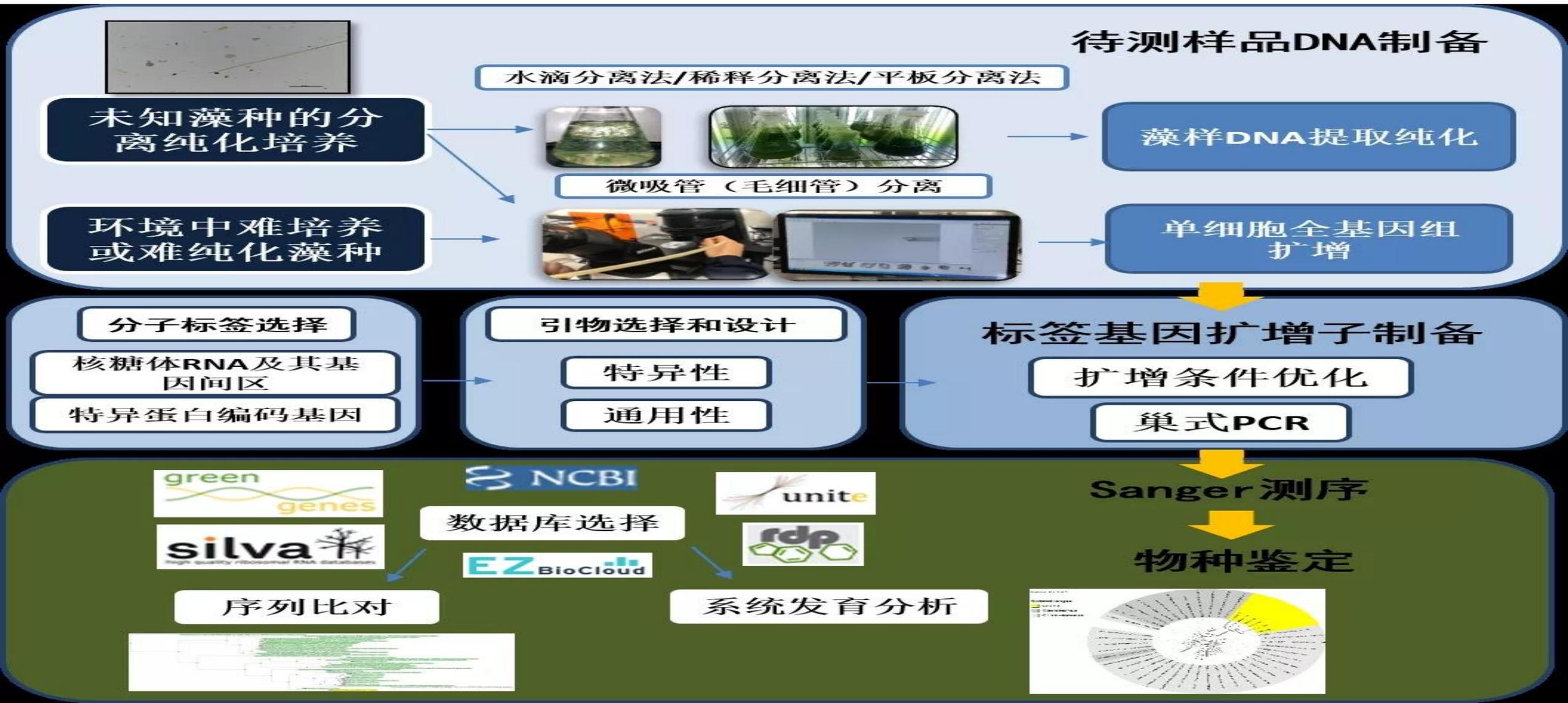
(i) F1 Homozygous indel (3/93)

CCATTGGCTCAGATTGACG-----CGGCAACTAAAAGA -5
 CCATTGGCTCAGATTGACGAG-----CAACTAAAAGA -7
 CCATTGGCTCAGA-----AACTAAAAGA -17

(ii) F1 indel / indel (2/93)

CCATTGGCTCAGATTGACGAG-----AACTAAAAGA -7
 CCATTGGCTCAGATTGACGAG-----CAACTAAAAGA -7
 CCATTGGCTCAGATTGACGAG-----AACTAAAAGA -7
 CCATTGGCTCAGAggg-----GCCGGCAACTAAAAGA -7(-10,+3)

藻种分子鉴定



分子标签的选择

分类

蓝藻

绿藻

硅藻

甲藻

隐藻

黄藻

裸藻

金藻

核糖体RNA&ITS

16S rRNA/ITS

18S rRNA/ITS

18S rRNA/28S
rRNA/ITS

18S rRNA/28S rRNA/
ITS/23S rRNA

18S rRNA/23S rRNA

18S rRNA

18S rRNA/28S rRNA

18S rRNA/ITS

管家基因

*cpcBA-IGS / rpoC1/ nifH/ cpcB/
rbcL/hetR/psbA/rpoD1/rbcS*

rbcL/tufA/psbA

rbcL/COI

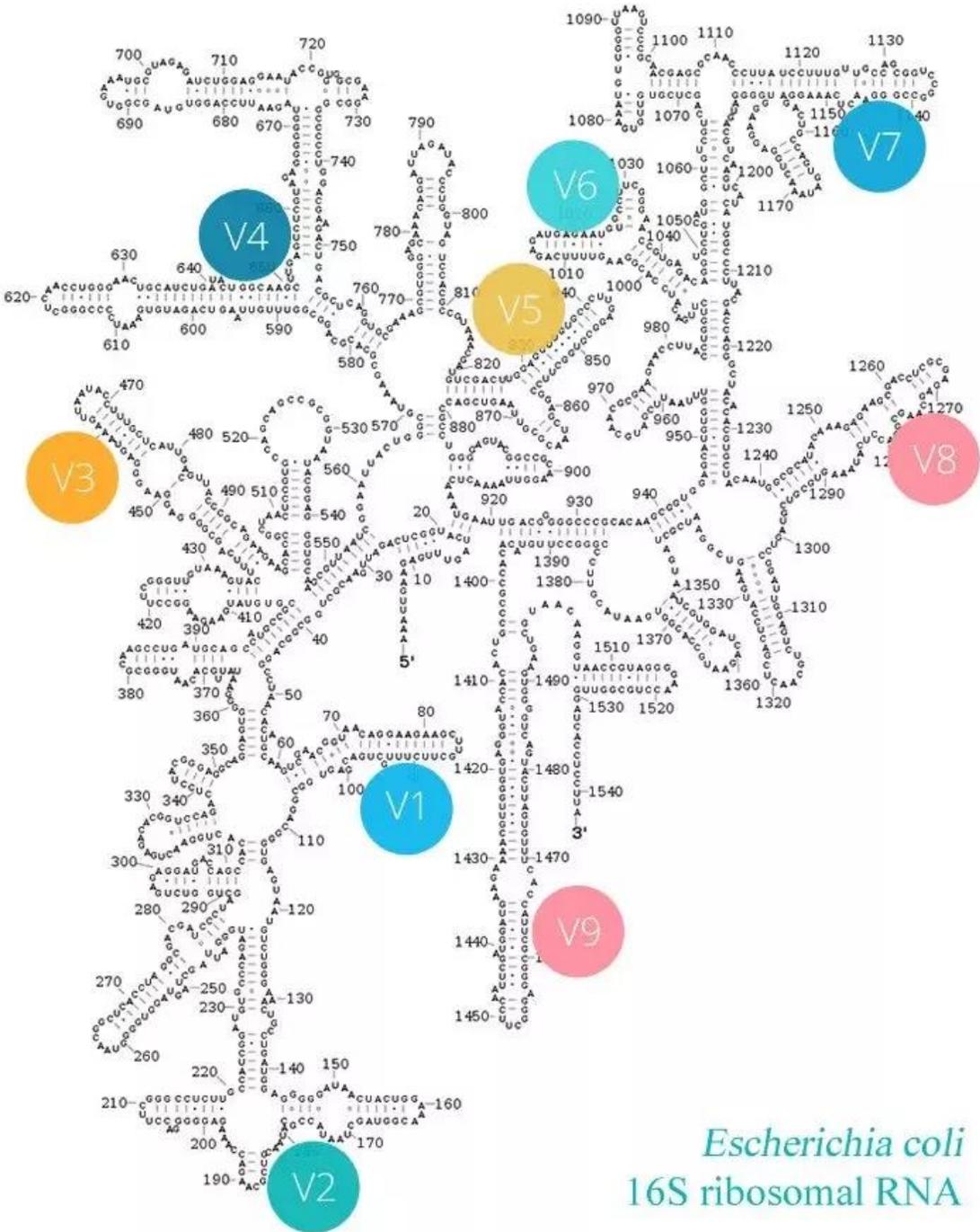
cob

/

rbcL/COI

/

/



16S rRNA 基因是编码原核生物核糖体小亚基的基因，长度约为1542bp，其分子大小适中，突变率小，是细菌系统分类学研究中最常用和最有用的标志。

16S rRNA基因序列包括9个可变区和10个保守区，保守区序列反映了物种间的亲缘关系，而可变区序列则能体现物种间的差异。

16S rRNA基因测序以细菌16S rRNA基因测序为主,核心是研究样品中的物种分类、物种丰度以及系统进化。

我们知道，16S rRNA基因普遍存在于细菌细胞，在细菌基因组中位于核糖体小亚基（约1540 bp），该区域兼顾保守性和高变性，含有10个保守区域(Conserved Regions)和9个高变区域(variable Regions)，保守区可用于设计引物进行目的片段的扩增，而通过对高变区的分析可以辨别细菌种类。

因此，16S rRNA基因被认为是最适于细菌系统发育学研究和物种分类鉴定。



中国科学院水生生物研究所

INSTITUTE OF HYDROBIOLOGY, CHINESE ACADEMY OF SCIENCES

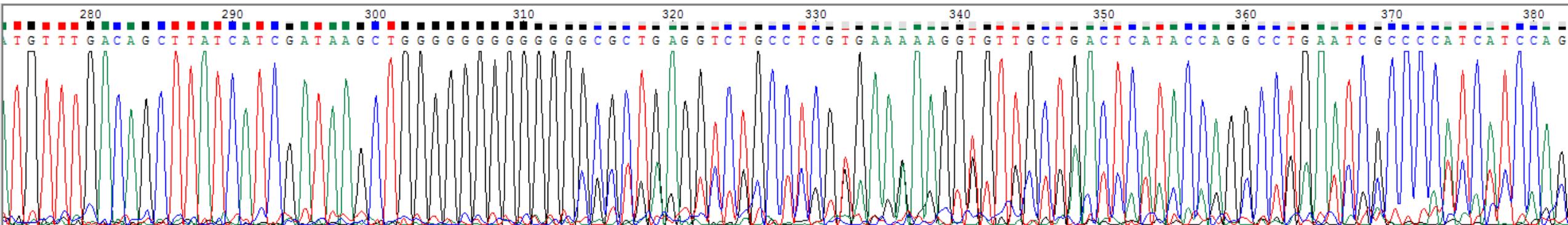
分析测试中心

常见问题案例解析

**一切以峰图为准！
一切以峰图为准！
一切以峰图为准！**

在测序结果中，峰图文件 (.ab1) 最能真实地反映样本情况，序列文件是分析软件根据峰图自动导出的，若出现峰图与序列不一致的情况，请以峰图为准。

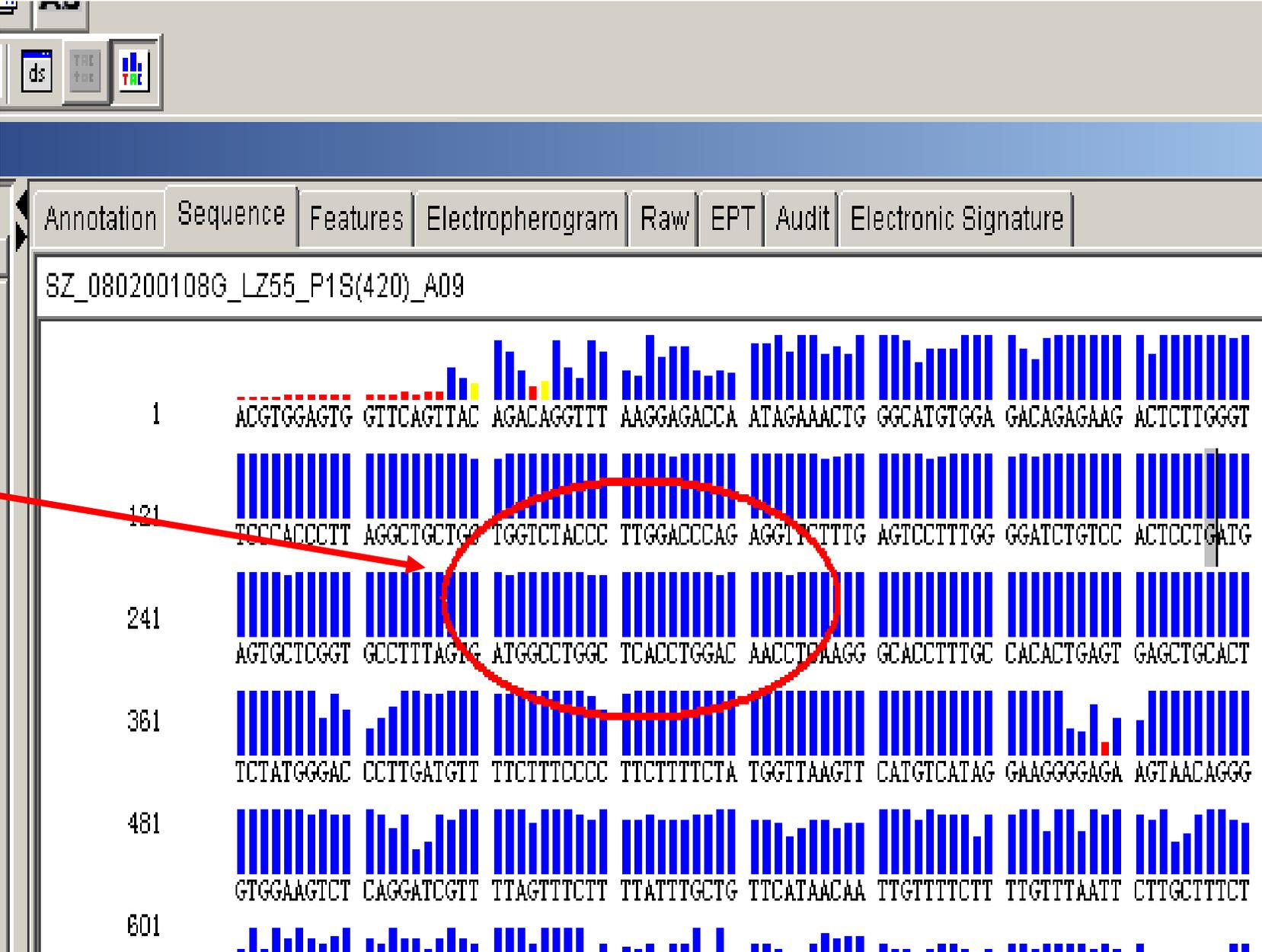
这样的峰图导出的序列可信么？



CATGTTT GACAGCTTATCATCGATAAGCTGGGGGGGGGGGGGGGGCGCTGAGGTCTGCCTC
GTGAAAAGGTGTTGCTGACTCATAACCAGGCCTGAATCGCCCCATCATCCAGCCAGAAA

怎样判别测序数据的质量

Base QV
for each base in
each sample file

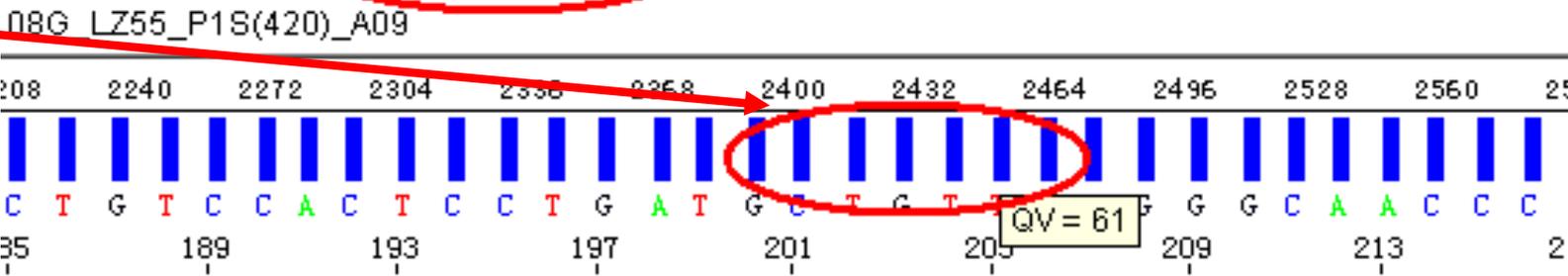


怎样判别测序数据的质量

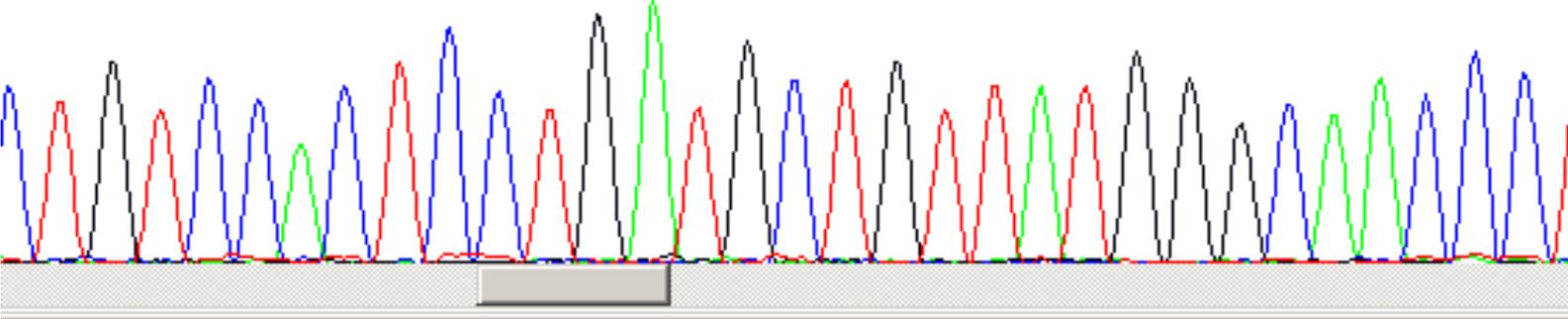
- **Base QV for each base in each sample file**

SZ_080200108G_LZ55_...	1	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	KB.bcp
SZ_080200108G_LZ55_...	1	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	KB.bcp
SZ_080200109G_LZ65_...	1	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	KB.bcp
SZ_080200109G_LZ65_...	1	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	KB.bcp

Sequence | Features | **Electropherogram** | Raw | EPT | Audit | Electronic Signature

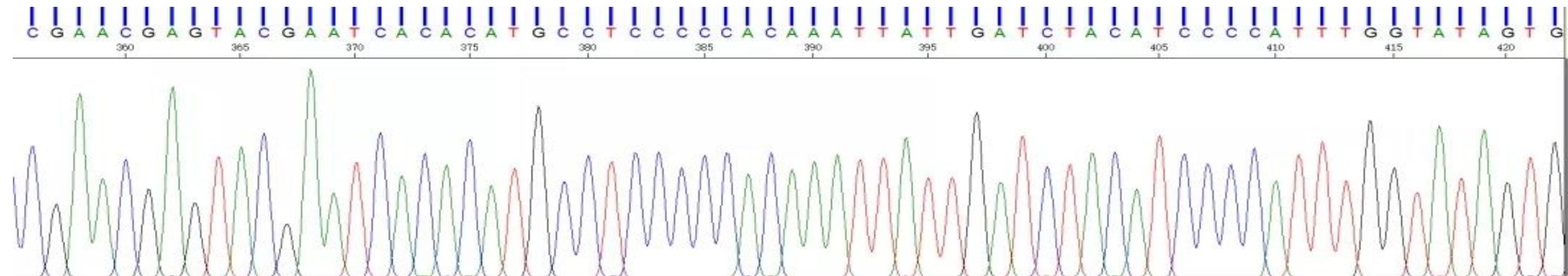


$QV = -10 \times \lg Pe$
 上面公式中，
 Pe代表碱基读取的错误概率
 QV10 代表Pe=10%；
 QV20 : Pe=1%；
 QV30 : Pe=0.1%
 QV40 : Pe=0.01%

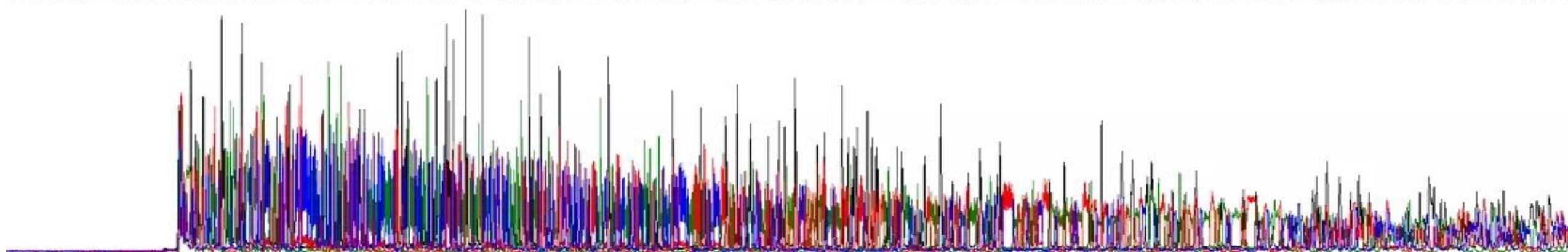


正常峰图特征

峰形比较完整，基本为单一峰形

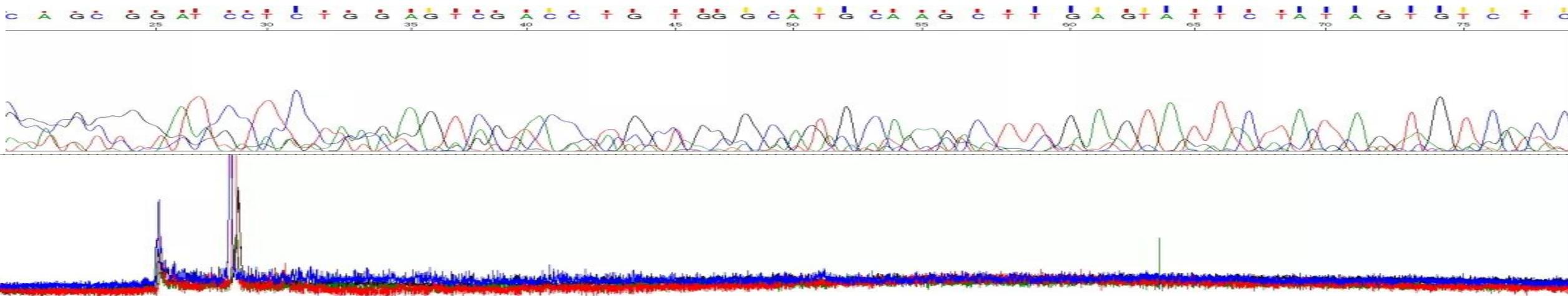


正常峰图示例



Sequence Scanner软件中Raw图实例

无信号



峰图特征

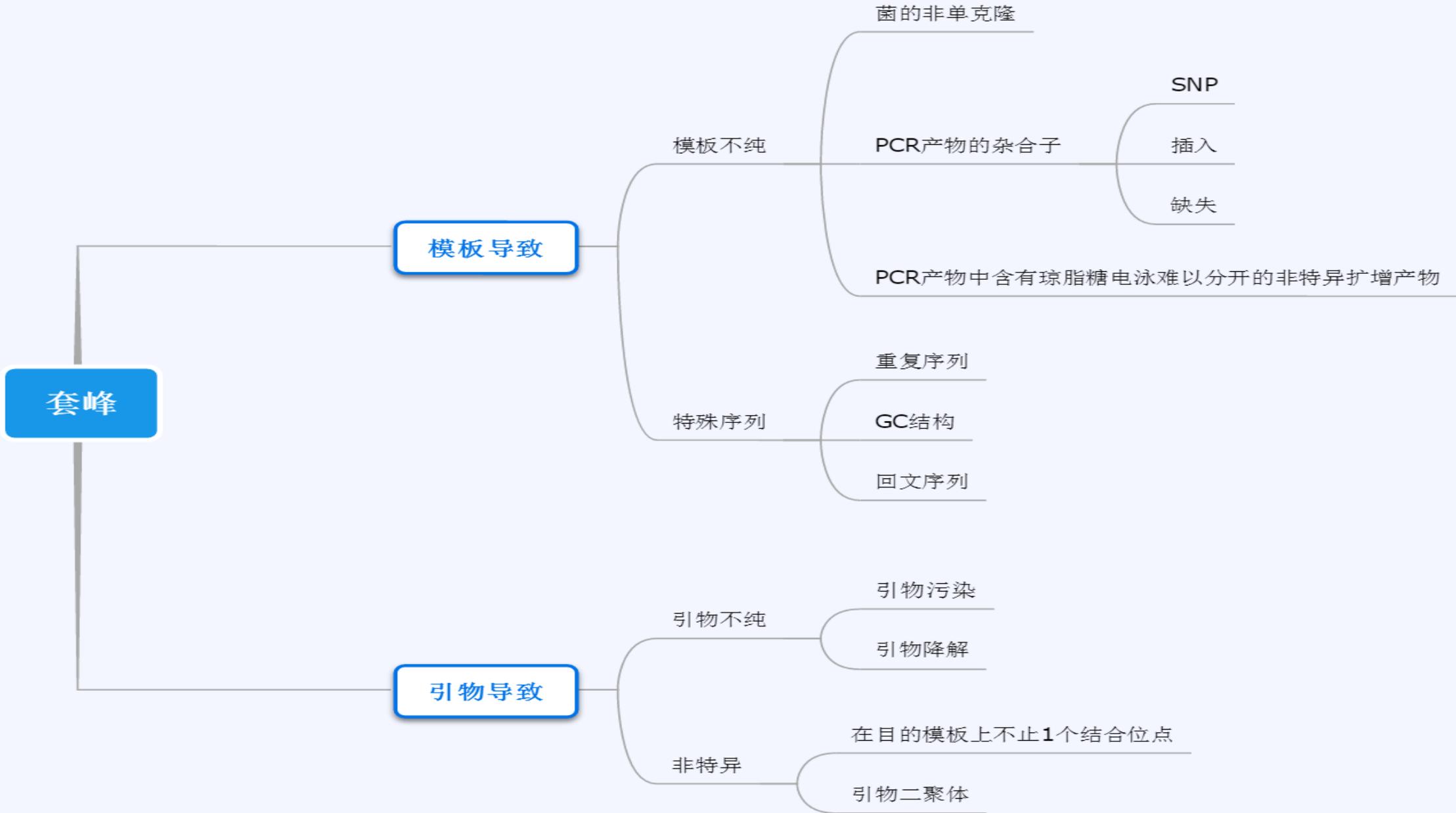
序列峰图杂乱无章，测序干扰大，没有明显主峰

成因分析

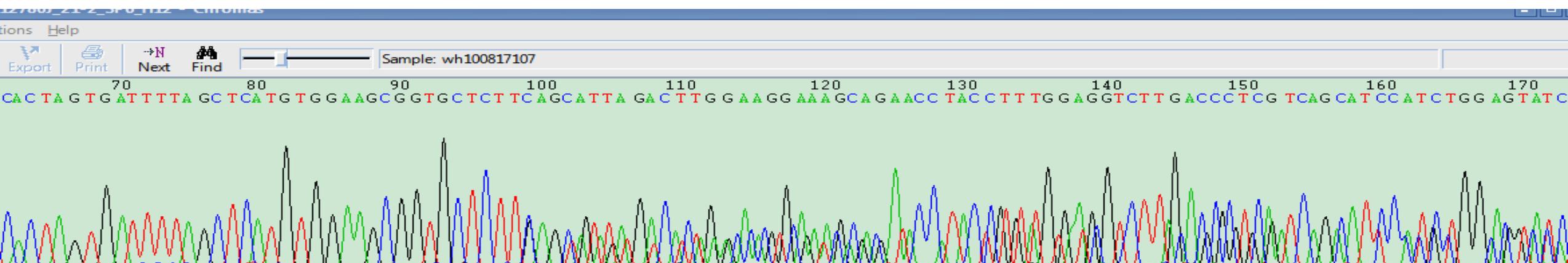
- 模板浓度太低
- 可能在引物与模板结合处存在突变导致不结合
- 空载，无插入片段上引物结合位点

解决方案

- 重新制备模板
- 换其他引物测序
- 用载体上通用引物测序



非单克隆



峰图特征

一般在峰图开始70-100bp位置会出现套峰，也有在峰图中部或后面才出现套峰的

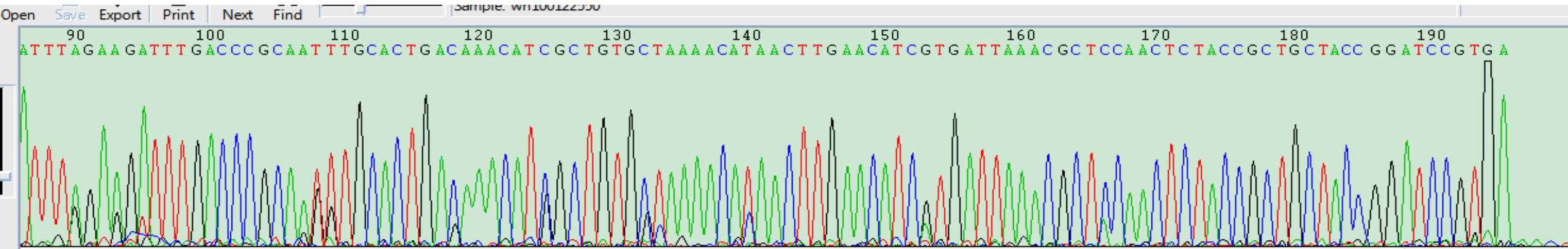
成因分析

挑到了2个不同的克隆子，2个克隆子的载体序列是相同的，所以前面一般会有70-80bp载体序列相同的单峰。2个不同的克隆子之间的插入片段的差异就体现在套峰出现的位置

解决方案

重新挑取单克隆

杂合-SNP



峰图特征

个别碱基位点出现套峰

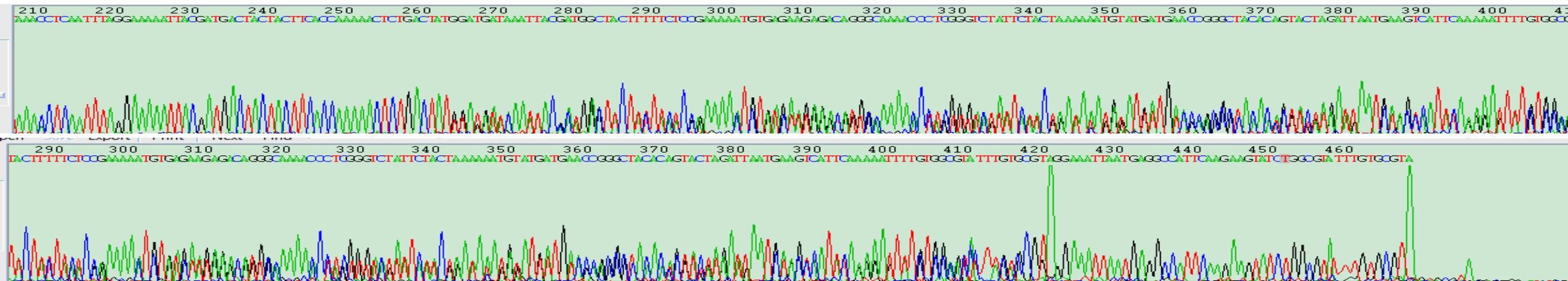
成因分析

样品模板本身就是一个处于杂合状态的基因，出现套峰的位置可能就是基因的天然突变后形成的突变位点

解决方案

如需获得单一产物
连接载体克隆分离

杂合—InDel



峰图特征

峰图前半部分正常，
后面某个位置开始
出现套峰

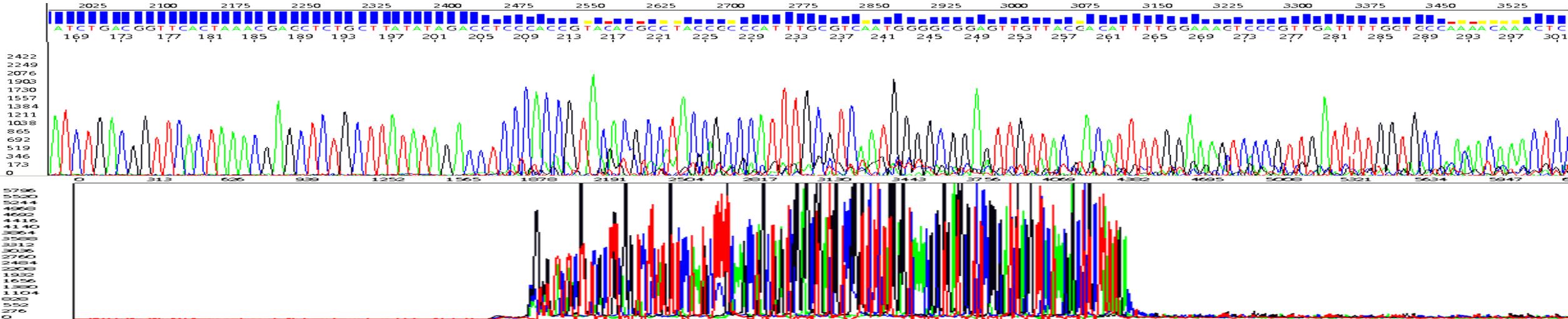
成因分析

在265bp的位置有部分模板由于碱基缺失（或插入），导致在这个位置开始套峰。因为由于碱基缺失（或插入），变成了长短有差异的2条链。在峰图结束的地方，有45bp的单峰，长链比短链长45个碱基

解决方案

如需获得单一产物
连接载体克隆分离

高级结构



峰图特征

原始测序信号开始正常，到某个地方突然中断或者减到一个比较低的水平

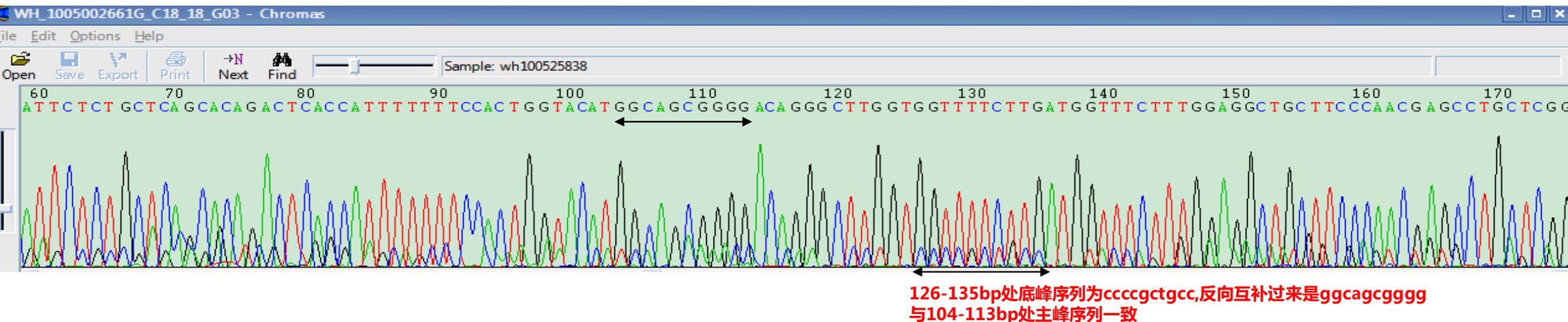
成因分析

局部GC含量高，回文序列等区域，在普通的退火温度条件下双链不易打开或者打开不充分，导致扩增的效率突然降低，得到的产物片段少，信号强度突然变低。测序仪在读取信号时候一般有色比，高级结构后面信号太低，跟背景信号混在一起形成套峰

解决方案

反向测序拼接

引物污染



峰图特征

连续且与主峰等长的底峰，底峰的反向互补序列与主峰一致

成因分析

正反向混合引物都能与模板结合扩增出相互反向互补的单链片段测序产物

解决方案

重新稀释得到单向纯引物

引物降解

图1

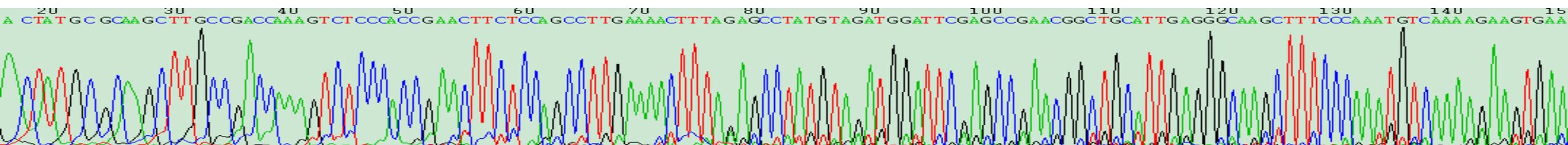
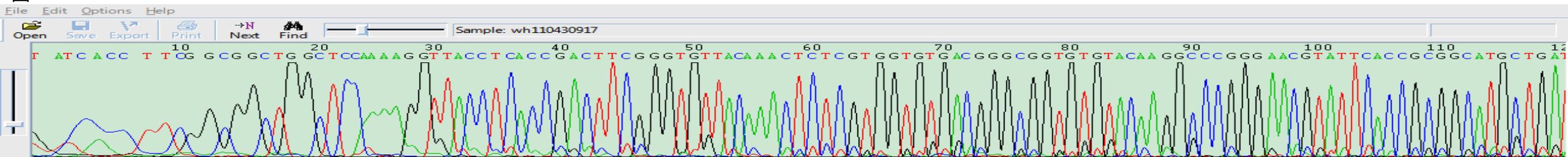


图2



峰图特征

从峰图开始就有轻微向前移码的底峰

成因分析

部分测序引物轻微降解，5' 端轻微降解的引物还是能够正常跟模板结合，由于5' 端短了n个碱基，此降解引物扩增得到的是比正常引物短n个碱基的测序片段产物

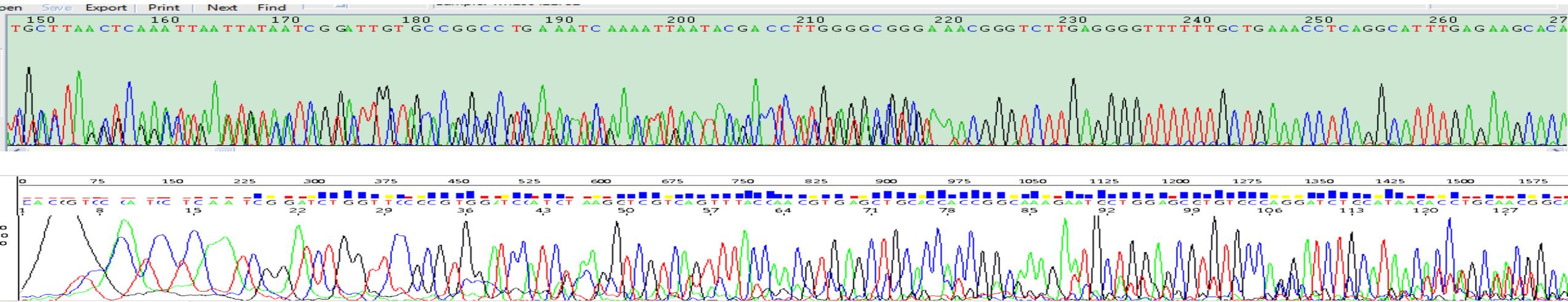
解决方案

重新稀释干粉引物
或重新合成纯引物

思考

峰图1和峰图2有本质上有区别吗？出现这样的峰图都是引物降解造成的吗？如果引物是3'端降解呢？

双引物结合



峰图特征

有2种情况。图1，从开始出现套峰，到中间某一个位置，套峰忽然消失。图2，从头到尾一直套峰，而且2套峰的高度差不多

成因分析

2种峰图成因本质是一样的，都是由于同一条引物在模板的不同位置结合

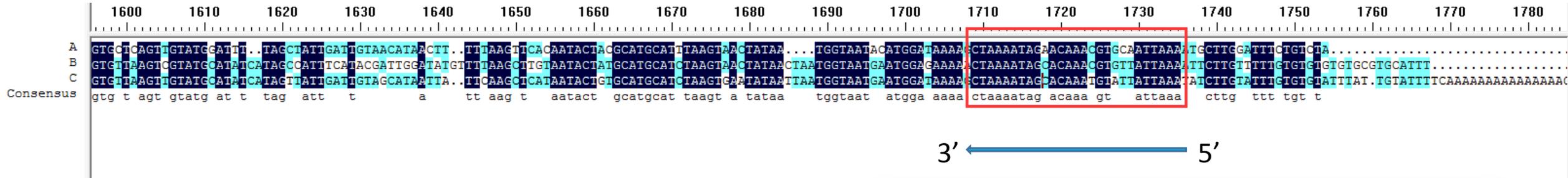
解决方案

更换反向测序引物或者重新设计特异性引物测序

思考

为什么图1套峰突然“消失”？图2的套峰却是从开始一直持续到结束？

案例分析：如何用近源物种转录组数据获取基因



问题

只有3个近源物种的转录组数据想要得到石蜡保存物种标本里基因信息，设计了很多引物都扩增不出来产物

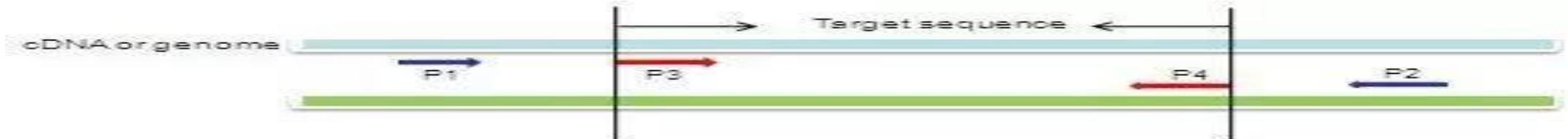
解决方案

- 相对保守区域设计引物
- 引物3' 端保守序列尽量长
- 使用兼并碱基
- 困难基因用巢式PCR增加特异性

巢式PCR (nested PCR, nPCR) 是一种变异的聚合酶链式反应(PCR)，使用两对（而非一对）PCR引物扩增完整的片段。扩增示意图如上：首先用目的基因两侧的P1和P2引物进行第一轮扩增，产物为初级扩增产物，然后取0.5-1μl初级扩增产物作为模板，用目的基因上下游的P3和P4进行第二轮扩增，就可以获得目的基因。

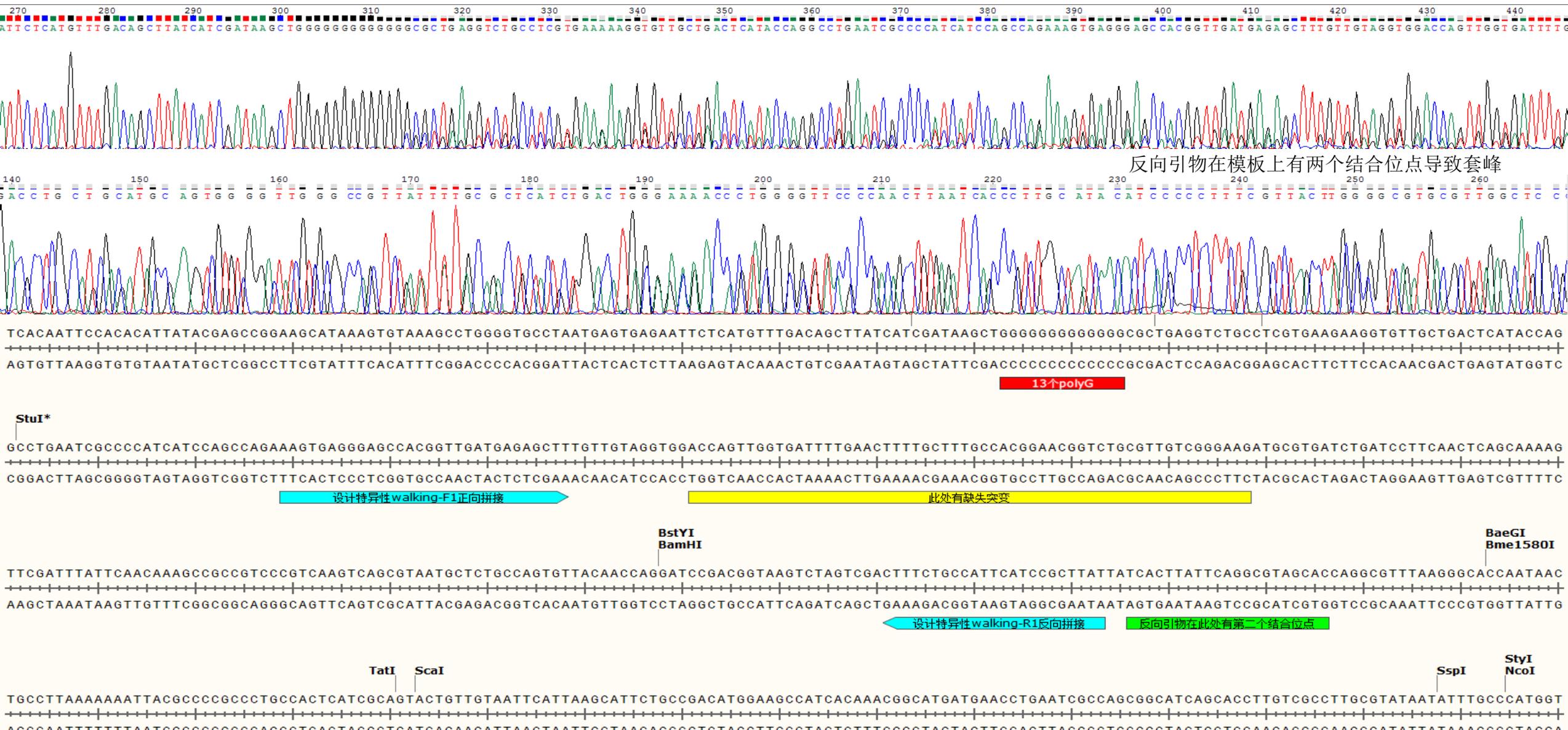
巢式PCR与普通PCR相比，有以下优势：

- 1、如果第一次扩增产生了错误片断，则第二次能在错误片段上进行引物配对并扩增的概率极低。因此，巢式PCR的扩增特异性非常高。
- 2、在巢式过程中往往第一轮PCR产物跑电泳检测看不到目的条带或者出现很多非特异性扩增条带，但是经过第二轮的PCR，目的条带就很单一、很亮了，这是因为巢式PCR很灵敏，尤其是样品中目的基因的表达丰度较低的情况下，巢式PCR的优势就展现出来了。



案例分析：前后夹击的困难基因如何测通

正向测序在310bp处遇到polyG导致后面套峰



武汉艾康健生物科技有限公司

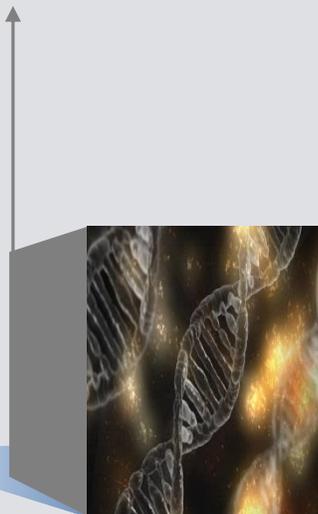
分析测试中心

一站式服务



Today Get

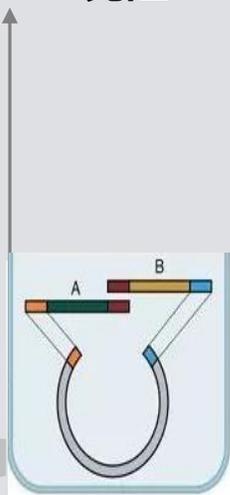
PCR重测序



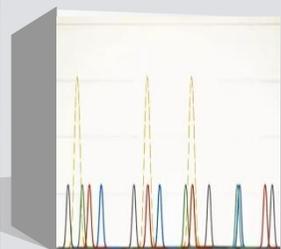
Sanger测序



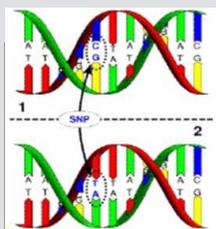
TA克隆



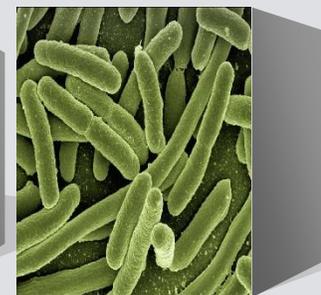
STR/SSR



SNP检测



菌种鉴定/Barcoding





分析测试中心419室



027-87530080