

# 轻松学用生物软件(2) 利用SeqMan进行序列拼接

主讲人: huaxing

2008年4月11日

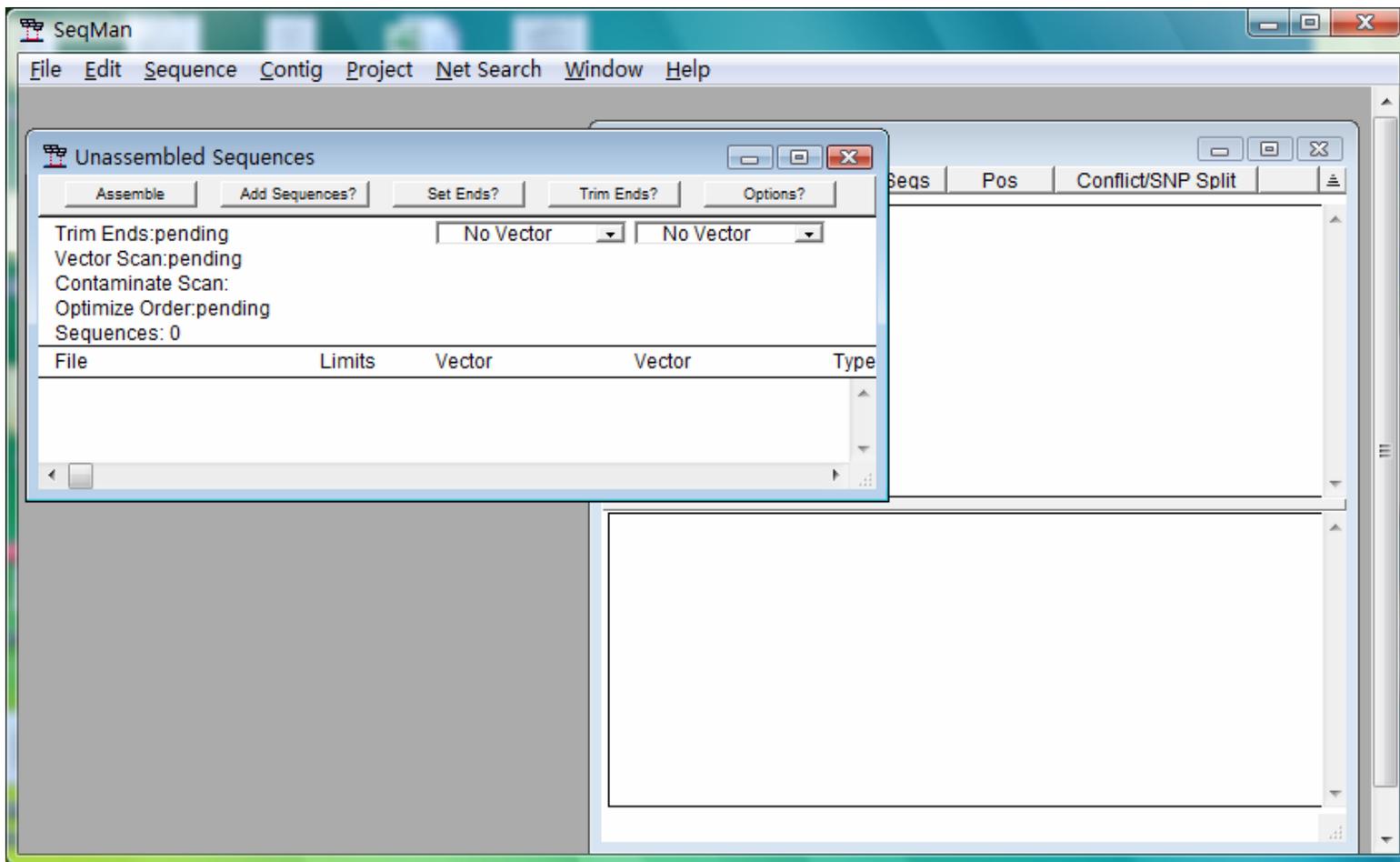


# 主要内容

- 大家好，首先感谢大家关注、支持螺旋网！今天我们一起一步一步的学习利用Dnastar 6.0中的子程序Seqman进行序列拼接，希望能对大家有所帮助。
- 注意：Dnastar 6.0的安装程序及破解补丁大家可以到螺旋网的“**生物信息学及生物软件交流**”版块免费下载！



# Step1:打开Seqman软件

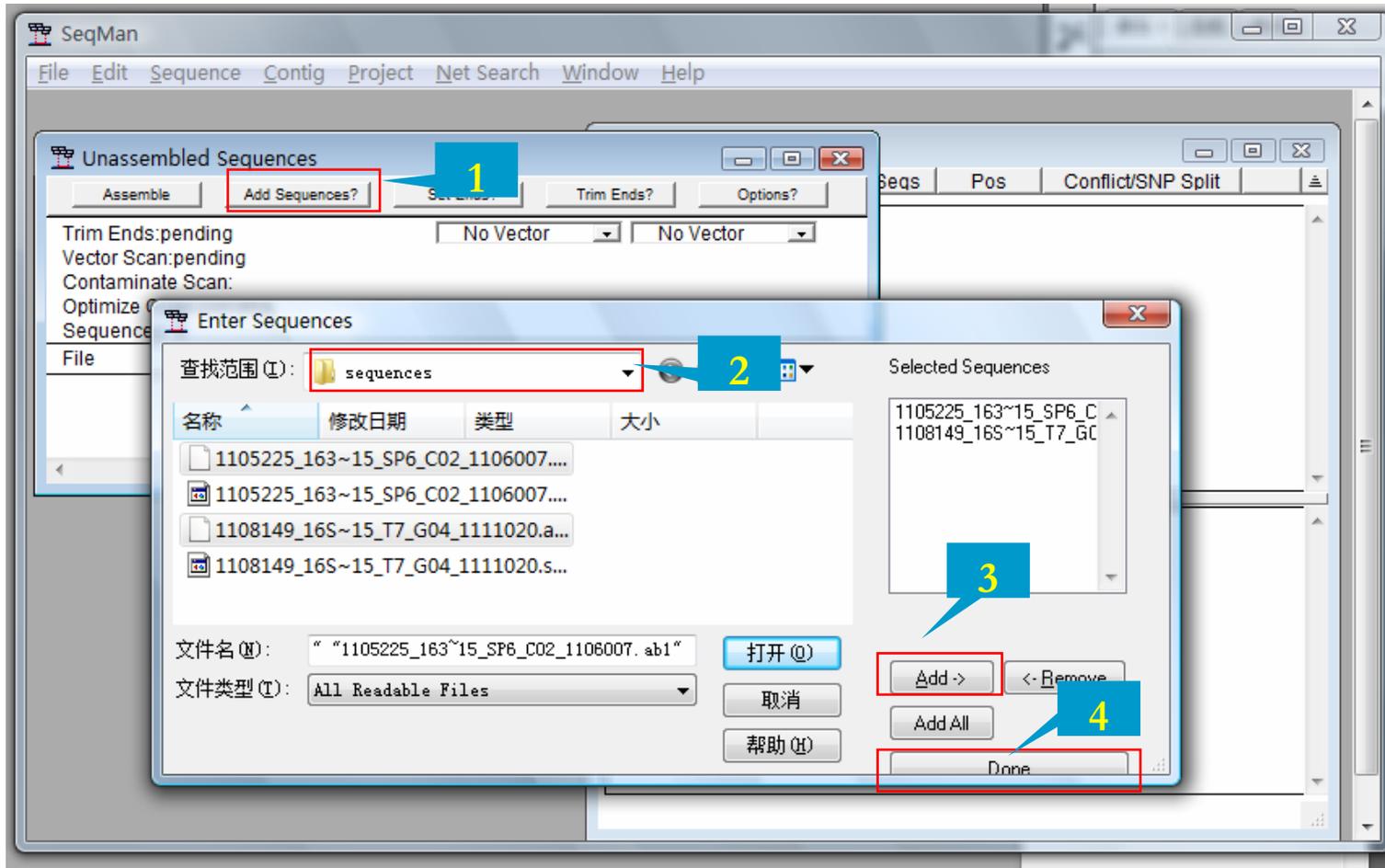




## Step2:加入你要拼接的序列

1. 点击Add sequences
  2. 查找并选中要拼接的序列（可按住control键进行多选）
  3. 点击Add按钮添加选择的序列
  4. 添加完后点击done
- 注：最好用测序的图谱尽量不要直接用测序得到的序列

# Step2:加入你要拼接的序列





## Step3:去除末端序列

- 主要是去除序列末端测序质量差或是载体序列
  - 有两种方法可以用来去除这类末端序列
    - 其一：利用Seqman自带的去除工具自动去除（利用Trim ends按钮进行）
    - 其二：手工去除
- 个人感觉手工去除方法最有效，因此下边我们以后手工去除为例进行演示

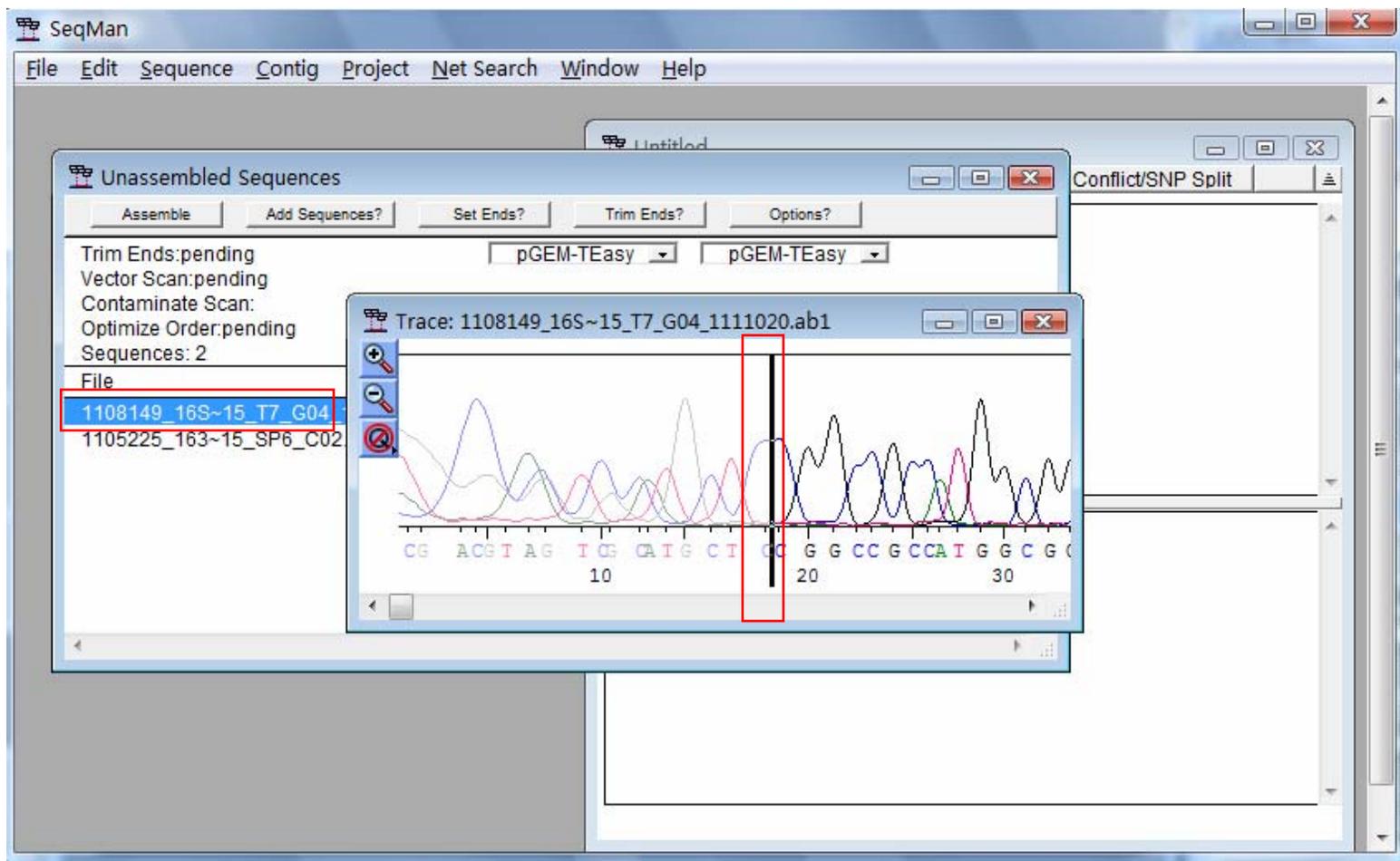


# 手工去除侧翼序列

1. 双击要去除侧翼序列的目标序列
2. 将鼠标放到测序图谱左边的一个黑色的竖线上，此时鼠标会变成一个有两个箭头的水平线
3. 按住左键拖动黑竖线，那么你就会发现侧翼序列的颜色变浅，这部分变浅的序列则就被去除，不再参加后面的拼接
4. 此步请将测序不准确或认为是载体的序列用这种方法去除。



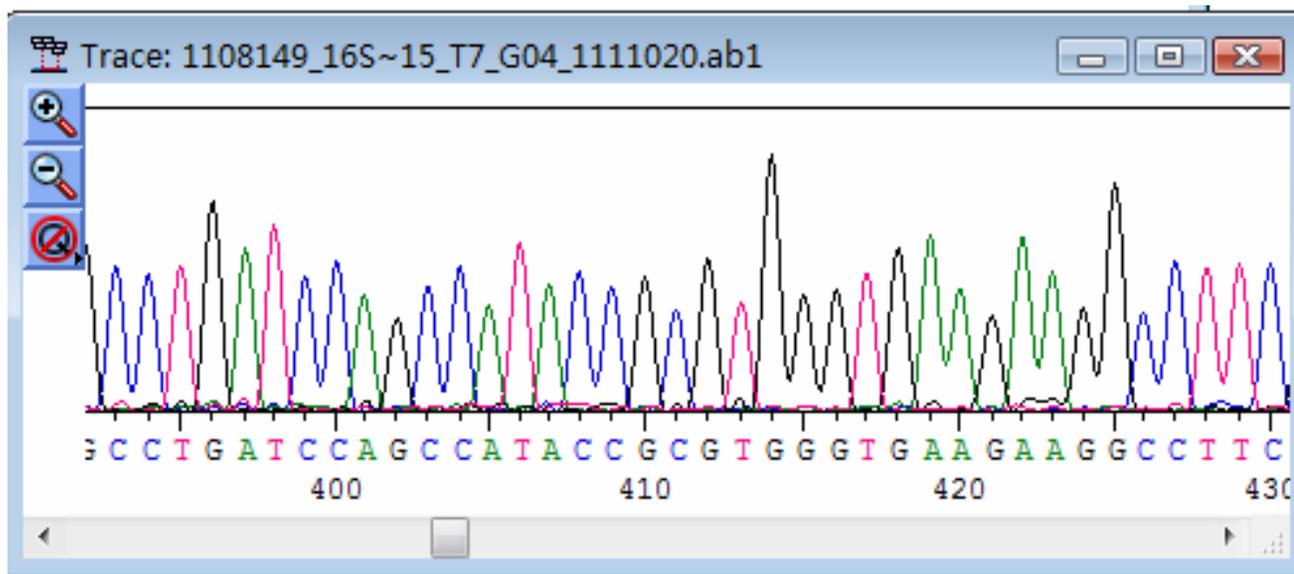
# 手工去除侧翼序列





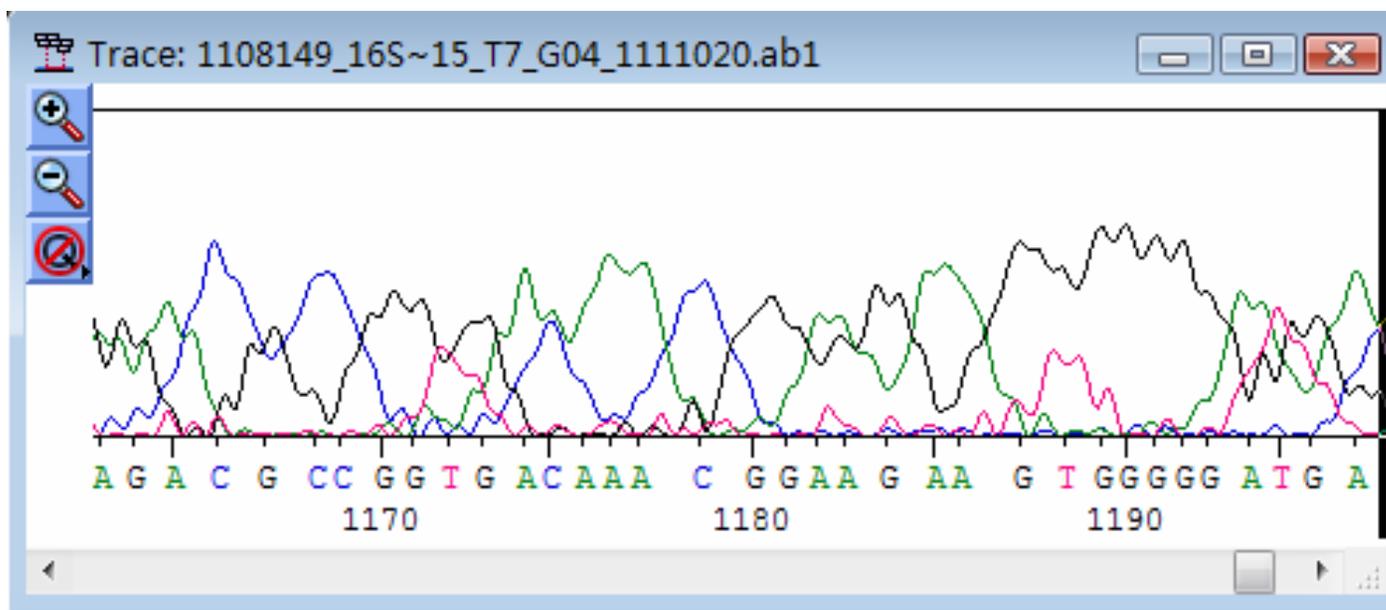
# 测序准确的峰形图

- 峰形规则，一般在序列的中部，如下图所示



# 测序不准确的峰形图

- 峰形较乱，很难判断是哪个碱基，一般位于序列两端，如下图所示



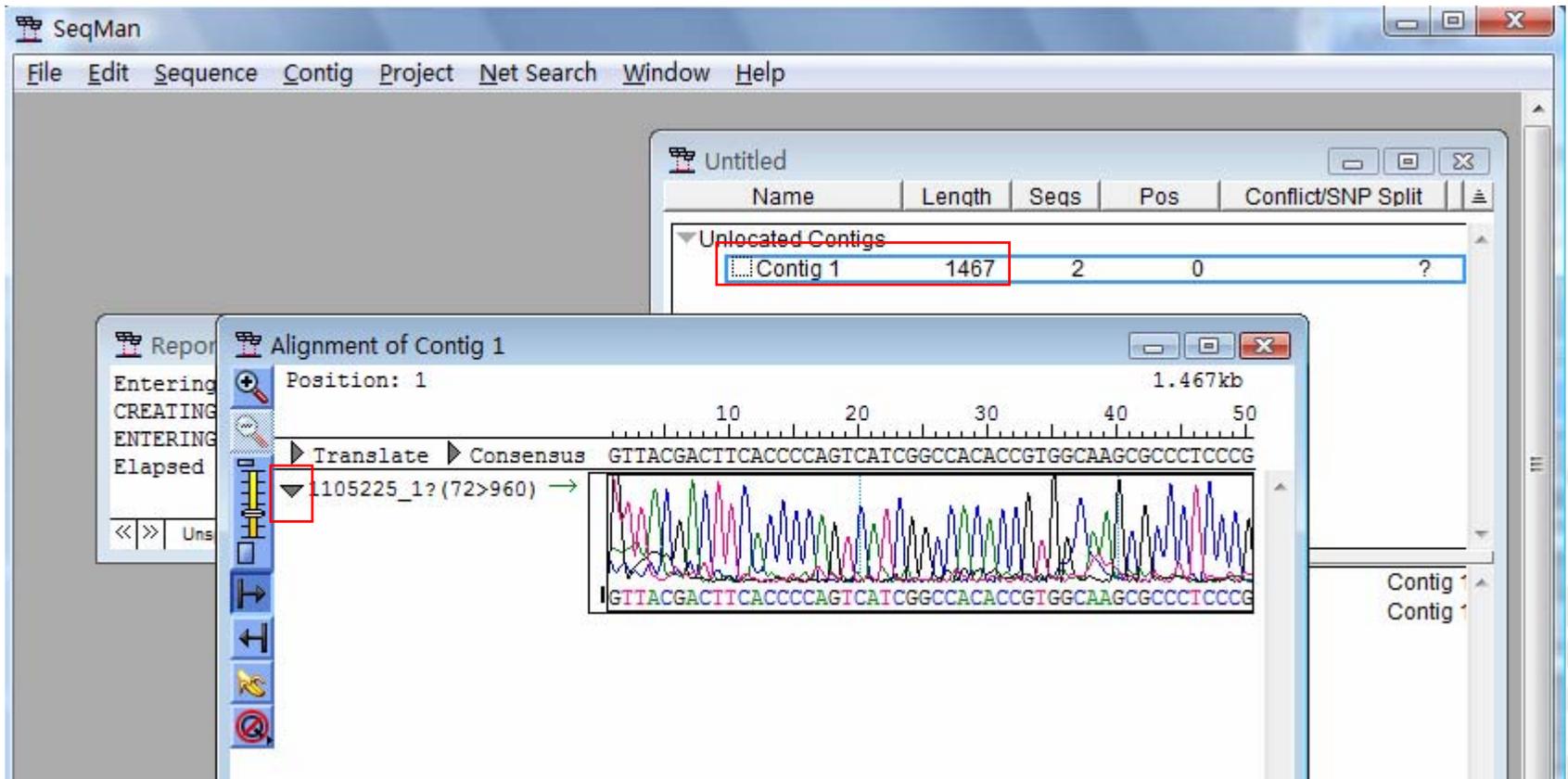


# Step4:进行序列拼接

- 点击Assemble按钮
- 在新出现窗口处点击拼接好的contig1
- 在出现的Alignment of contig1 窗口中点击左三角显示序列的测序图谱
- 点击菜单contig → strategy view可以观察序列拼接的宏观图



# 序列拼接图





# Step5:查找拼接错误

- 点击菜单Edit → find conflict
- 点击Find Previous或Find Next查找拼接中出现的错误
- 还可以通过Seqman左下角的快捷按钮查找错误的拼接



# 查找错误的拼接

SeqMan - [Alignment of Contig 1]

File Edit Sequence Contig Project Net Search Window Help

Selection: 954 -> 954 = 1 1.579kb

930 940 950 960

Translate Trace  
Majority ACACCTCTAGTTGKCCAGTTTCCACTGCAGTTCCAGGTTGAGCCCAGGG

1105225\_163?.ab1 (1>960) →  
ACACCTCTAATTTGCCAGTTTCCACTGTAGTTCCAGGTTGAG

1108149\_163? →  
ACACCTCTAATTTGCCAGTTTCCACTGTAGTTCCAGGTTGAG

AGGTTGAGCCCAGGG

AGGTTGAGCCCAGGG

Search

Conflict Search

Literal Search in: Consensus

Allow Ambiguity

Conflict

Conflict/SNP Split

Consensus Difference

Find Previous Find Next Cancel

<< >> 'Conflict'

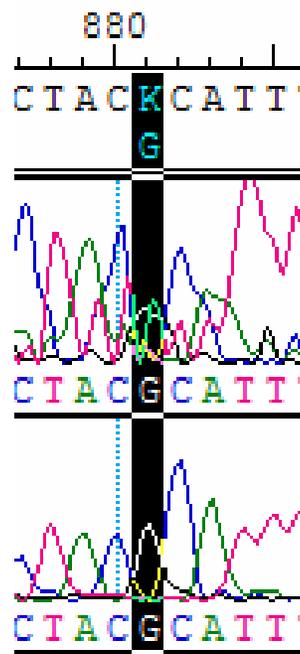


# 错误的拼接的类型

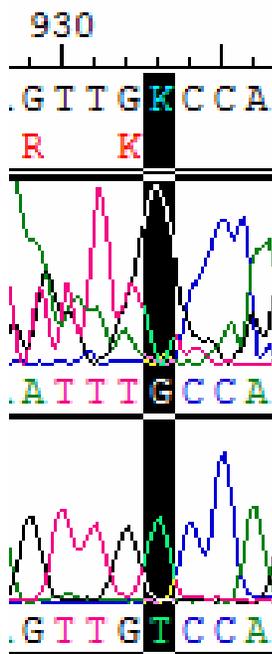
1. 两条序列的测序结果不一致并明显一条测序质量好而另一条质量差  
处理：直接将该处修改为正确的碱基
2. 两条序列的测序结果不一致并两条测序质量都比较差  
处理：重新测序或用新的合适引物重新测定
3. 两条序列的测序结果不一致并明显两条测序质量都好  
处理：测序过程出现问题，重新测定



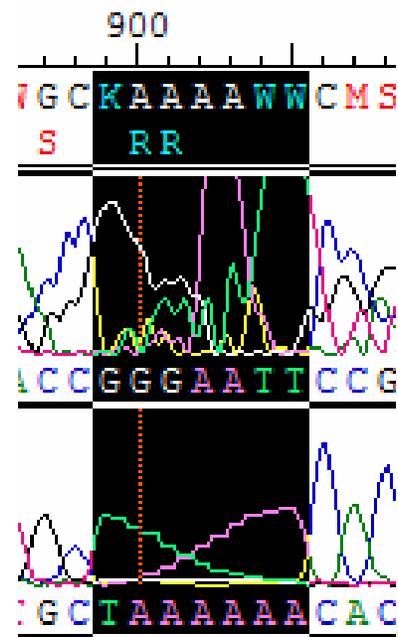
# 错误的拼接的类型



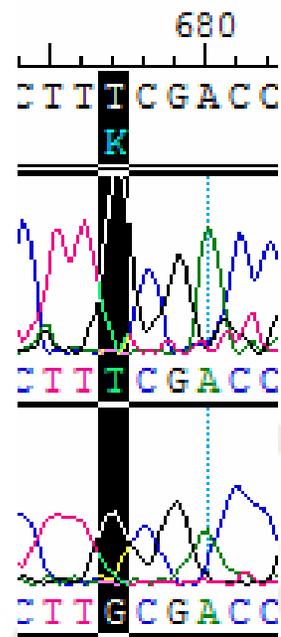
类型1



类型1



类型2

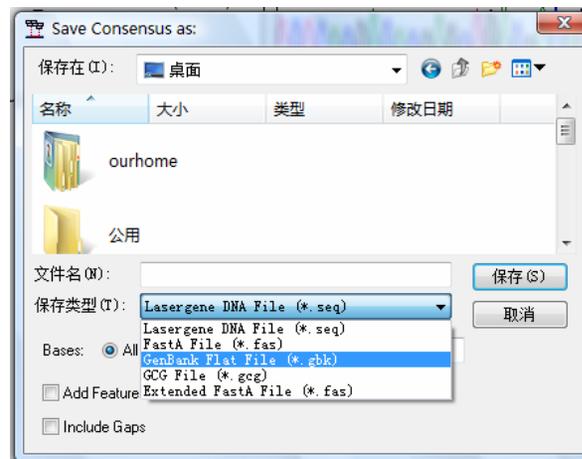
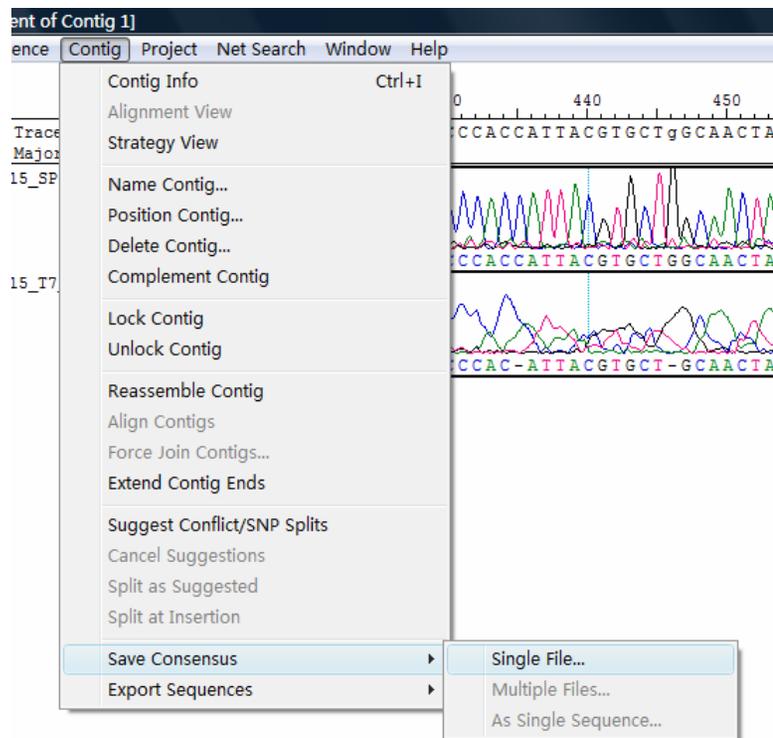


类型3



# Step6:导出拼接的序列

- 可选择合适的格式，导出拼接好的序列

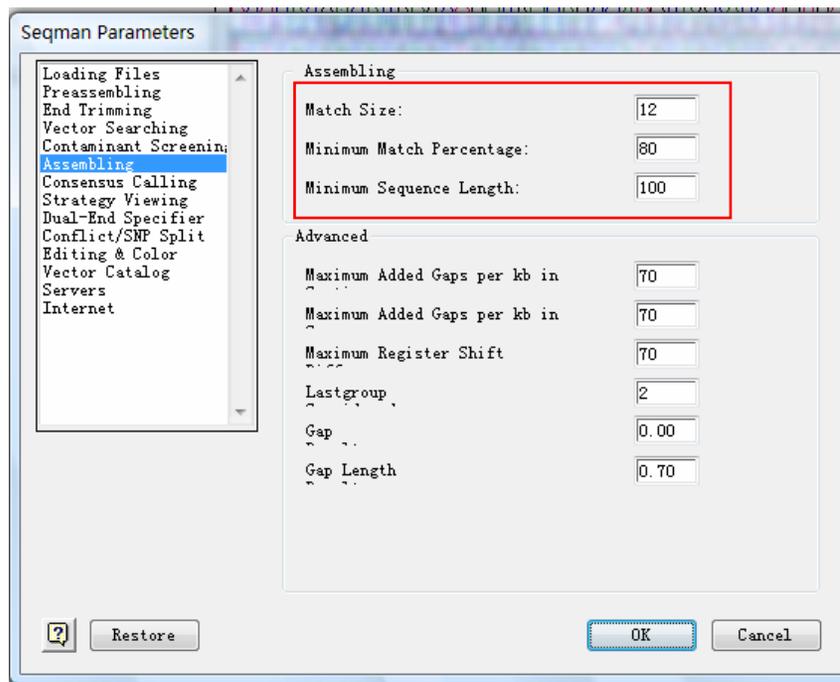
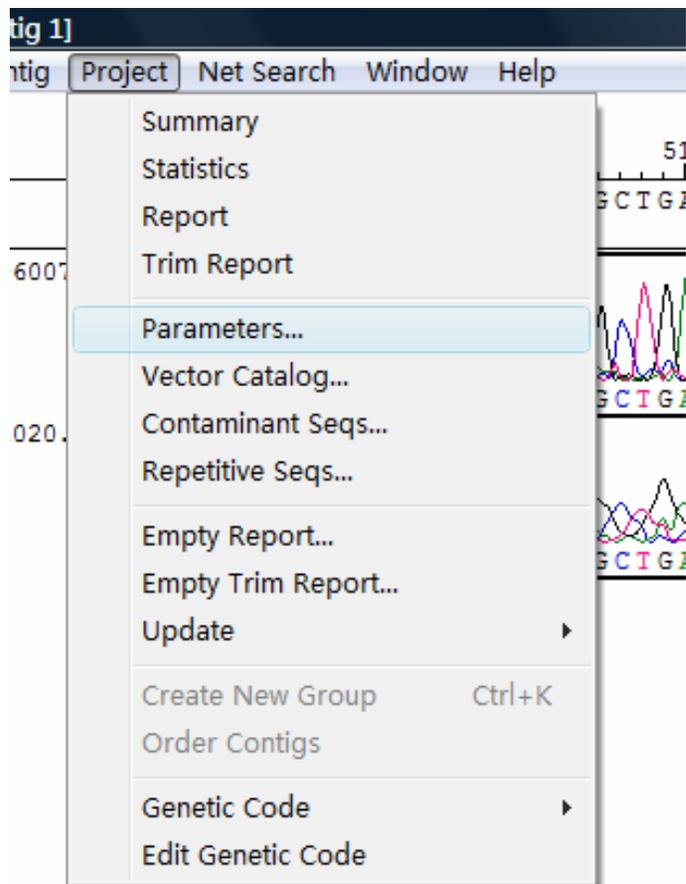




- 通过以上几步我们就能很快将几个测序片段进行拼接，大家可以拿着自己的序列试试!
- 当然如果两个测序片段的拼接片段太短可能利用默认的参数不能完成拼接，大家可以试着修改一下拼接参数试试!如降低Match size 及Minimum Match Percentage的值!



# 修改参数命令





- 关于Seqman的简单介绍就到这里，另外由于资料准备仓促，如有不完善或不准确的地方，请大家指出，谢谢！
- 下期讲座预告“轻松学用生物软件(3)”，具体时间及内容将在螺旋课堂上做出预告，欢迎大家参加！