利用SeqMan进行序列拼接

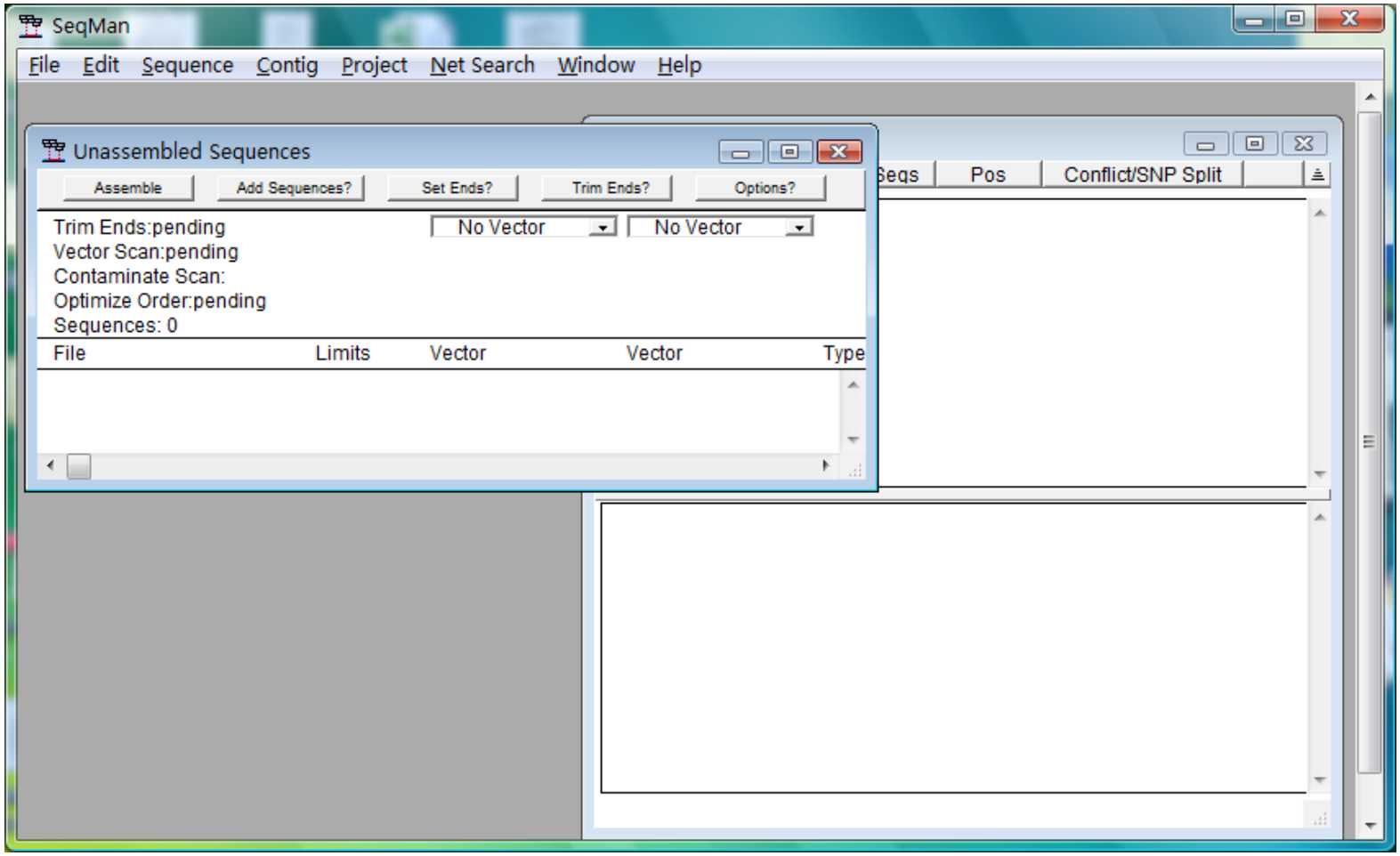
简介

- DNAStar是分子生物实验室常用的序列分析软件。 该软件由EditSeq， MegAlign，GeneQuest， MapDraw， PrimerSelect，Protean， SeqMan七个模块组成，功能主要有：序列的格式转换，序列拼接和重叠克隆群的处理；基因寻找；蛋白质结构域的查找；多重序列的比较和两两序列比较；寡核苷酸设计（PCR引物，测序引物，探针）。

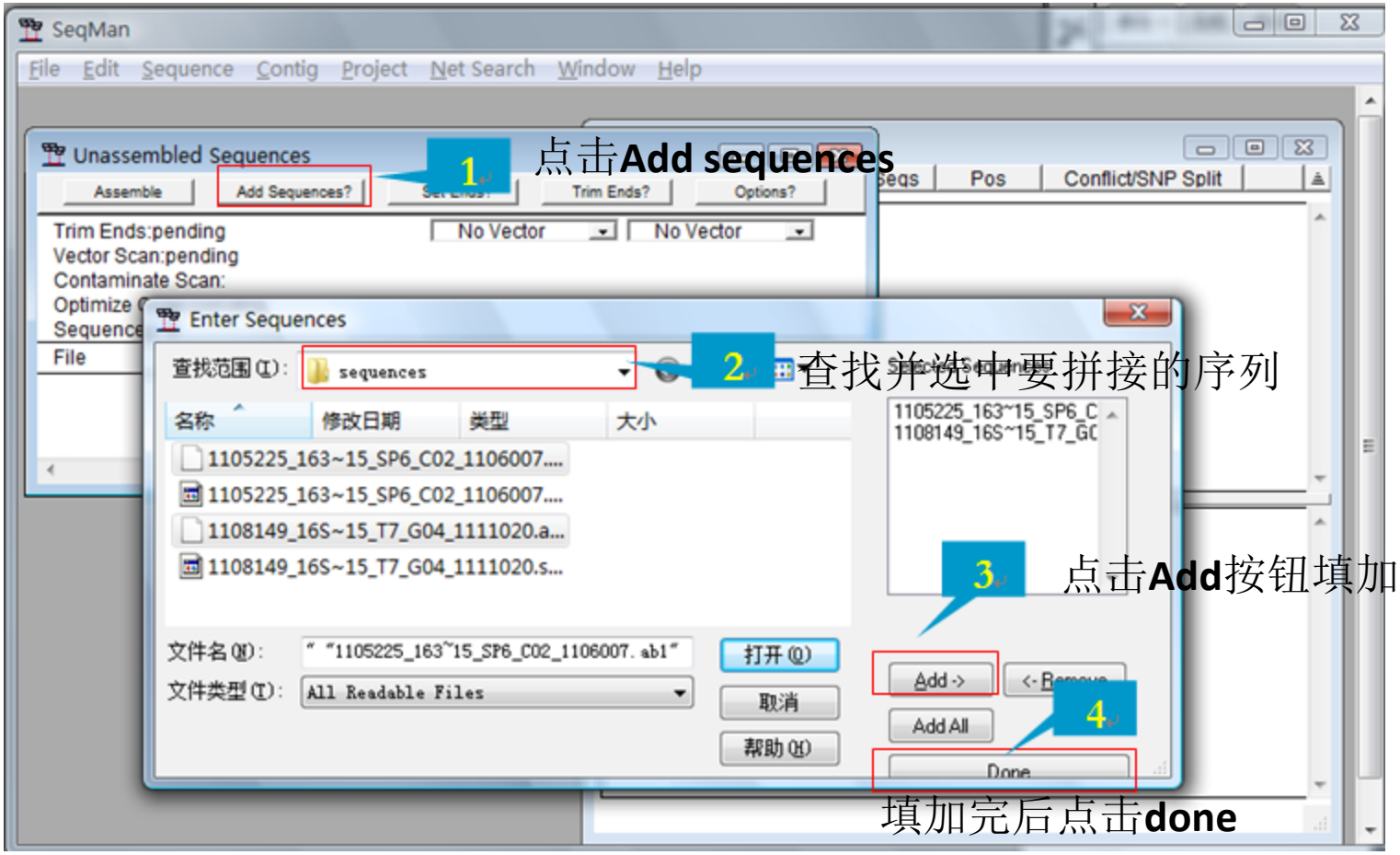
SeqMan

- SeqMan主要用于多序列的拼接，可以将成千上万的序列（最多可以支持64000多序列）装配成contigs，同时在拼接前，可以修整质量差的序列以及清除自动测序的序列结果中的污染序列或载体序列。seqman还提供完善的编辑和输出功能。

Step1:打开Seqman软件

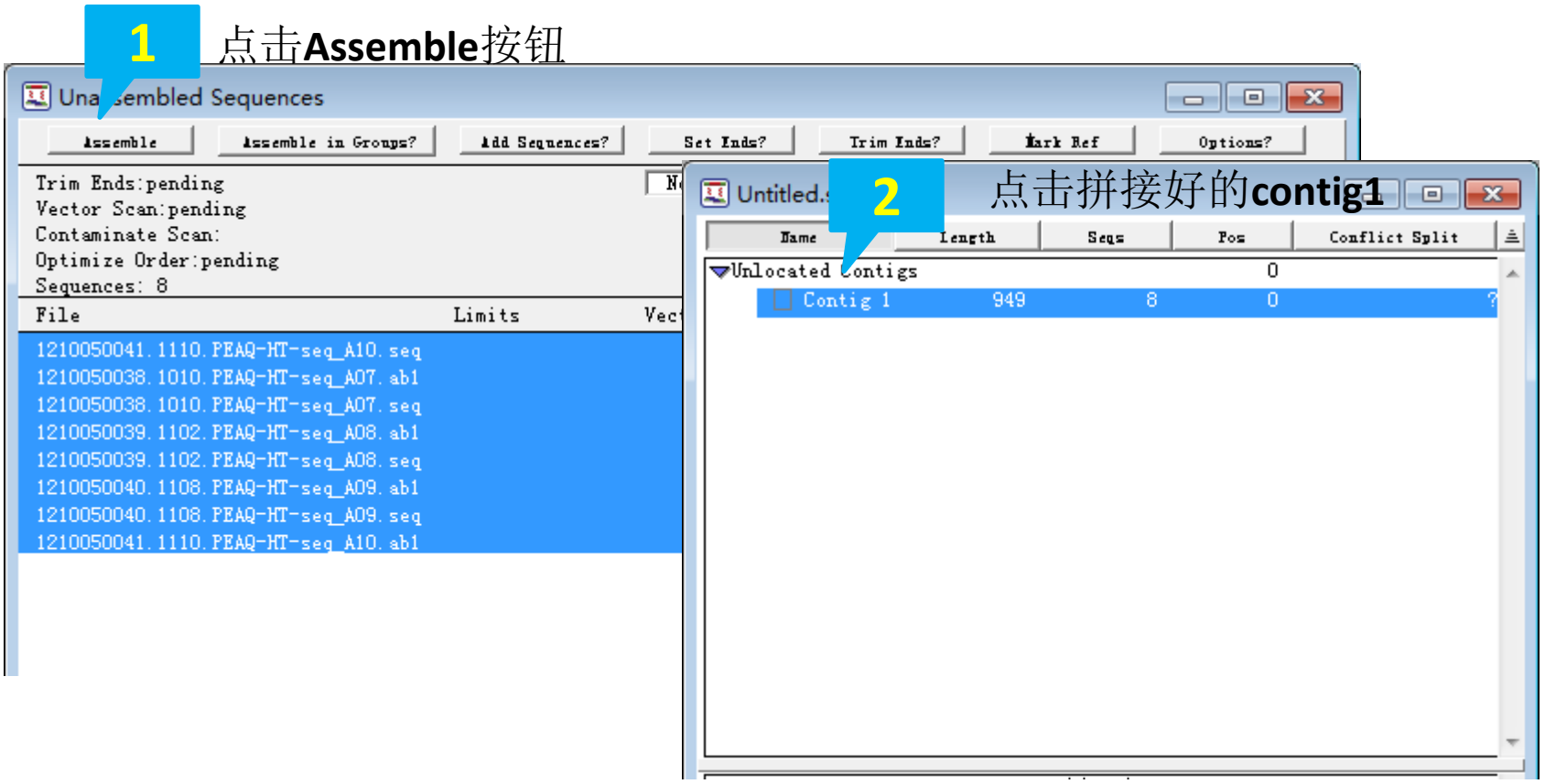


Step2:加入你要拼接的序列

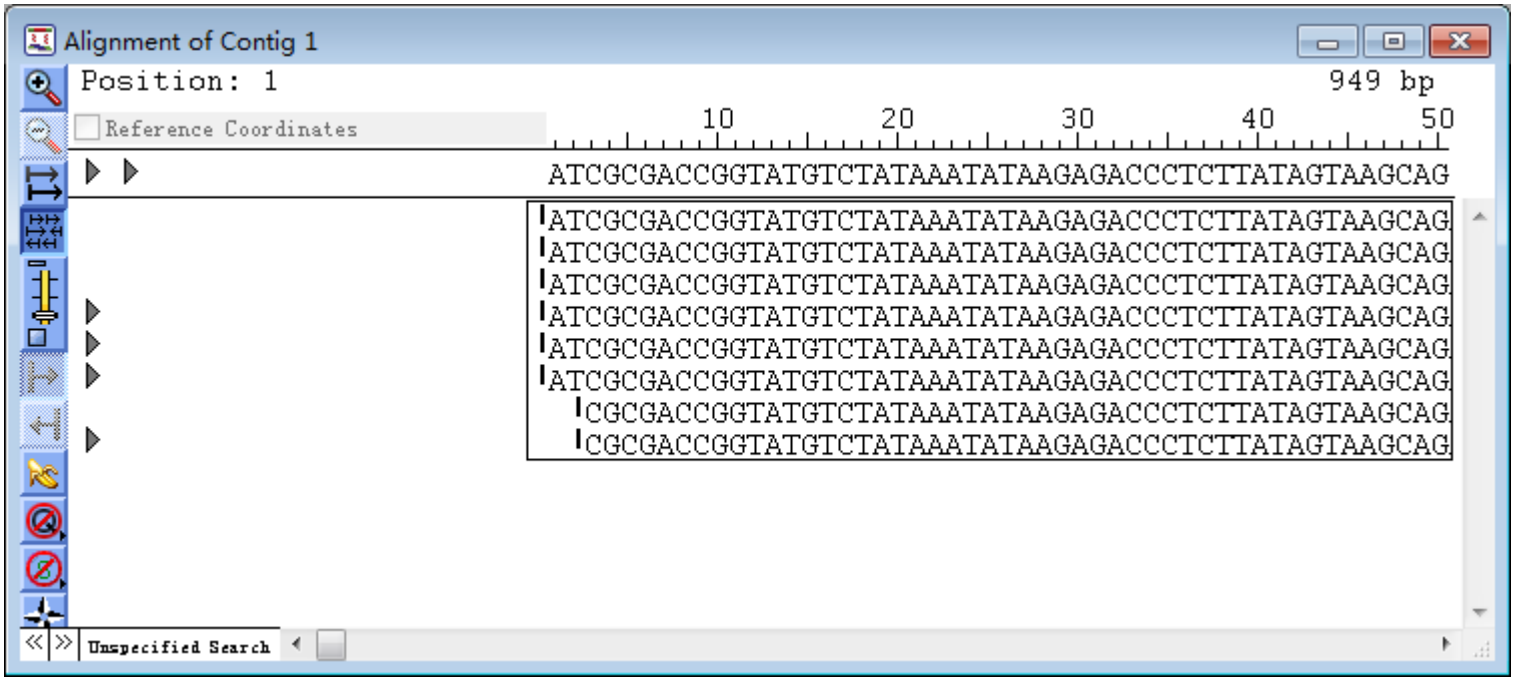


注：最好用测序的图谱（*.abi)尽量不要直接用测序得到的序列(.seq)

Step3:进行序列拼接



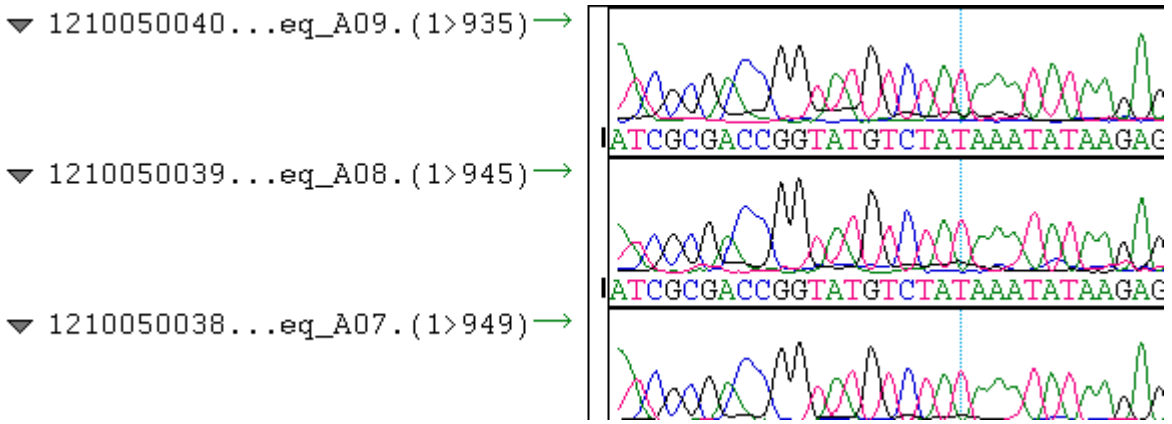
Step3:进行序列拼接



Alignment of contig1

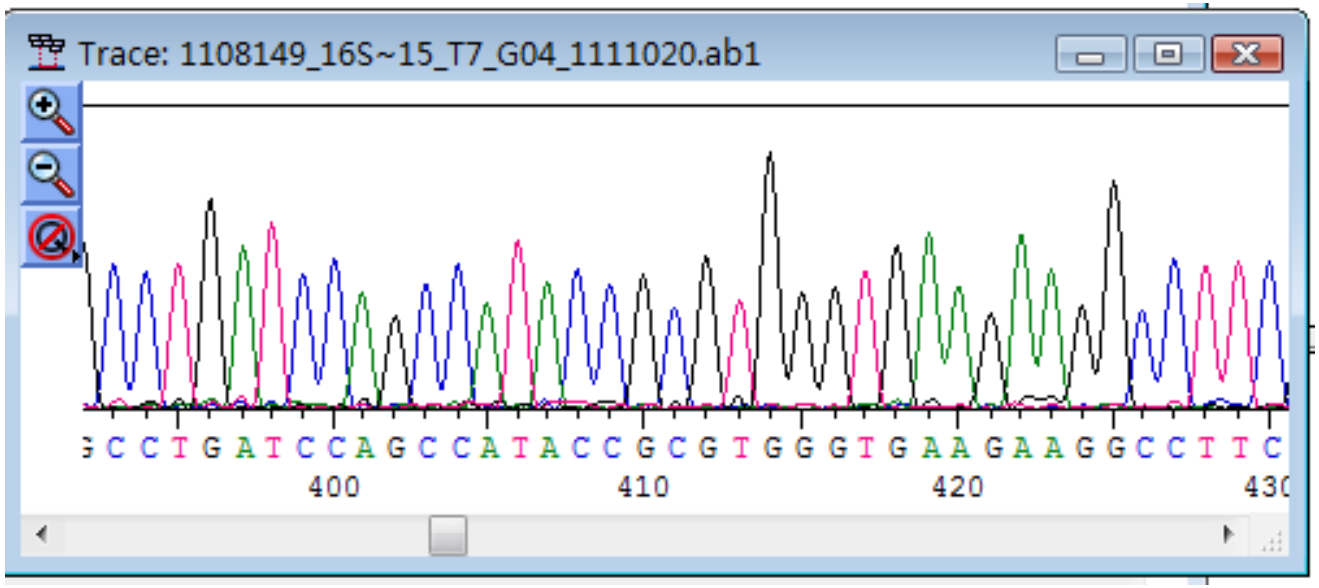
窗口中点击 左三角

显示序列的测序图谱



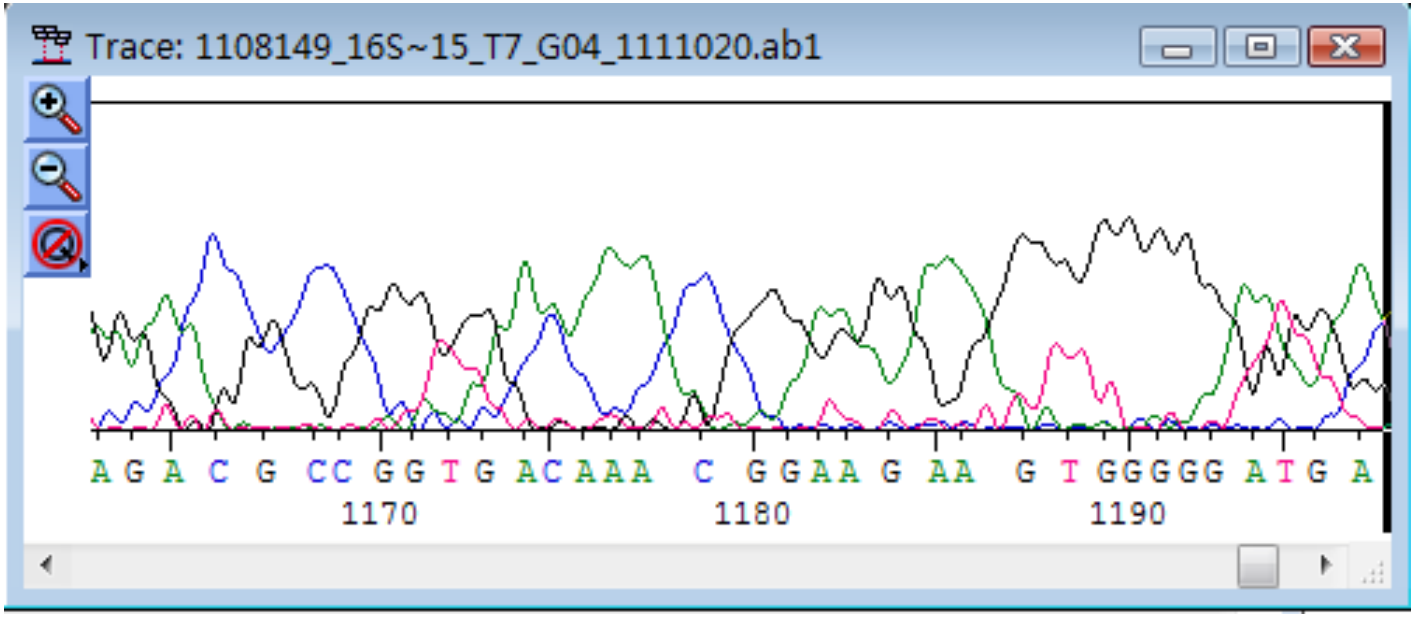
测序准确的峰形图

- 峰形规则，一般在序列的中部，如下图所示

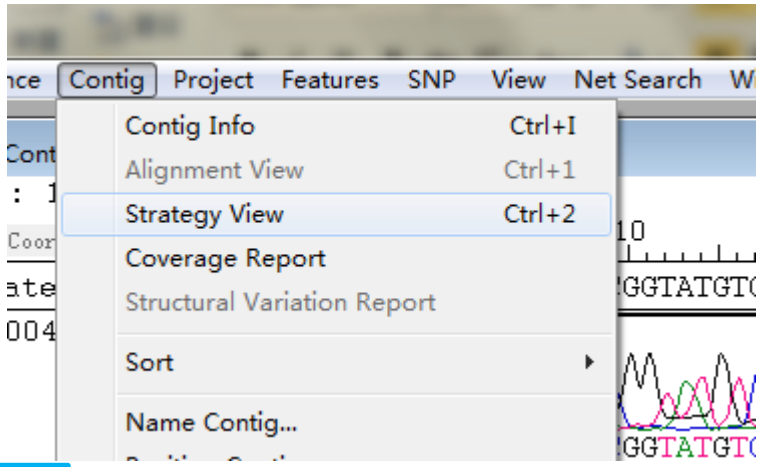


测序不准确的峰形图

- 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

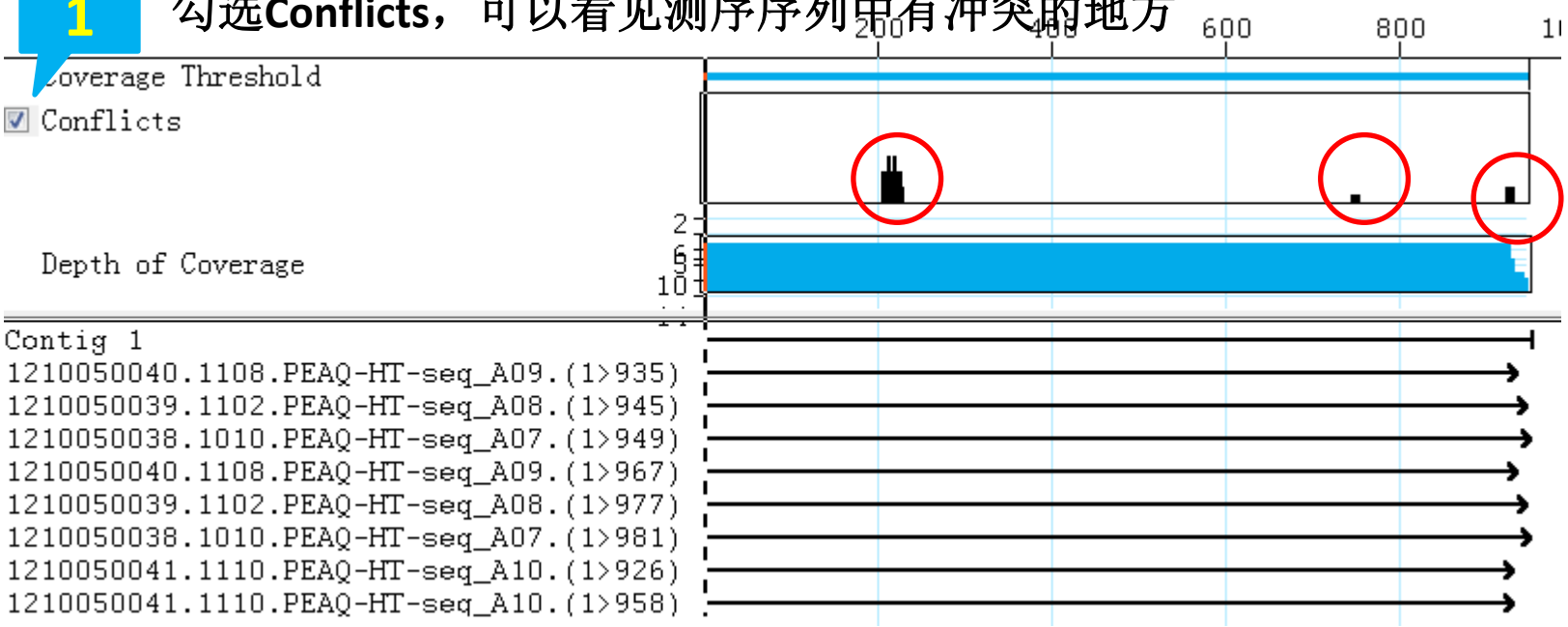


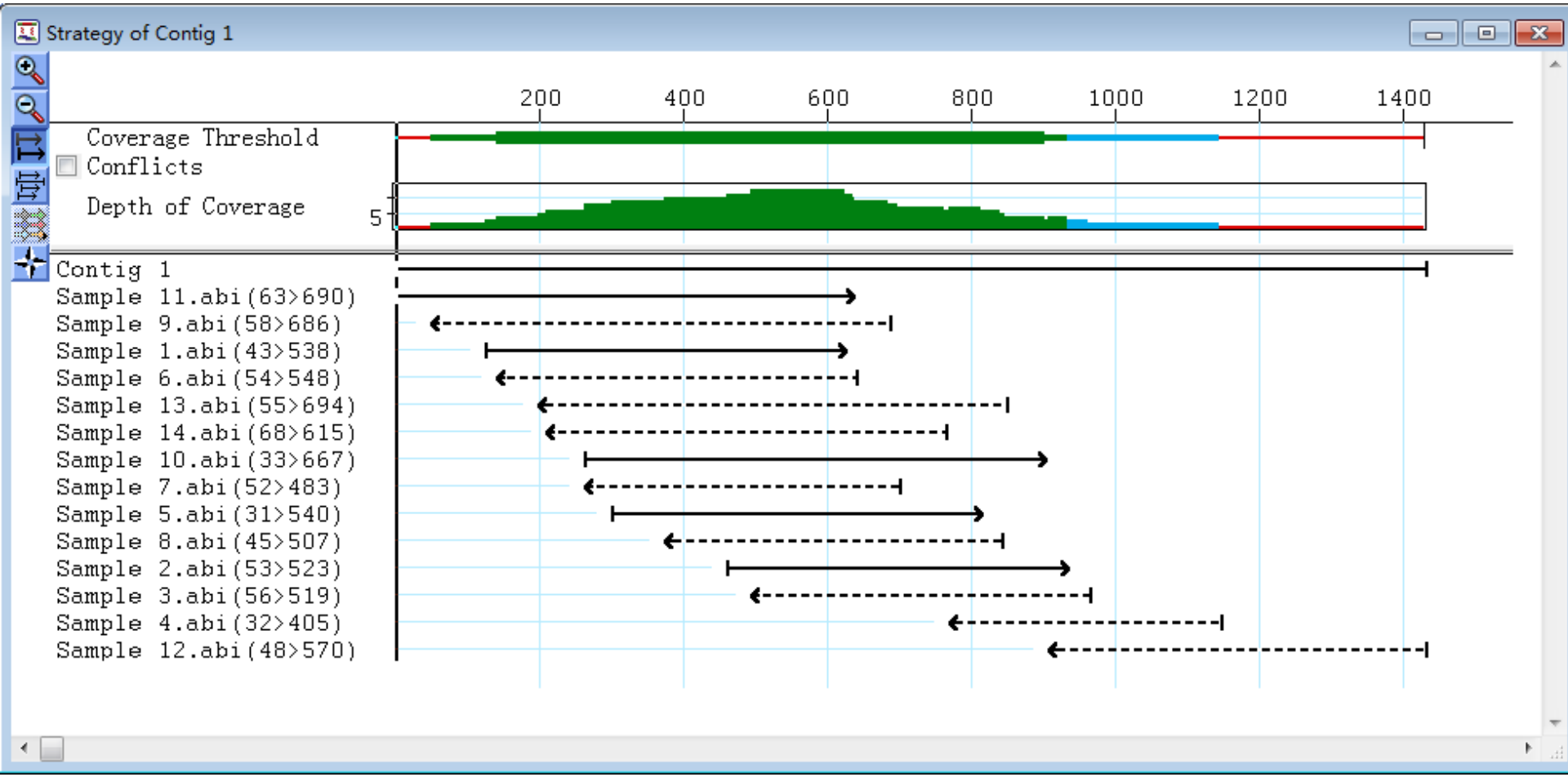
Step4:查找拼接错误



- 点击菜单**contig-strategy view**可以观察序列拼接的宏观图

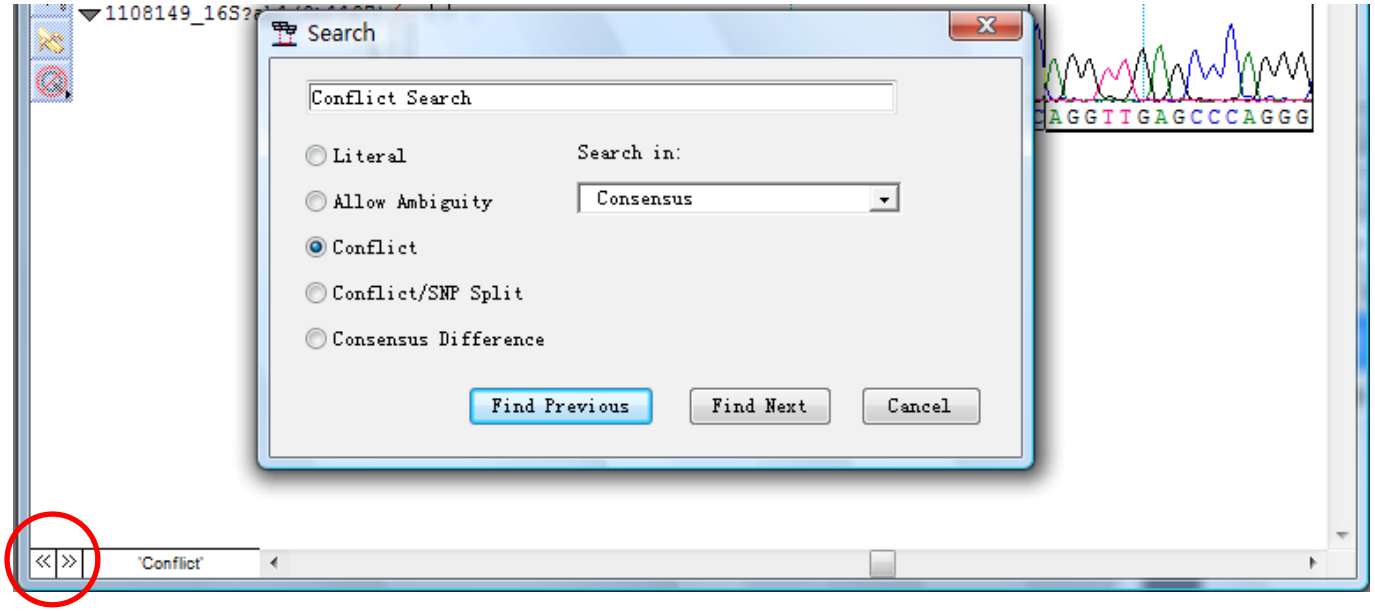
1 勾选**Conflicts**，可以看见测序序列中有冲突的地方



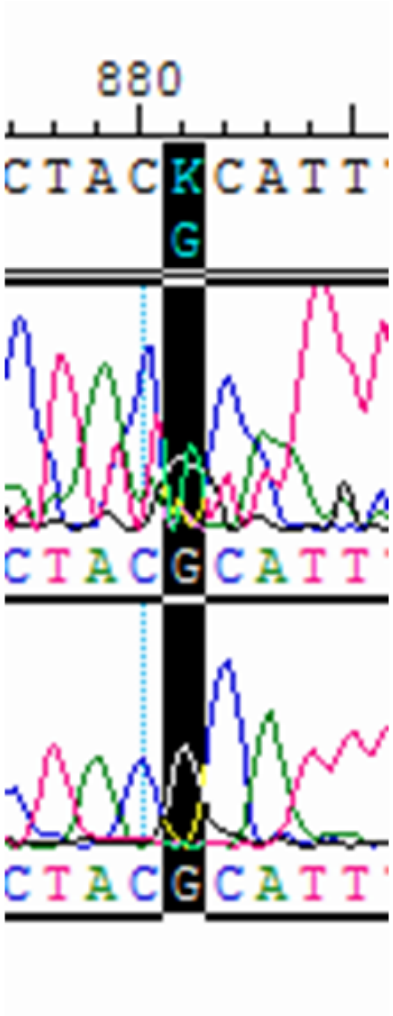


- 标尺下面的条带，显示了序列的数量和覆盖程度。条带的颜色和厚度代表覆盖序列的数量。中间的蓝线表示只有一条链被测序的区域。出现于 contig 两边的细红线表示这个区域只测过一次序。

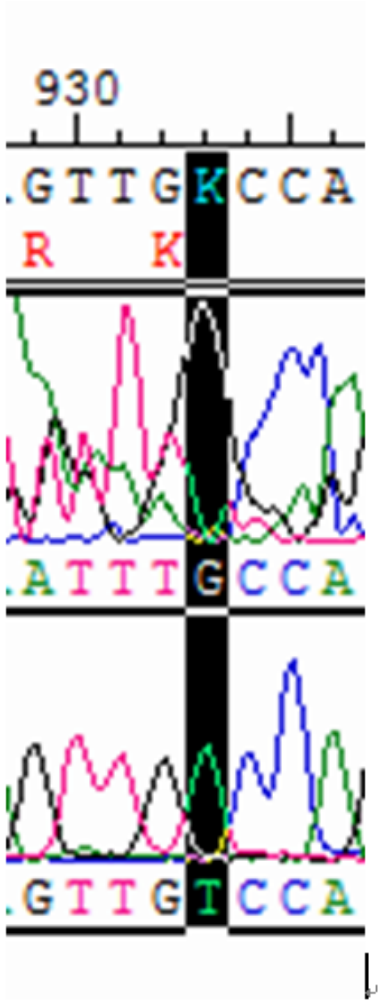
- 点击菜单**Edit -find conflict**或**Find Next**查找接接中 出现的错误
- 还可用左下角的快捷按钮查找错误的拼接



Step5：修改拼接错误



类型1



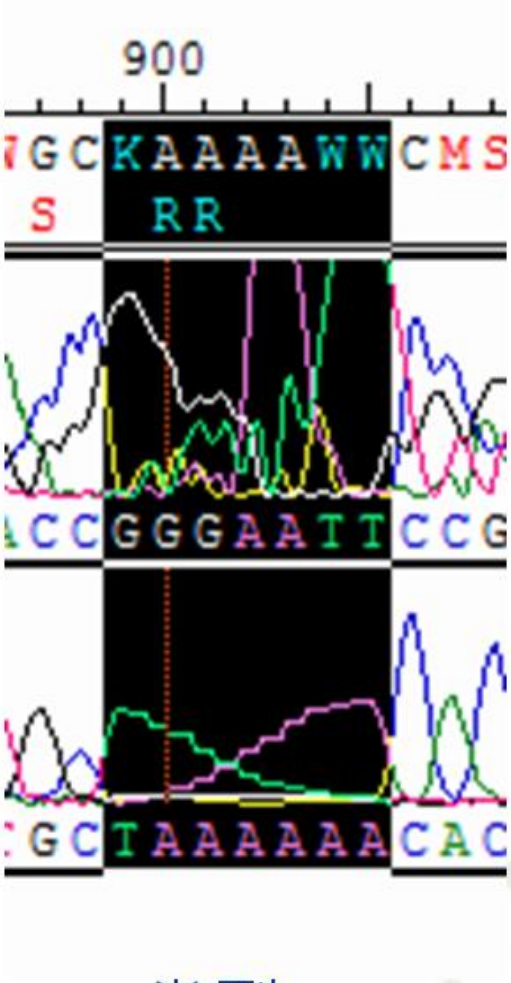
类型1

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

错误拼接的类型

Step5：修改拼接错误



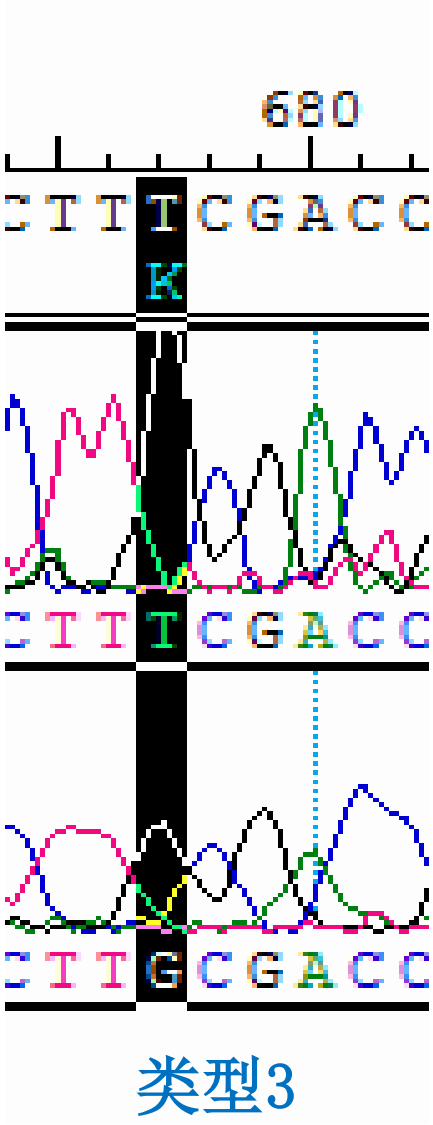
类型2

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

错误拼接的类型

Step5：修改拼接错误



类型3

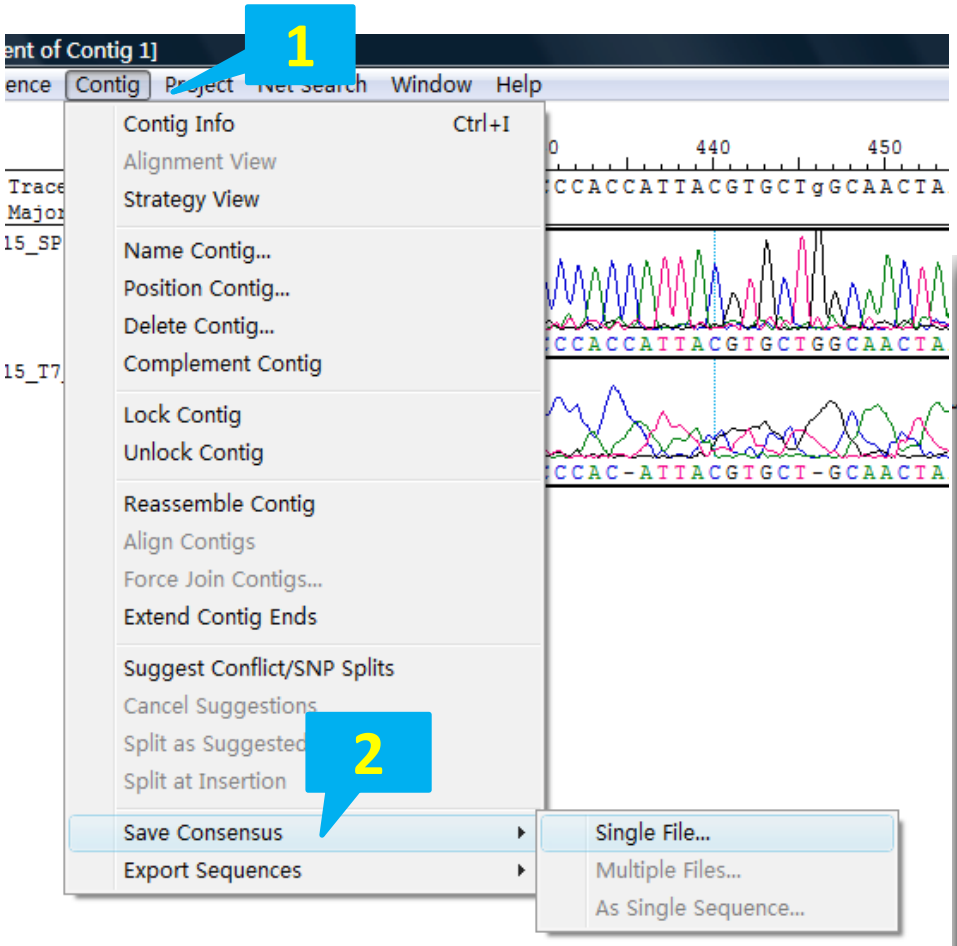
3. 两条序列的测序结果不一致并明显两条测序质量都好

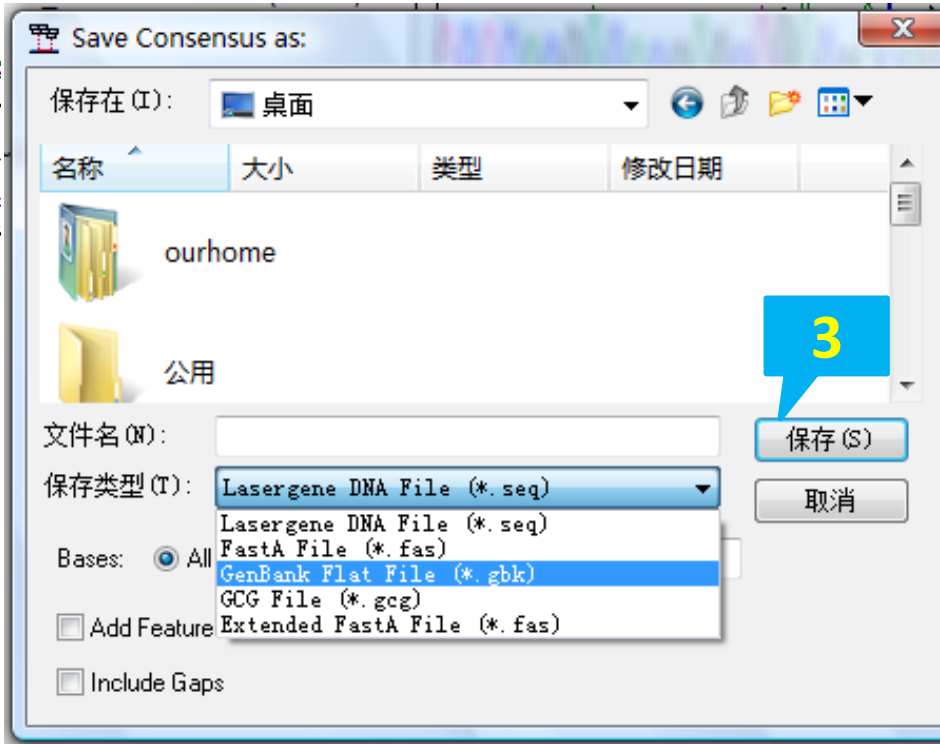
处理：测序过程出现问题，重新测定

错误拼接的类型

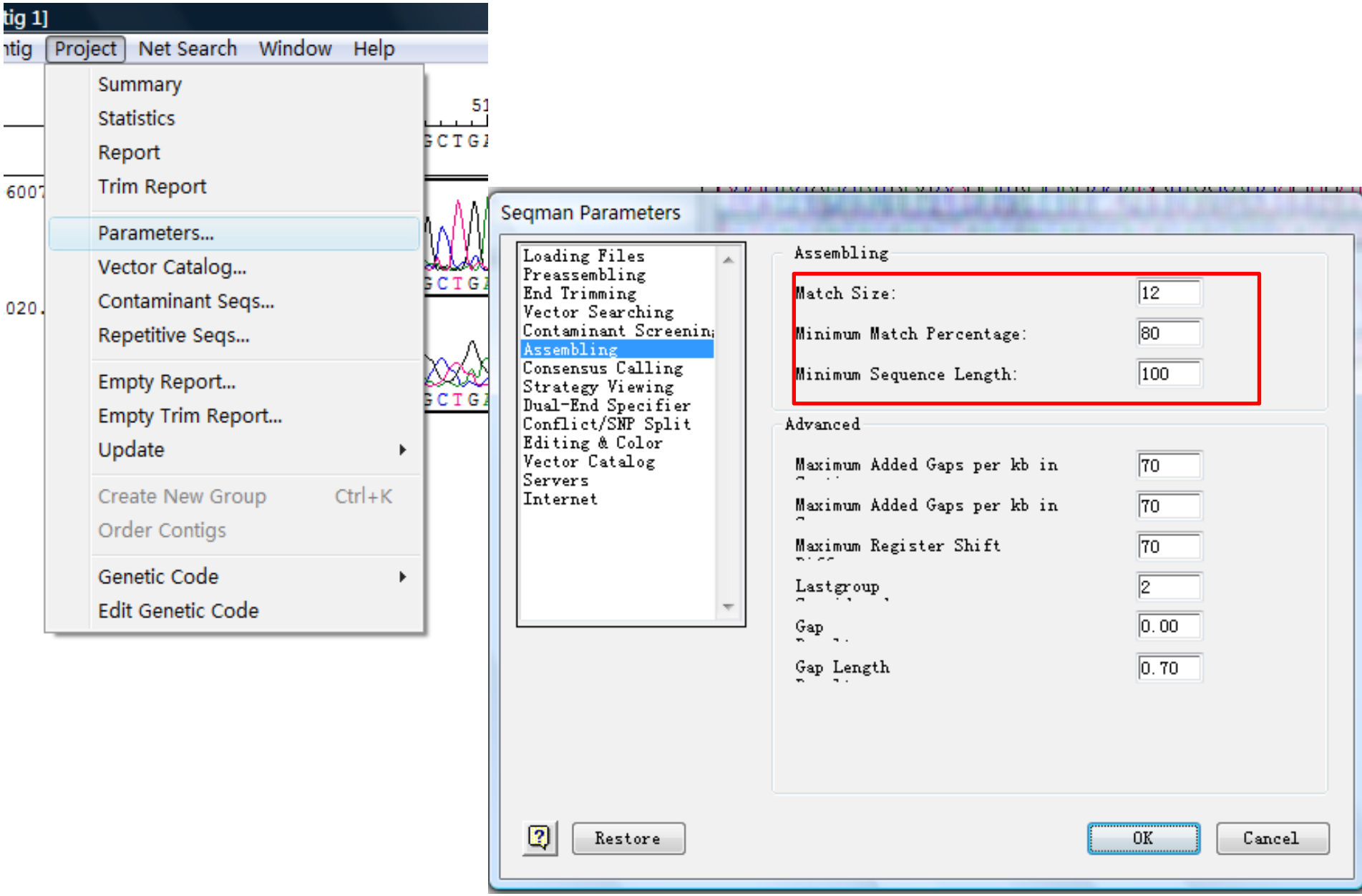
Step6:导出拼接的序列

- 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!


tig 1]
Project
Net Search
Window
Help
Summary
Statistics
Report
Trim Report
Parameters...
Vector Catalog...
Contaminant Seqs...
Repetitive Seqs...
Empty Report...
Empty Trim Report...
Update
Create New Group Ctrl+K
Order Contigs
Genetic Code
Edit Genetic Code
Seqman Parameters
Loading Files
Preassembling
End Trimming
Vector Searching
Contaminant Screenin
Assembling
Consensus Calling
Strategy Viewing
Dual-End Specifier
Conflict/SNP Split
Editing & Color
Vector Catalog
Servers
Internet
Assembling
Match Size: 12
Minimum Match Percentage: 80
Minimum Sequence Length: 100
Advanced
Maximum Added Gaps per kb in
70
Maximum Added Gaps per kb in
70
Maximum Register Shift
70
Lastgroup
2
Gap
0.00
Gap Length
0.70
Restore
OK
Cancel

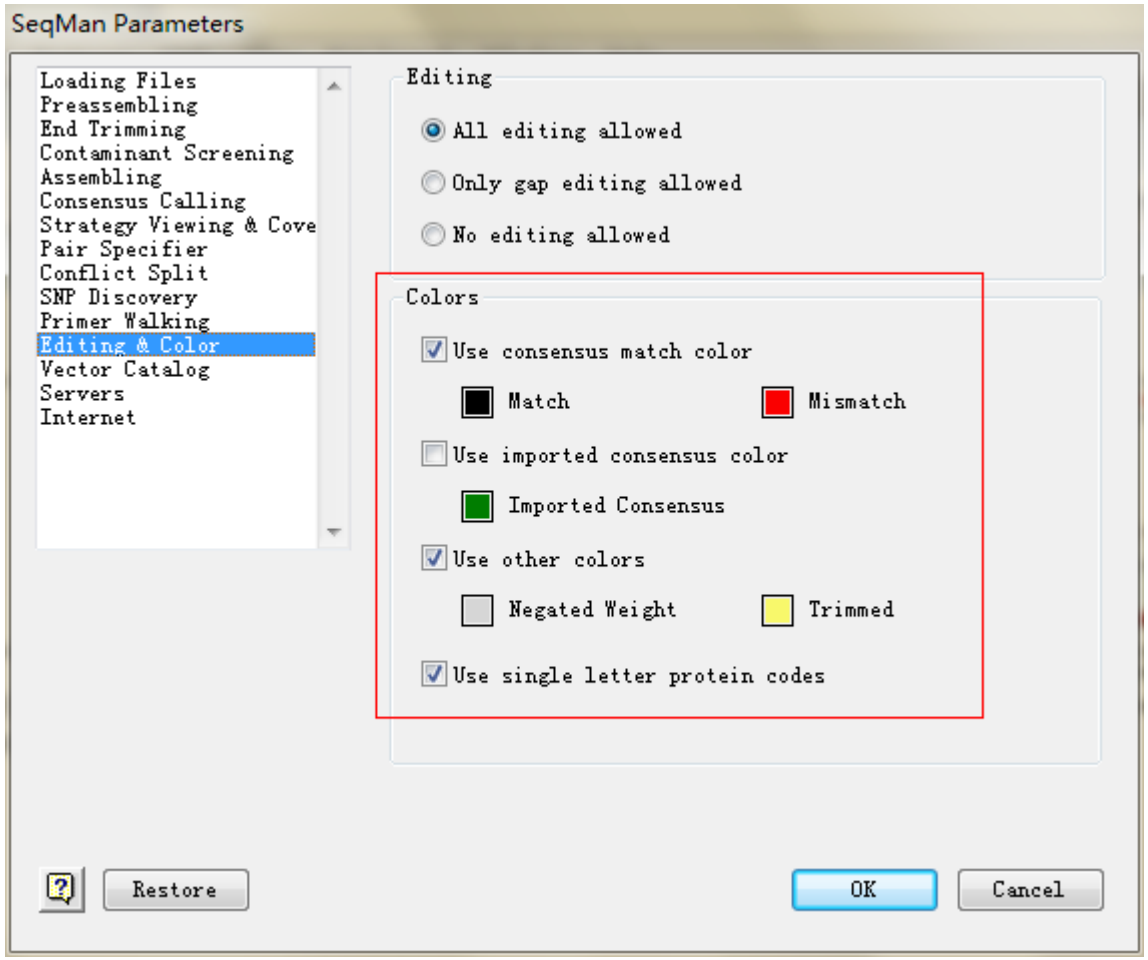
SeqMan进阶

- 手工及自动去除末端
- 消除载体或污染的序列
- 网络比对下载同源相似序列

手动修改序列末端

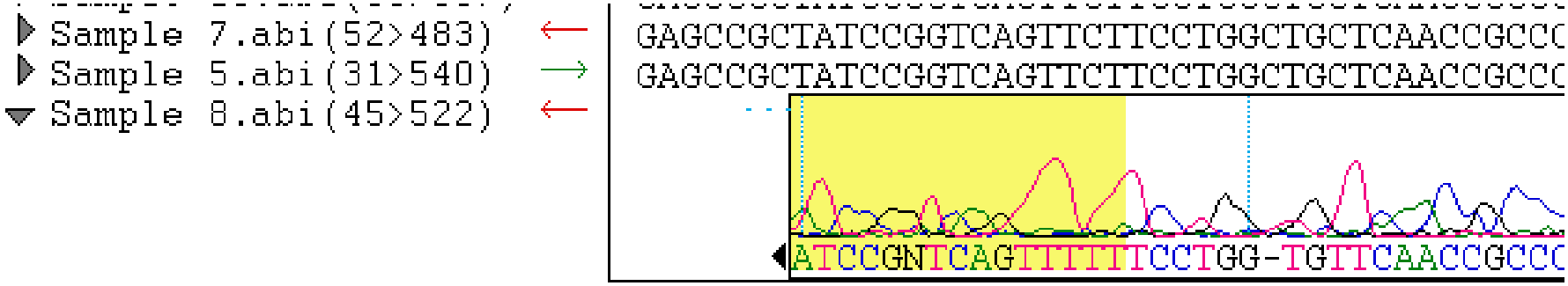
- SeqMan根据trace数据的质量和载体序列在装配之前可以自动地进行末端修整。然而有时候修改的程度难以掌握，下面我们将用手工的方法找回修整过的末端。

手动修改



- 为了区分修整过和没有修整过的数据，我们给修整过的数据加一个有颜色的背景。选择菜单Project→Parameters→Editing Color打开下面的对话框。确定use consensus match color和use other color已被选中。

手动修改

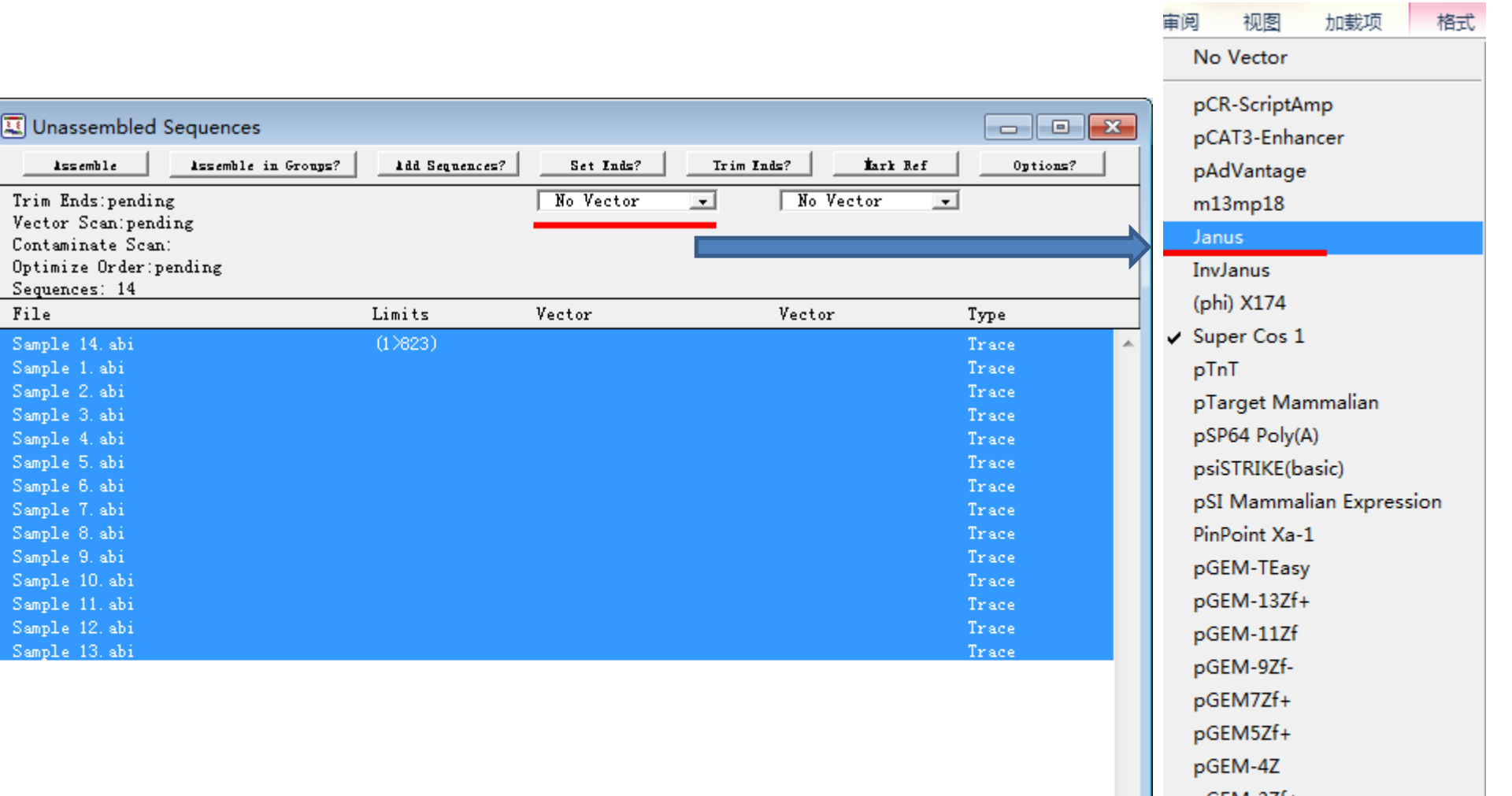


- 修整完毕后 Alignment View 中在序列的左边会有一个黑色的垂直棒，右边有一个小的黑三角形。
- 要找回修整去掉的序列末端，只需把垂直棒向序列的两端拖动即可，以前修整去掉的序列有明亮的黄色背景。

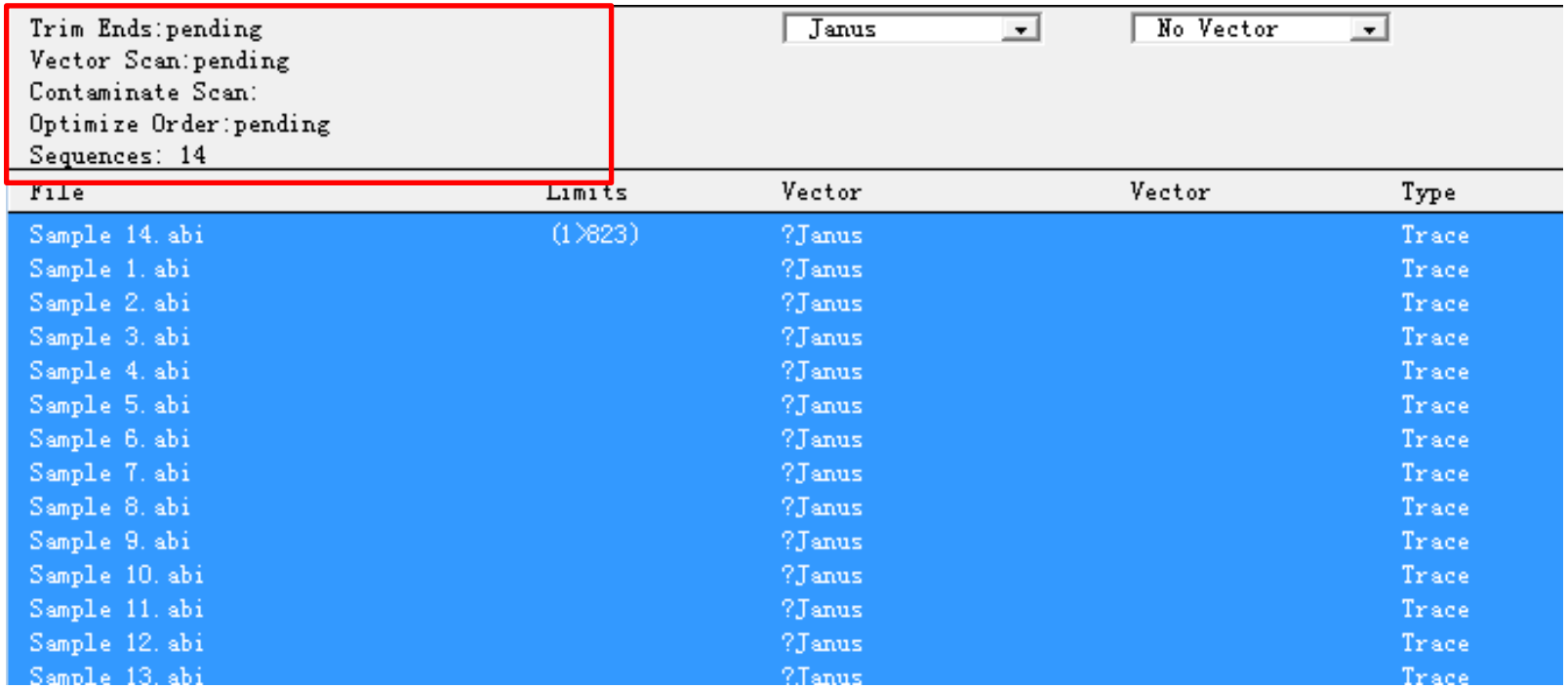
Pre-Assembly Options 操作及序列装配

- 在拼接前面，可以将所要拼接的片段中清除载体和污染序列，优化装配顺序，设定片段末端和标记重复序列

去除载体序列

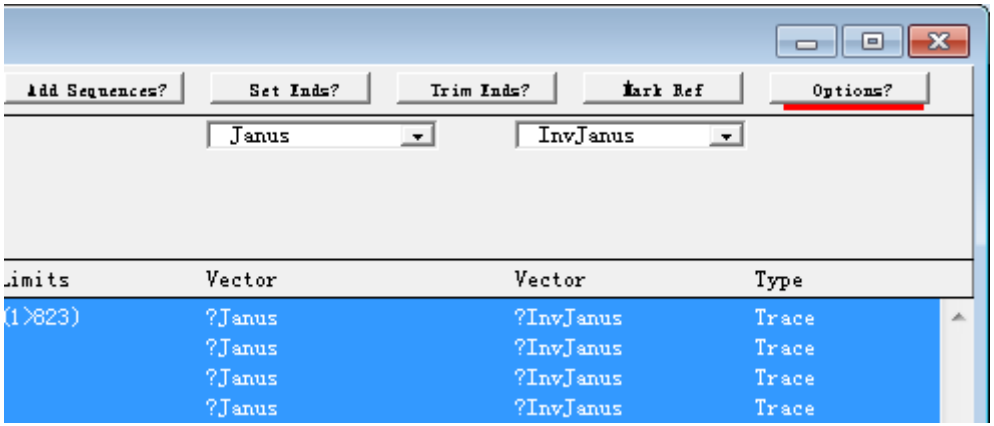

Unassembled Sequences
Assemble
Assemble in Groups?
Add Sequences?
Set Ends?
Trim Ends?
Mark Ref
Options?
Trim Ends:pending
Vector Scan:pending
Contaminate Scan:
Optimize Order:pending
Sequences: 14
No Vector
No Vector
File
Limits
Vector
Vector
Type
Sample 14.abi (1>823) Trace
Sample 1.abi Trace
Sample 2.abi Trace
Sample 3.abi Trace
Sample 4.abi Trace
Sample 5.abi Trace
Sample 6.abi Trace
Sample 7.abi Trace
Sample 8.abi Trace
Sample 9.abi Trace
Sample 10.abi Trace
Sample 11.abi Trace
Sample 12.abi Trace
Sample 13.abi Trace
审阅
视图
加载项
格式
No Vector
pCR-ScriptAmp
pCAT3-Enhancer
pAdVantage
m13mp18
Janus
InvJanus
(phi) X174
Super Cos 1
pTnT
pTarget Mammalian
pSP64 Poly(A)
psiSTRIKE(basic)
pSI Mammalian Expression
PinPoint Xa-1
pGEM-TEasy
pGEM-13Zf+
pGEM-11Zf
pGEM-9Zf-
pGEM7Zf+
pGEM5Zf+
pGEM-4Z

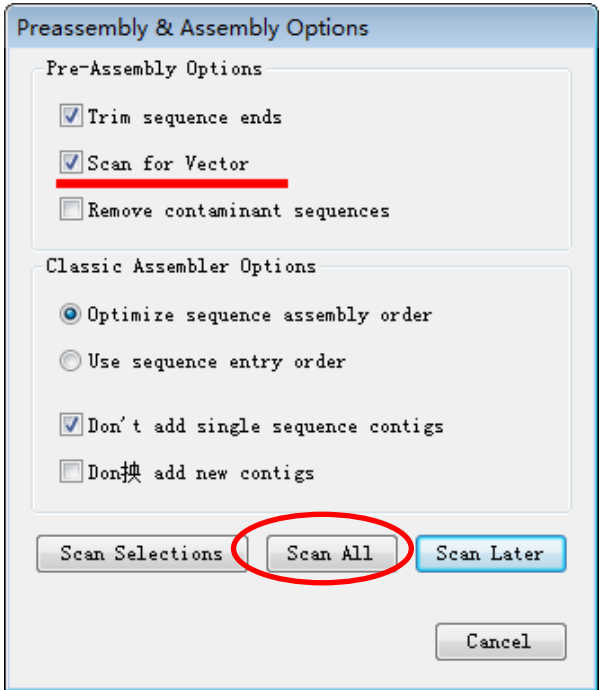
去除载体序列


Trim Ends:pending
Vector Scan:pending
Contaminate Scan:
Optimize Order:pending
Sequences: 14
Janus
No Vector
File
Limits
Vector
Vector
Type
Sample 14.abi
(1>823)
?Janus
Trace
Sample 1.abi
?Janus
Trace
Sample 2.abi
?Janus
Trace
Sample 3.abi
?Janus
Trace
Sample 4.abi
?Janus
Trace
Sample 5.abi
?Janus
Trace
Sample 6.abi
?Janus
Trace
Sample 7.abi
?Janus
Trace
Sample 8.abi
?Janus
Trace
Sample 9.abi
?Janus
Trace
Sample 10.abi
?Janus
Trace
Sample 11.abi
?Janus
Trace
Sample 12.abi
?Janus
Trace
Sample 13.abi
?Janus
Trace

去除载体序列

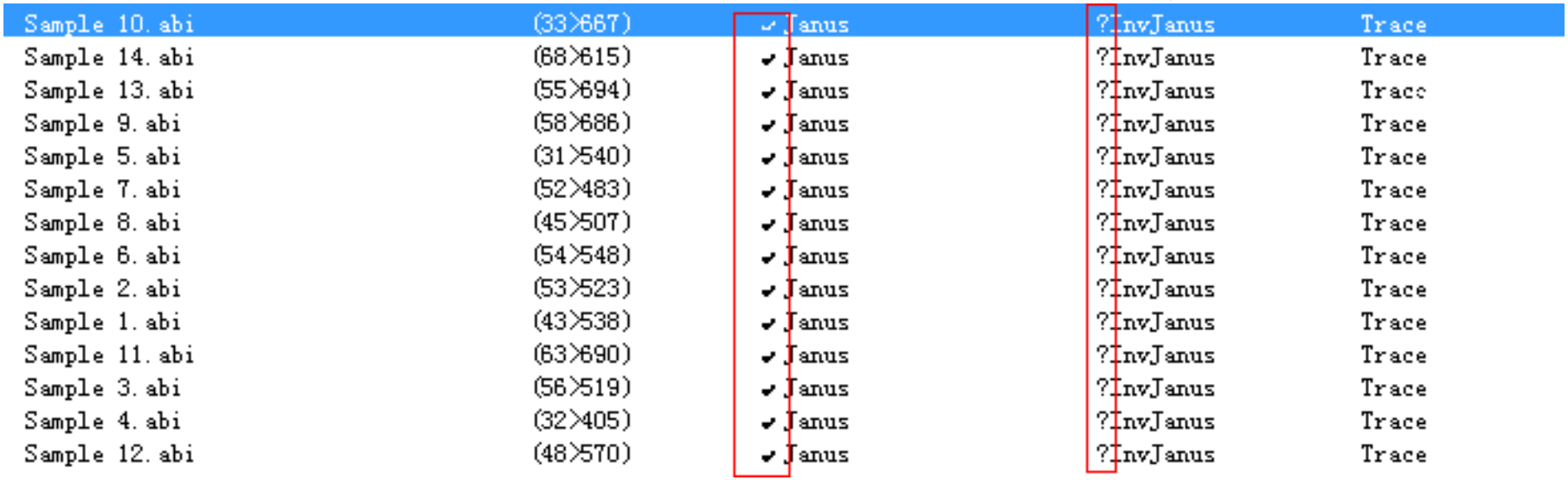
- seqman在拼接前，有--点击Unassembled Sequences窗口的右上角的“options“按钮选择相应功能。默认设定Trim sequence ends，scan for vector，optimize sequence assembly order.。





去除载体序列

- 单击 Scan All按钮，将出现一个report窗口。

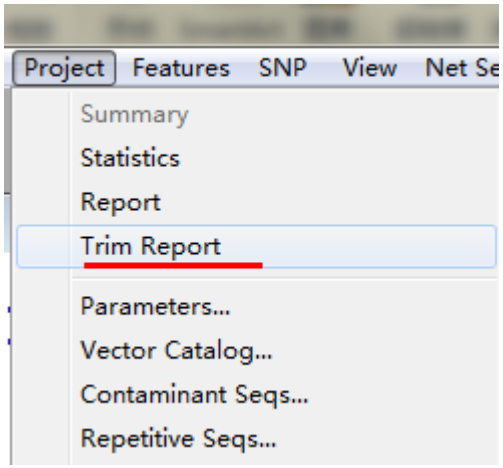


Sample 10.abi	(33>667)	✔Janus	?InvJanus	Trace
Sample 14.abi	(68>615)	✔Janus	?InvJanus	Trace
Sample 13.abi	(55>694)	✔Janus	?InvJanus	Trace
Sample 9.abi	(58>686)	✔Janus	?InvJanus	Trace
Sample 5.abi	(31>540)	✔Janus	?InvJanus	Trace
Sample 7.abi	(52>483)	✔Janus	?InvJanus	Trace
Sample 8.abi	(45>507)	✔Janus	?InvJanus	Trace
Sample 6.abi	(54>548)	✔Janus	?InvJanus	Trace
Sample 2.abi	(53>523)	✔Janus	?InvJanus	Trace
Sample 1.abi	(43>538)	✔Janus	?InvJanus	Trace
Sample 11.abi	(63>690)	✔Janus	?InvJanus	Trace
Sample 3.abi	(56>519)	✔Janus	?InvJanus	Trace
Sample 4.abi	(32>405)	✔Janus	?InvJanus	Trace
Sample 12.abi	(48>570)	✔Janus	?InvJanus	Trace

- 现在载体栏显示：载体名字前都有一个检测通过的标志，说明Janus 载体在全部14 序列中都已经检测到了。
- 单击assemble按钮，进行序列拼接。

查看末端修整和载体序列去除细节报告

- 选择Project 菜单的Trim Report打开Trim report窗口。



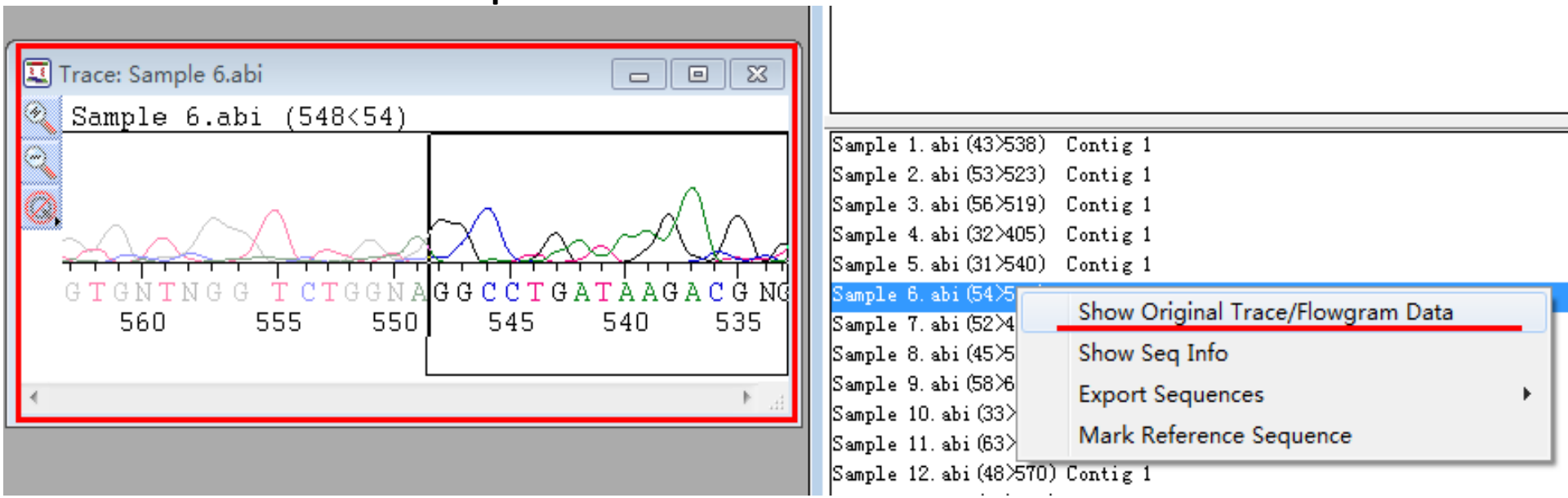
Trim Report

```
SCAN FOR VECTOR
    Janus
    InvJanus

TRIM ENDS
    MEDIUM TRIM PARAMETERS
        Trace:     Threshold    = 12
        Qual:      Threshold    = 12
         Non-Trace: Window Size = 50

                   Maximum Ns   = 2
                                AVERAGE  AMOUNT   TRIMMED  PRE-TRIM
            NAME                QUALITY  TRIMMED   LENGTH   LENGTH
"Sample 14.abi" (68>615)          26       275       548      823
"Sample 1.abi" (43>538)           33       245       496      741
"Sample 2.abi" (53>523)           28       389       471      860
"Sample 3.abi" (56>519)           32       318       464      782
"Sample 4.abi" (32>405)           34       443       374      817
"Sample 5.abi" (31>540)           30       283       510      793
"Sample 6.abi" (54>548)           28       337       495      832
"Sample 7.abi" (52>483)           26       375       432      807
"Sample 8.abi" (45>507)           30       389       463      852
"Sample 9.abi" (58>686)           26       141       629      770
"Sample 10.abi" (33>667)          29        67       635      702
"Sample 11.abi" (63>690)          28       209       628      837
"Sample 12.abi" (48>570)          30       269       523      792
"Sample 13.abi" (55>694)          26       168       640      808
-------------------------------------------------------------------
    14 Sequences      Average     29       279       522      801
```

Unspecified Search

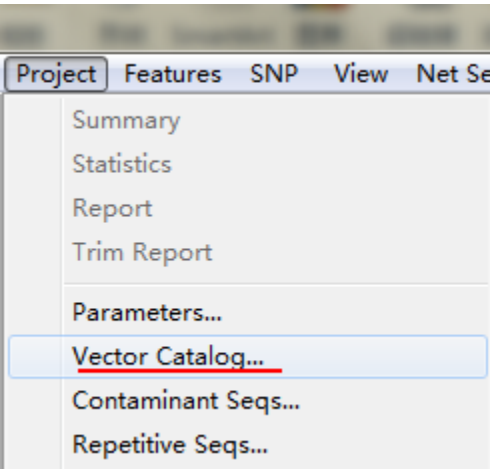
查看修整序列前后的跟踪数据

- 右键选择6 号样本，然后Show Original Trace Data,打开Trace：Sample 6.abi 窗口

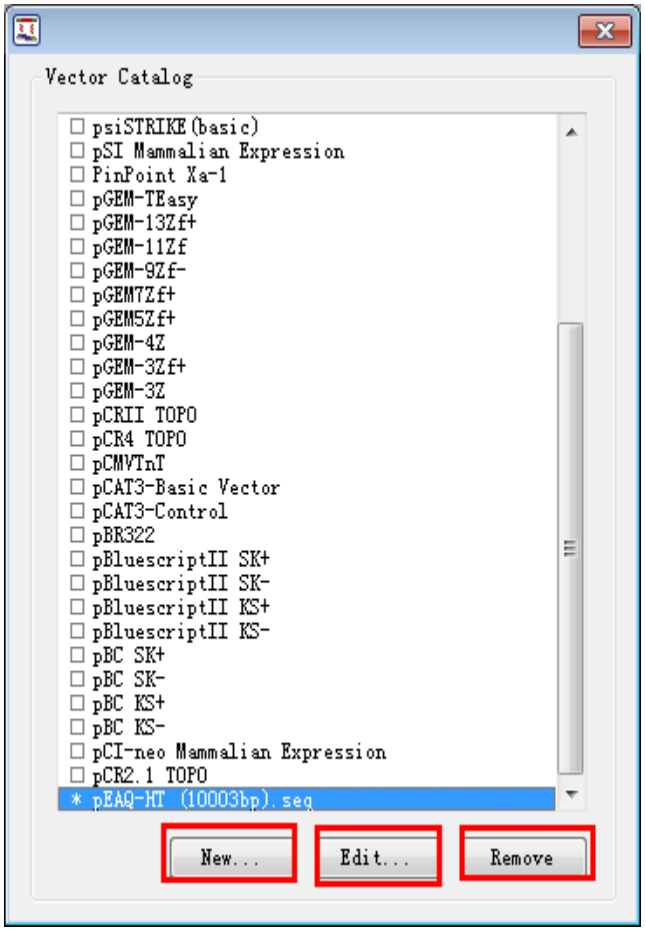


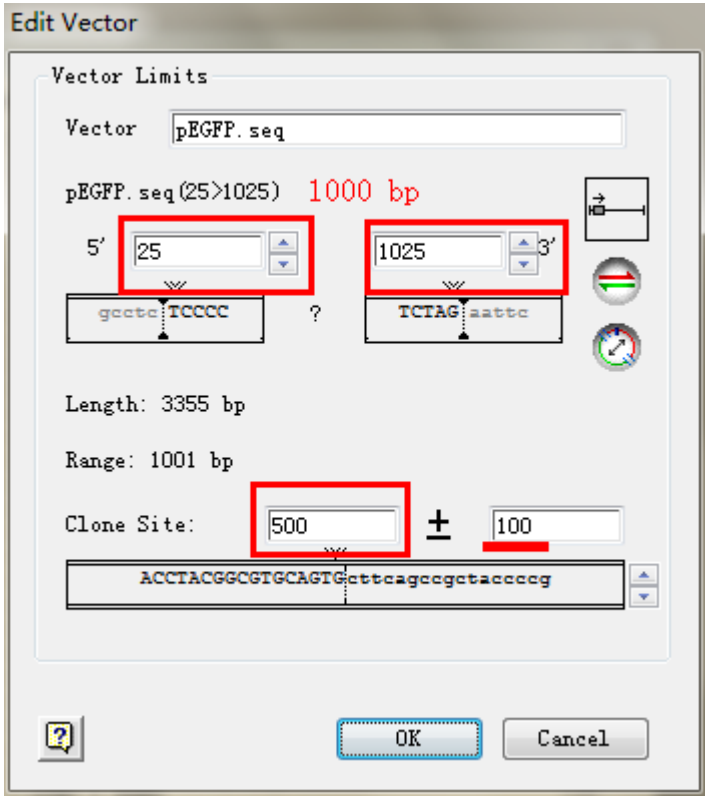
- 垂直的黑棒出现于修整和未修整的序列之间，根据需要拖动垂直黑棒，可以调整用于装配的序列末端。

自定义载体序列



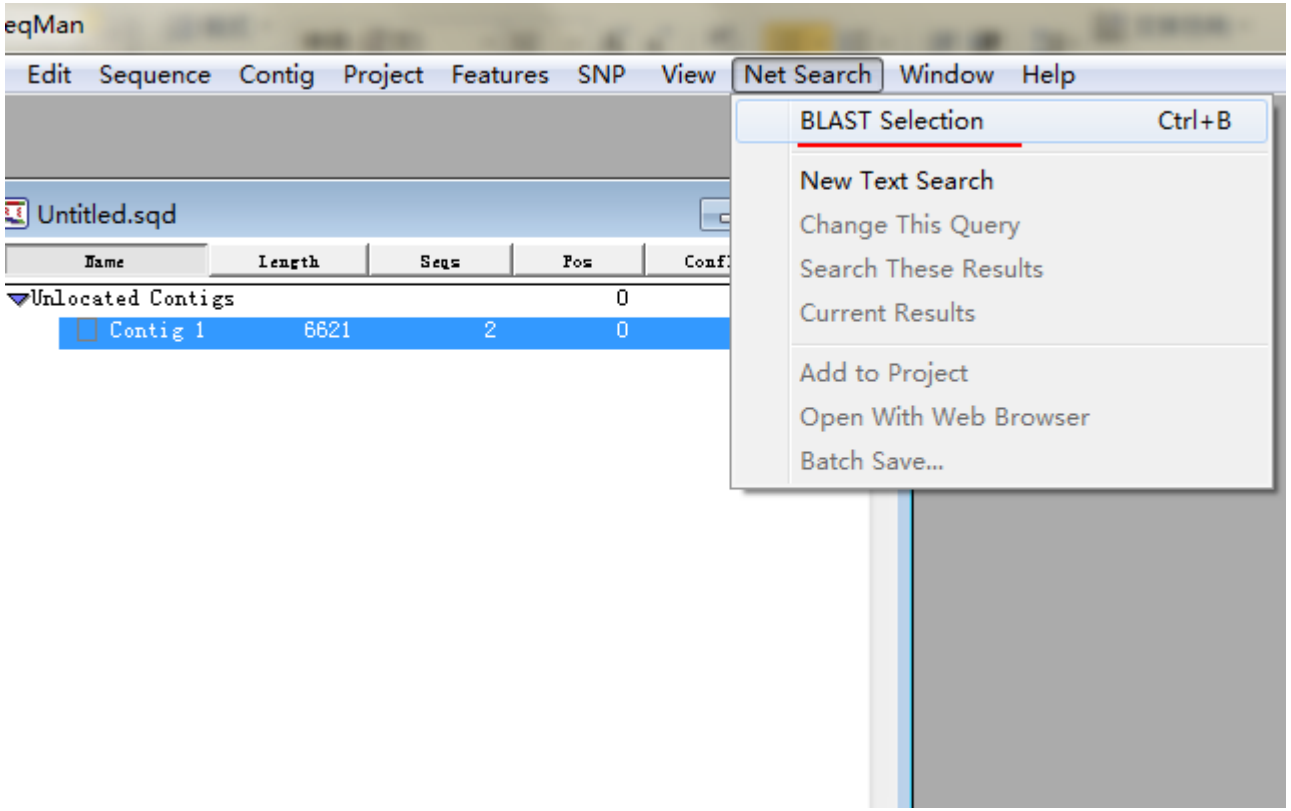
点击Vector Catalog选项自定义实验室常用的载体序列





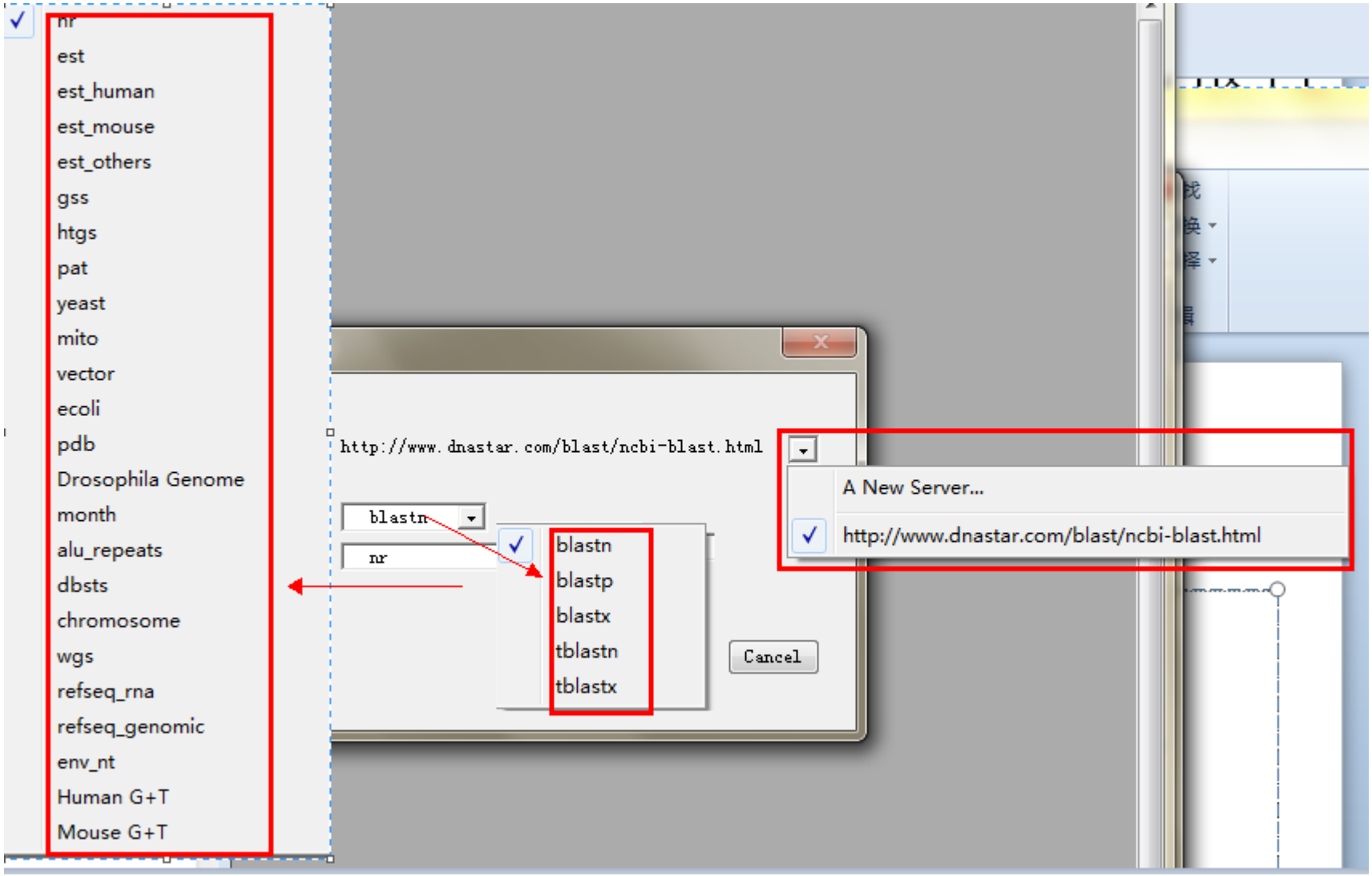
1、扫描载体插入位点上下500bp
2、多克隆位点位置及兼容位点

网络同源序列的下载和比对

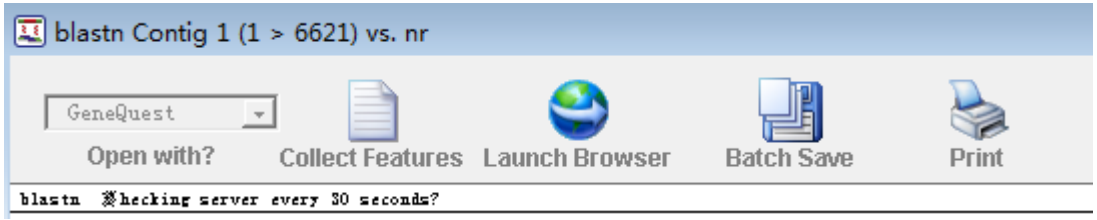


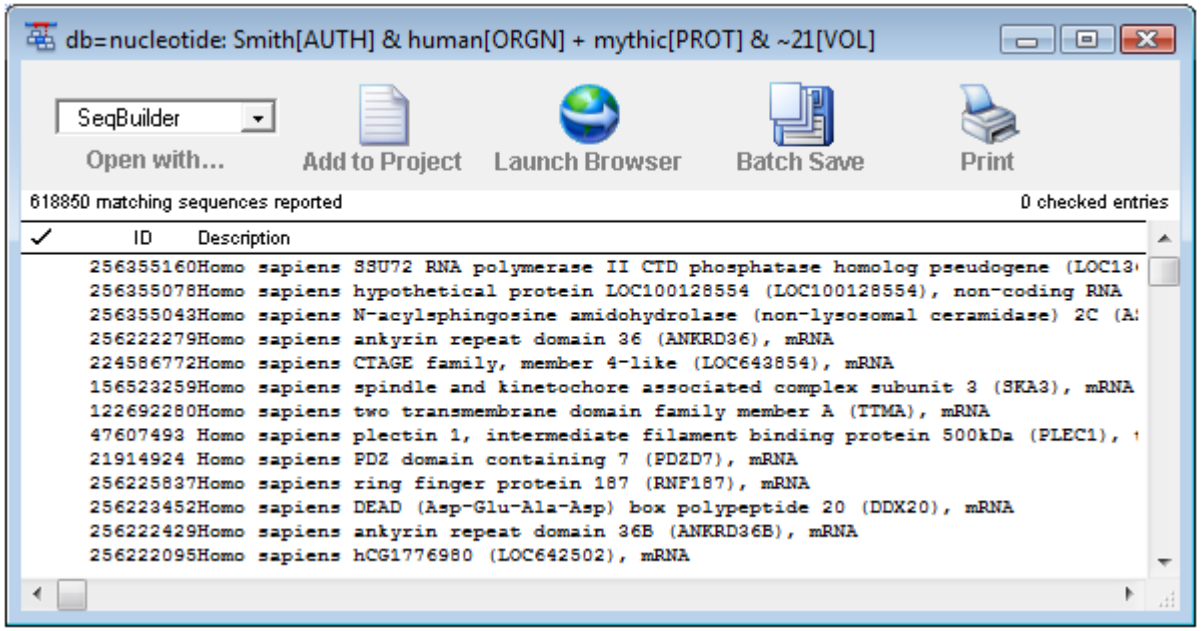
- 选定你的Contig 序列， 点击Net Search里的BLAST Selection

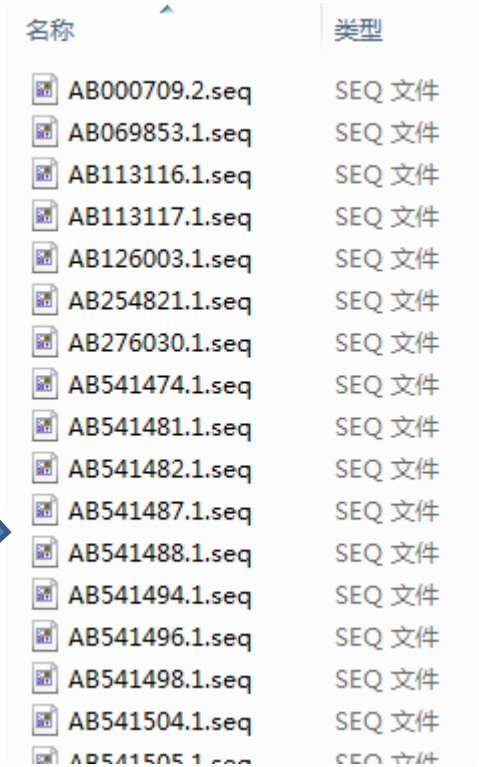
网络同源序列的下载和比对



网络同源序列的下载和比对







谢谢！