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, SegMan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

壁 SeqMan		
<u>File Edit Sequence Contig Project Net Search Wi</u>	ndow <u>H</u> elp	
,		^
T Unassembled Sequences		
Assemble Add Sequences? Set Ends? Tr	rim Ends? Options? Deqs	
Trim Ends:pending No Vector	✓ No Vector ✓	
Contaminate Scan:		
Optimize Order:pending Sequences: 0		
File Limits Vector	Vector Type	
	<u>^</u>	
	Ψ.	E
		^
		+
		🚽 E.

Step2:加入你要拼接的序列

The SeqMan
<u>File Edit S</u> equence <u>C</u> ontig <u>P</u> roject <u>N</u> et Search <u>W</u> indow <u>H</u> elp
Trim Ends:pending Vector Scan:pending Contaminate Scan: Contaminate Scan: Contaminate Scan: Contaminate Scan: Contaminate Scan:
Optimize Enter Sequences
File
名称 修改日期 类型 大小 1105225_163~15_SP6_C 1108149_16S~15_T7_GC
I105225_163~15_SP6_C02_1106007 I105225_163~15_SP6_C02_1106007
□ 1108149_16S~15_T7_G04_1111020.a □ 1108149 16S~15_T7_G04_1111020.a □ 1108149 16S~15_T7_G04_1111020.a
文件名 谜): ""1105225_163~15_SP6_C02_1106007. ab1" 打开 ②
XIT突空 ①: All Keadable Files
填加完后点击done

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

点击Assemble按钮

1

🗉 Una sembled Sequences				
Assemble Assemble in Groups?	dd Sequences? Set Inds?	Trim Inds? Hark Ref	Options?	
Issemble Issemble in Grongs? Id Trim Ends:pending Contaminate Scan: Optimize Order:pending Contaminate Scan: Optimize Order:pending Sequences: 8 File Limi 1210050041.1110.PEAQ=HT=seq_A10.seq 1210050038.1010.PEAQ=HT=seq_A07.ab1 1210050038.1010.PEAQ=HT=seq_A07.seq 1210050039.1102.PEAQ=HT=seq_A08.ab1 1210050039.1102.PEAQ=HT=seq_A08.seq 1210050040.1108.PEAQ=HT=seq_A09.ab1 1210050040.1108.PEAQ=HT=seq_A09.ab1 1210050040.1108.PEAQ=HT=seq_A09.seq 1210050040.1108.PEAQ=HT=seq_A10.ab1 1210050041.1110.PEAQ=HT=seq_A10.ab1	its Vec	Initial Initia In		Lict Split
				Ŧ

Step3:进行序列拼接

r							7	
I Alignment of Contig 1					[- 0 -	<	
💽 Position: 1						949 bp		
Reference Coordinates		10 	20 1	30 	40	50 		
	ATCGCGA	CCGGTATGI	CTATAAATA	TAAGAGACCO	CTCTTATA	GTAAGCAG		
	ATCGCGA	CCGGTATGI	CTATAAATA	TAAGAGACCO	CTCTTATA	GTAAGCAG	*	
	IATCGCGA	CCGGTATGI CCGGTATGI	СТАТАААТА ГСТАТАААТА	TAAGAGACC(TAAGAGACC	СТСТТАТА ТСТТАТАТ	GTAAGCAG		
I I I I I I I I I I I I I I I I I I I	ATCGCGA	CCGGTATGI	CTATAAATA	TAAGAGACC	CTCTTATA	GTAAGCAG		
	ATCGCGA	CCGGTATGI	CTATAAATA	TAAGAGACCO	TCTTATA TCTTATA	GTAAGCAG		
	ICGCGA	CCGGTATGI	CTATAAATA	TAAGAGACCO	CTCTTATA	GTAAGCAG		
	ICGCGA	CCGGTATGI	CTATAAATA	TAAGAGACCO	CTCTTATA	GTAAGCAG		
<u>N</u>								
≪≫ Unspecified Search ◀						۱.		
		▼ 121005	∪U4Ueq	_AU9.(1>9)	35)→		۸۸	۸ ۸
						Mark	^w W	XMMM-MMX
Alignment of cor	ntig1					ATCGCGACC	GGTA	TGTCTATAAATATAAGA(
	- <u>b</u> .	▼ 121005	50039eg	_AO8.(1>94	45)→ [۸۸	Δ
囱口甲息击 左二	1用					LANA	ſ₩ ∧/	Marama A
日二古石山山							<u>aata</u>	TCTCTATAAATATAACA
亟不户列的测户	11111111111111111111111111111111111111	▼ 121005	50038eg	_A07.(1>94	49)→	11COCOACC.	0011	
			-				M.	A
						/ 🗶 XXXX X/ 🔪	X	~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

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





 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

▼1108149_16S?	The Search
	 Literal Search in: Allow Ambiguity Consensus Conflict Conflict/SNP Split
 Conflict'	Find Previous Find Next Cancel

Step5: 修改拼接错误







 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



Step5: 修改拼接错误



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



Step5: 修改拼接错误



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



Step6:导出拼接的序列

• 可选择合适的格式,导出拼接好的序列

ent of	Con	tig 1]											
ence	Cor	itig Project iver search	Window Help										
Trace Majoi		Contig Info Alignment View Strategy View	Ctrl+I	0 44 	40 450 	L. A.							
15_SP 15 T7		Name Contig Position Contig Delete Contig Complement Contig					 野 Save Conse 保存在 Œ):	ensus as:	1.5 Tant	•	G Ø	P	×
_		Lock Contig Unlock Contig				~	名称	大小	类型	修改	如日期		-
		Reassemble Contig Align Contigs Force Join Contigs Extend Contig Ends					our	home				3	
		Suggest Conflict/SNP Spli Cancel Suggestions Split as Suggested Split at Insertion	ts				文件名 00): 保存类型 (T):	Lasergene DNA Lasergene DNA	File (*.seq) File (*.seq)			保存に取消	»
		Save Consensus Export Sequences	• [Single File Multiple File As Single Se	equence		Bases: O A	FastA file (* GenBank Flat GCG File (*.g Extended Fast ps	. ras) File (*.gbk) cg) A File (*.fas)				

- 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!



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。

丰动修改

SegMan Parameters Editing Loading Files Preassembling End Trimming All editing allowed Contaminant Screening Assembling Only gap editing allowed Consensus Calling Strategy Viewing & Cove No editing allowed Pair Specifier Conflict Split Colors SNP Discovery Primer Walking Editing & Color Vse consensus match color Vector Catalog Servers Match Mismatch Internet Use imported consensus color Imported Consensus $\overline{\mathbf{v}}$ ✓ Use other colors Negated Weight Trimmed Vse single letter protein codes 2 OK Cancel Restore

的背景。 ÉŤ. 甲 $Project \rightarrow Paramete$ rs→Editing Color 下面的对话 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



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.。

1dd Sequences?	Set Inds?	Trim Inds? Hark Re	ef Options?
	Janus	 InvJanus 	-
Limits	Vector	Vector	Туре
(1>823)	?Janus	?InvJanus	Trace 🔺
	?Janus	?InvJanus	Trace
	?Janus	?InvJanus	Trace
	?Janus	?InvJanus	Trace



• 单击 Scan All 按钮,将出现一个report 窗口。

Sample 10. abi	(33>667)	🗸 anus	?InvJanus	Trace
Sample 14. abi	(68>615)	🖌 Janus	?InvJanus	Trace
Sample 13. abi	(55>694)	🖌 🕽 anus	?InvJanus	Trace
Sample 9. abi	(58>686)	🖌 🕽 anus	?InvJanus	Trace
Sample 5. abi	(31>540)	🖌 🕽 anus	?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

> Unspecified Search

Irim Report					
SCAN FOR VECTOR					
Janus					
InvJanus					
TRIM ENDS					
MEDIUM TRIM PARAMETERS					
Trace: Thresho	ld = 12				
Oual: Thresho	ld = 12				
Non-Trace: Window	/Size = 50				
Maximum	. Ns = 2				
TON INC.	AVERAGE	AMOUNT	TRIMMED	PRE-TRIM	
NAME	OUALITY	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 Avera	.ge 29	279	522	801	

Summary Statistics Report Trim Report Parameters... Vector Catalog... Contaminant Seqs... Repetitive Seqs...

Project Features SNP

report窗口

View

Net Se

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

Project Features SNP View Net Se		Edit Vector
Project Features SNP View Net Se Summary Statistics Report Trim Report Parameters Vector Catalog	Vector Catalog	Vector Limits Vector pEGFP. seq pEGFP. seq(25>1025) 1000 bp 5' 25 gcctc TCCCC ? TCTAG satte
Contaminant Seqs Repetitive Seqs 点击Vector Catalog选项 自定义实验室常用的载 体序列	□ pGEM-42 □ pGEM-32f+ □ pGEM-3Z □ pCRII TOPO □ pCR4 TOPO □ pCAT3-Basic Vector □ pCAT3-Control □ pBR322 □ pBluescriptII SK+ □ pBluescriptII SK- □ pBluescriptII KS+ □ pBluescriptII KS+ □ pBluescriptII KS- □ pBC SK+	Length: 3355 bp Range: 1001 bp Clone Site: 500 <u>+</u> 100 ACCTACGGCGTGCAGTGcttcagccgctaccccg
	pBC SK- pBC KS+ pCI-neo Mammalian Expression pCR2.1 TOPO * pEAQ-HT (10003bp).seq New Edit Remove	 OK Cancel 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

Edit Sequence	Contig Pro	ect Features	s SNP	View	Net Search Window He	lp
					BLAST Selection	Ctrl+B
Untitled.sqd					New Text Search Change This Query	
Jame	Length	Segs	Pos	Conf:	Search These Results	
Contig 1	s 662 <u>1</u>	2	0		Current Results	
					Add to Project	
					Open With Web Brow	ser
					Batch Save	
				-		

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

💷 blastn Contig 1 (1 > 6621) vs. nr						
GeneQuest Image: Collect Features Image: C						
blastn Ahecking server every 30 seconds?						
db=nucleotide: Smith[AUTH] & human[ORGN] + mythic[PROT] & ~21[VOL]	• •					
SeqBuilder 🗉 🤮 📲 🦕						
Open with Add to Project Launch Browser Batch Save Print						
618850 matching sequences reported D chec	ked entries					
✓ ID Description						
256355160Homo sapiens SSU72 RNA polymerase II CTD phosphatase homolog pseudogene (LOC13/ 256355078Homo sapiens hypothetical protein LOC100128554 (LOC100128554), non-coding RNA 256355043Homo sapiens N-acylsphingosine amidohydrolase (non-lysosomal ceramidase) 2C (A/ 256252279Homo sapiens ankyrin repeat domain 36 (ANKRD36), mRNA 224586772Homo sapiens CTAGE family, member 4-like (LOC643854), mRNA 156523259Homo sapiens spindle and kinetochore associated complex subunit 3 (SKA3), mRNA 122692280Homo sapiens two transmembrane domain family member A (TTMA), mRNA 47607493 Homo sapiens plectin 1, intermediate filament binding protein 500kDa (PLEC1), 1 21914924 Homo sapiens ring finger protein 187 (RNF187), mRNA 256223452Homo sapiens DEAD (Asp-Glu-Ala-Asp) box polypeptide 20 (DDX20), mRNA 256222429Homo sapiens ankyrin repeat domain 36B (ANKRD36B), mRNA						
	► a					

	~	
	名称	类型
	📓 AB000709.2.seq	SEQ 文件
	📧 AB069853.1.seq	SEQ 文件
	📧 AB113116.1.seq	SEQ 文件
	📧 AB113117.1.seq	SEQ 文件
	📧 AB126003.1.seq	SEQ 文件
	📧 AB254821.1.seq	SEQ 文件
	📧 AB276030.1.seq	SEQ 文件
	📧 AB541474.1.seq	SEQ 文件
	📧 AB541481.1.seq	SEQ 文件
	📧 AB541482.1.seq	SEQ 文件
Z	📧 AB541487.1.seq	SEQ 文件
	📧 AB541488.1.seq	SEQ 文件
	📧 AB541494.1.seq	SEQ 文件
	📧 AB541496.1.seq	SEQ 文件
	📧 AB541498.1.seq	SEQ 文件
	📧 AB541504.1.seq	SEQ 文件
	AP5/1505 1 cog	CEO 77/4

谢谢!