An Introduction to Finishing Using the Consed Platform (Part II)
Developed by Andrew Nylander from notes by Chad Tomlinson.
Revised by William Barshop and Wilson Leung.

Files for this Tutorial:
All files for this tutorial are contained within XBAA7G24.tgz

Note: For this part of the tutorial we will be using the XBAA7G24 project. Open a new xterm and navigate to the 'edit_dir' of the XBAA7G24 directory. Type 'consed &' to launch the program and open XBAA7G24.fasta.screen.ace.1 to view the initial assembly.

Resolving Gaps

From the initial Assembly View in the ace.1 file we can see that this fosmid is in four separate contigs, and so a large part of the work for finishing this project will focus on closing the gaps to assemble a single contig (Figure 25). Of course, we will also need to resolve high quality discrepancies and weak regions before we are finished.

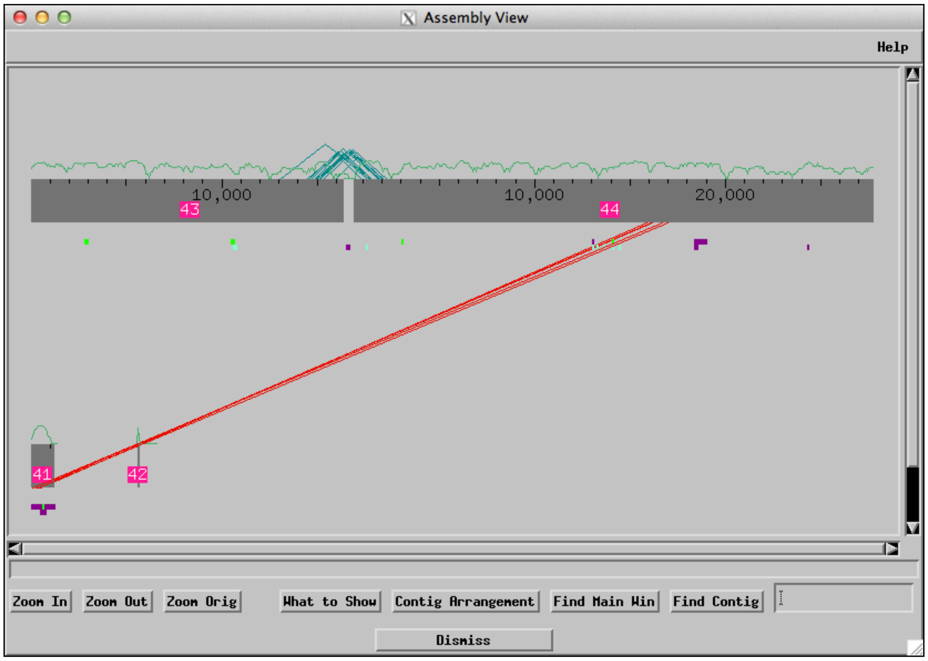


Figure 25. Initial assembly

The green triangles between contigs 43 and 44 indicate consistent forward / reverse pairs and so it is probably best to start by trying to close this gap first and save the misassembly between 41 and 44 for later (Figure 26). First we will run *crossmatch* to visualize the repeat structure in the two contigs. Knowledge of the repeat structure will be important when we start designing oligos since we need our primers to anneal to a unique region. We can also use the green ‘high quality match elsewhere’ tags and the blue ‘repeat’ tags in the Aligned Reads Window to identify repetitive regions.

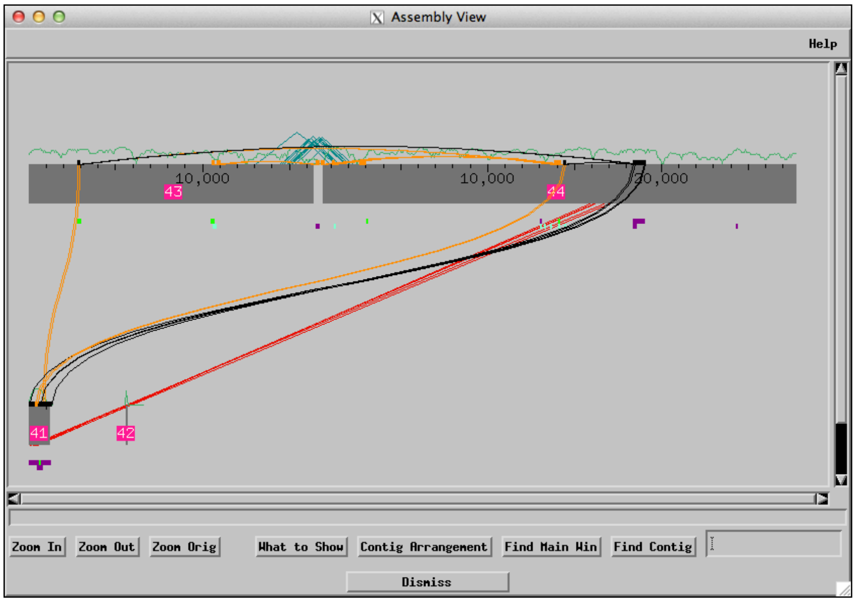
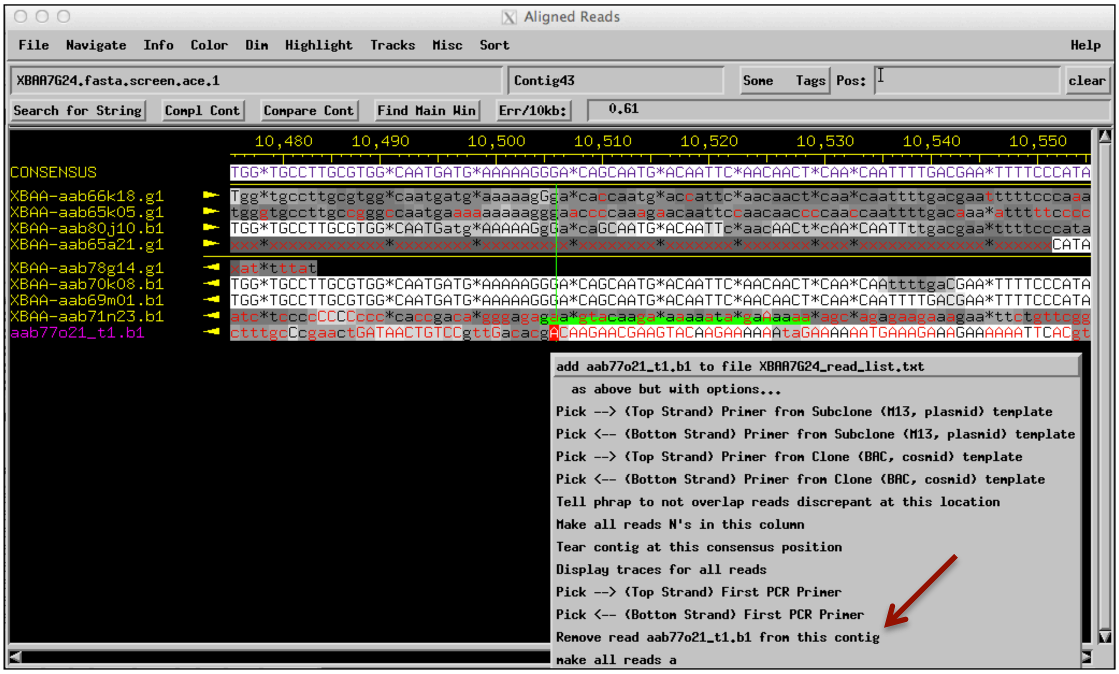


Figure 26. Assembly View showing repeat structures

To ensure that there are no misassemblies around the gap that may have prevented proper joining of the contigs, and to look for a possible overlap it is important to select a string of high quality bases near to the gap and do a ‘Search for String’ using those bases as described previously in this tutorial. If there are multiple matches, a misassembly is possible. If a sequence at the right end of contig 43 matches the sequence at the left end of contig 44, it may be possible to manually join these regions together (i.e. a ‘forced join’). If there are no matches (as in this case) then we can be fairly certain that the gap cannot be forced joined and that additional sequencing data will be required.

In order to close this gap we will need to order reactions for more data and so we will create oligos for this region as described previously in this tutorial. Often it is helpful to use spanning subclones that are shown by the purple lines in the assembly view. You can call oligos for forward and reverse reactions at the closest unique sequence. For this project the primers have already been designed and the reaction results are available in the exercise package. We can incorporate the new data into the assembly using the ‘Add New Reads’ function. Select the gaprxns.fof file and select the options to put unaligned reads into its own contig and to recalculate the consensus quality scores based on the additional reads. When the new reads have been added a navigation window will appear that show us where the reads were added into the assembly.

To see all the available data from the new reads, change the Dim option to “Dim Nothing”. If we look at the position of the reads in the Aligned Reads window it is obvious that the reads were inserted in the wrong place because there are many high quality discrepancies between the new reads and the consensus sequence. To correct this, we will pull out the misaligned reads and place them in a new contig. To put the read into its own contig, right-click on the read and select ‘Remove read aab77o21_t1.b1 from this contig’ (Figure 27). In the “Remove Reads” window, select the “Just Put Each Read Into Its Own Contig” option (underneath the Contigs list) and then click on the “Do It” button. Do this for both reads and then save the assembly.

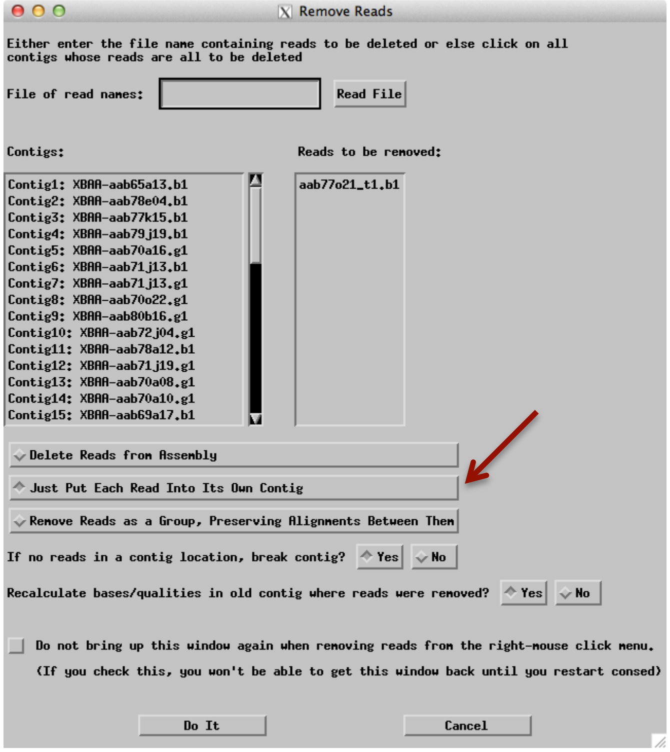

Aligned Reads
File
Navigate
Info
Color
Dim
Highlight
Tracks
Misc
Sort
Help
Some
Tags
Pos:
clear
Search for String
Compl Cont
Compare Cont
Find Main Win
Err/10kb:
add aab77o21_t1.b1 to file XBAA7G24_read_list.txt
as above but with options...
Pick --> (Top Strand) Primer from Subclone (M13, plasmid) template
Pick <-- (Bottom Strand) Primer from Subclone (M13, plasmid) template
Pick --> (Top Strand) Primer from Clone (BAC, cosmid) template
Pick <-- (Bottom Strand) Primer from Clone (BAC, cosmid) template
Tell phrap to not overlap reads discrepant at this location
Make all reads N's in this column
Tear contig at this consensus position
Display traces for all reads
Pick --> (Top Strand) First PCR Primer
Pick <-- (Bottom Strand) First PCR Primer
Remove read aab77o21_t1.b1 from this contig
make all reads a

XBAA7G24.fasta.screen.ace.1

Contig43

Err/10kb: 0.61

```
                   10,480    10,490     10,500    10,510    10,520     10,530    10,540    10,550
CONSENSUS          TGG*TGCCTTGCGTGG*CAATGATG*AAAAAGGGA*CAGCAATG*ACAATTC*AACAACT*CAA*CAATTTTGACGAA*TTTTCCCATA
XBAA-aab66k18.g1   Tgg*tgccttgcgtgg*caatgatg*aaaaagGga*caccaatg*accattc*aacaact*caa*caattttgacgaattttcccaaa
XBAA-aab65k05.g1   tgggtgccttgccgggccaatgaaaaaaaaagggaaccccaaagaacaattccaacaaccccaaccaattttgacaaa*atttttcccc
XBAA-aab80j10.b1   TGG*TGCCTTGCGTGG*CAATGatg*AAAAAGgGa*caGCAATG*ACAATTc*aacAACt*cAA*CAATTttgacgaa*ttttcccata
XBAA-aab65a21.g1   xxx*xxxxxxxxxxxx*xxxxxxxx*xxxxxxxxx*xxxxxxxx*xxxxxxx*xxxxxxx*xxx*xxxxxxxxxxxxx*xxxxxxCATA
XBAA-aab78g14.g1   xat*tttat
XBAA-aab70k08.b1   TGG*TGCCTTGCGTGG*CAATGATG*AAAAAGGGA*CAGCAATG*ACAATTC*AACAACT*CAA*CAAttttgaCGAA*TTTTCCCATA
XBAA-aab69m01.b1   TGG*TGCCTTGCGTGG*CAATGATG*AAAAAGGGA*CAGCAATG*ACAATTC*AACAACT*CAA*CAATTTTGACGAA*TTTTCCCATA
XBAA-aab71n23.b1   atc*tccccCCCCccc*caccgaca*gggagagaa*gtacaaga*aaaaata*gaAaaaa*agc*agagaagaaagaa*ttctgttcgg
aab77o21_t1.b1     ctttgcCcgaactGATAACTGTCCgttGacacgACAAGAACGAAGTACAAGAAAAAAtaGAAAAAATGAAAGAAGAAAAAATTCACgt
```



Either enter the file name containing reads to be deleted or else click on all contigs whose reads are all to be deleted

File of read names:

Contigs:

Contig1: XBAA-aab65a13.b1
Contig2: XBAA-aab78e04.b1
Contig3: XBAA-aab77k15.b1
Contig4: XBAA-aab79j19.b1
Contig5: XBAA-aab70a16.g1
Contig6: XBAA-aab71j13.b1
Contig7: XBAA-aab71j13.g1
Contig8: XBAA-aab70o22.g1
Contig9: XBAA-aab80b16.g1
Contig10: XBAA-aab72j04.g1
Contig11: XBAA-aab78a12.b1
Contig12: XBAA-aab71j19.g1
Contig13: XBAA-aab70a08.g1
Contig14: XBAA-aab70a10.g1
Contig15: XBAA-aab69a17.b1

Reads to be removed:

aab77o21_t1.b1

○ Delete Reads from Assembly
● Just Put Each Read Into Its Own Contig
○ Remove Reads as a Group, Preserving Alignments Between Them

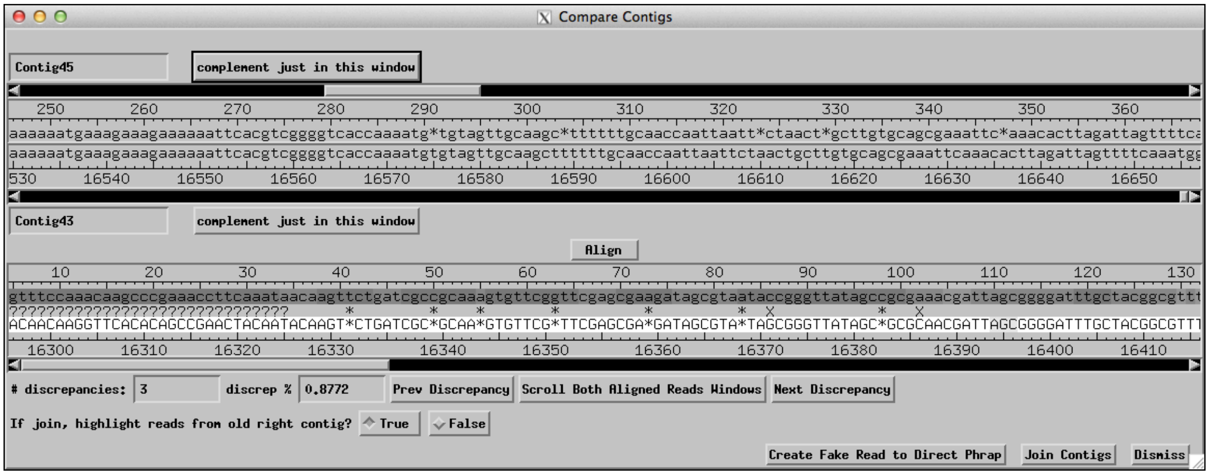
If no reads in a contig location, break contig? ● Yes ○ No

Recalculate bases/qualities in old contig where reads were removed? ● Yes ○ No

□ Do not bring up this window again when removing reads from the right-mouse click menu.
(If you check this, you won't be able to get this window back until you restart consed)

Figure 27. Pulling out a misaligned read.

Go back to the Consed Main Window, in the ‘Contig List’ you should see that contigs 45 and 46 correspond to the reads that were just pulled out. In the Assembly View window right click on the right end of contig 43 and open the Aligned Reads window. Highlight about 20 high quality bases near the start of the gap before the quality of the sequence drops. Use these bases to perform a search for string. You should return matches only to Contig43 and one of your new single read contigs. Go to the Aligned Reads window for both matches and click on ‘Compare Cont’ in both of the windows. A ‘Compare Contig’ window will appear with both sequences lined up in the top box. Click on ‘Align’ to make an alignment between the two sequences in the lower box and see how well they match (Figure 28). At the beginning of the alignment you will see numerous mismatches, but since the quality is low at the start of one of the sequences we will not concern ourselves with these mismatches. What is important is to see if there are any high quality discrepancies of the alignment. Looking through the alignment there are no high quality mismatches, and so we can now join the contigs together by clicking the ‘Join Contigs’ button.


Compare Contigs
complement just in this window
complement just in this window
Align
Prev Discrepancy
Scroll Both Aligned Reads Windows
Next Discrepancy
◇ True
◇ False
Create Fake Read to Direct Phrap
Join Contigs
Dismiss

Contig45

```
  250       260       270       280       290       300       310       320       330       340       350       360
aaaaaatgaaagaaagaaaaaattcacgtcggggtcaccaaaatg*tgtagttgcaagc*ttttttgcaaccaattaatt*ctaact*gcttgtgcagcgaaattc*aaacacttagattagtttca
aaaaaatgaaagaaagaaaaaattcacgtcggggtcaccaaaatgtgtagttgcaagcttttttgcaaccaattaattctaactgcttgtgcagcgaaattcaaacacttagattagttttcaaatgg
530      16540     16550     16560     16570     16580     16590     16600     16610     16620     16630     16640     16650
```

Contig43

```
    10        20        30        40        50        60        70        80        90       100       110       120       130
gtttccaaacaagcccgaaaccttcaaataacaagttctgatcgccgcaaagtgttcggttcgagcgaagatagcgtaataccgggttatagccgcgaaacgattagcggggatttgctacggcgttt
??????????????????????????????      *        *    *       *         *         *  X           *   X
ACAACAAGGTTCACACAGCCGAACTACAATACAAGT*CTGATCGC*GCAA*GTGTTCG*TTCGAGCGA*GATAGCGTA*TAGCGGGTTATAGC*GCGCAACGATTAGCGGGGATTTGCTACGGCGTTT
  16300     16310     16320     16330     16340      16350      16360      16370     16380      16390     16400     16410
```

discrepancies: 3 discrep % 0.8772

If join, highlight reads from old right contig?

Figure 28. Compare Contigs window.

Repeat the same procedure to incorporate contig46 into the main assembly as well. Then join the end of the new contig48 with the beginning of contig44 to resolve the gap. Open the ace.gap3 file and compare it to your final assembly. (Depending on the exact steps taken to close the gap your contig numbers may differ from the screenshot below.)

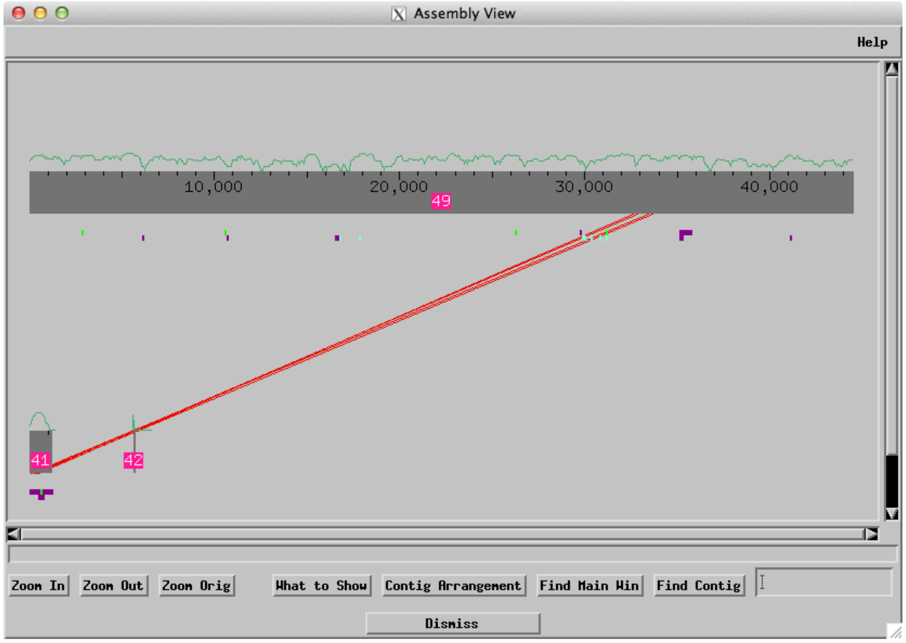


Figure 29. Assembly View from ace.gap3.

Resolving the Misassembly:

To make sure that the contig numbers are consistent for the next part of this tutorial, we will start the project again with a new ace file 'ace.mis1.' Quit Consed and then launch Consed again using the xterm. Open the file with the suffix "ace.mis1" and do not apply changes from the work file. Next open the Assembly View (Figure 30).

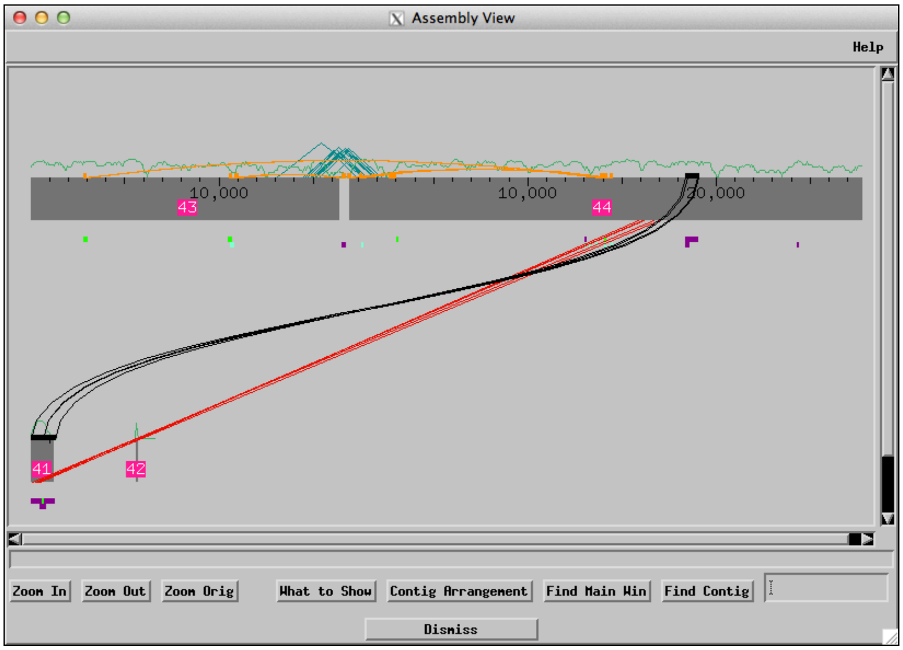
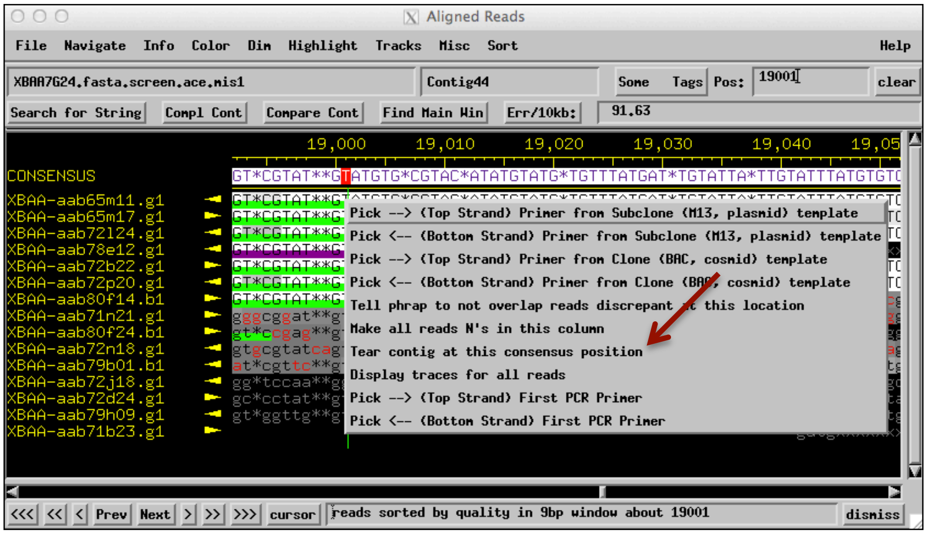


Figure 30. ace.mis1 Assembly View.

There is clearly a problem between contigs 41 and 44. The red lines represent inconsistent forward/reverse read pairs. The mate pairs are inconsistent in this case because they are in separate contigs that are too far apart compared to the expected insert size. There is also an inverted repeat match between contigs 41 and 44. Based on these observations it appears that phrap has collapsed multiple copies of a repeat in contig 44. This means we need to tear contig 44 apart in order to insert the reads from contig 41 into the repeat region.

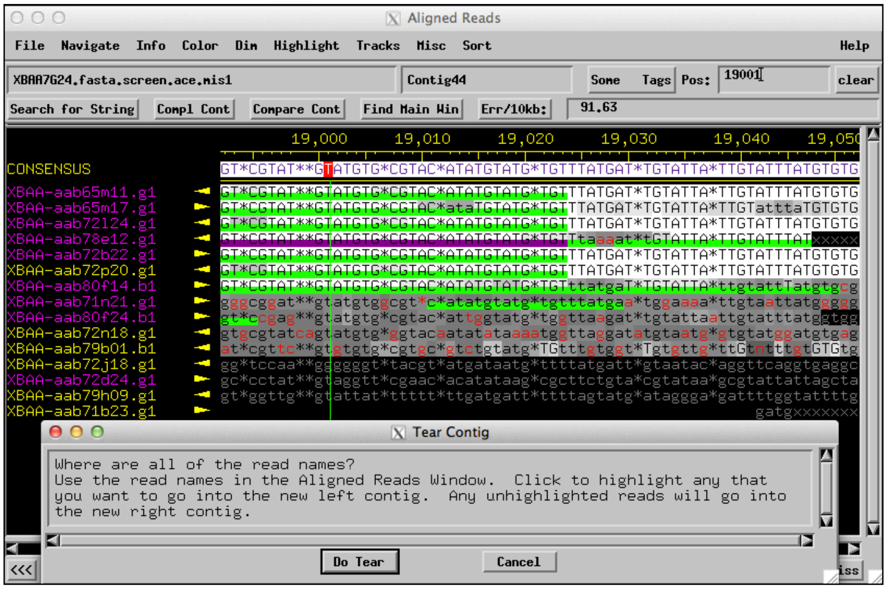
To resolve the misassembly we first tear contig 44 near the inverted repeat at about 19,000. To do the tear, go to the position 19,001 in the Aligned Reads window and right click on any of the reads. Select 'Tear contig at this consensus position' from the drop down menu (Figure 31).


Aligned Reads
File Navigate Info Color Dim Highlight Tracks Misc Sort Help
XBAA7G24.fasta.screen.ace.mis1
Contig44
Some Tags Pos: 19001 clear
Search for String Compl Cont Compare Cont Find Main Win Err/10kb: 91.63
19,000 19,010 19,020 19,030 19,040 19,05
CONSENSUS
XBAA-aab65m11.g1
XBAA-aab65m17.g1
XBAA-aab72124.g1
XBAA-aab78e12.g1
XBAA-aab72b22.g1
XBAA-aab72p20.g1
XBAA-aab80f14.b1
XBAA-aab71n21.g1
XBAA-aab80f24.b1
XBAA-aab72n18.g1
XBAA-aab79b01.b1
XBAA-aab72j18.g1
XBAA-aab72d24.g1
XBAA-aab79h09.g1
XBAA-aab71b23.g1
Pick --> (Top Strand) Primer from Subclone (M13, plasmid) template
Pick <-- (Bottom Strand) Primer from Subclone (M13, plasmid) template
Pick --> (Top Strand) Primer from Clone (BAC, cosmid) template
Pick <-- (Bottom Strand) Primer from Clone (BAC, cosmid) template
Tell phrap to not overlap reads discrepant at this location
Make all reads N's in this column
Tear contig at this consensus position
Display traces for all reads
Pick --> (Top Strand) First PCR Primer
Pick <-- (Bottom Strand) First PCR Primer
<<< << < Prev Next > >> >>> cursor
dismiss

reads sorted by quality in 9bp window about 19001

Figure 31. Tearing the contig.

In the new window that appears, click on 'Do Tear' (Figure 32). You will then be prompted to save your assembly.



Where are all of the read names?
Use the read names in the Aligned Reads Window. Click to highlight any that you want to go into the new left contig. Any unhighlighted reads will go into the new right contig.

Figure 32. Tearing the contig continued.

Open Assembly View again and we see that contig 44 has been split into two contigs (45 and 46), with contig 41c placed between them (Figure 33).

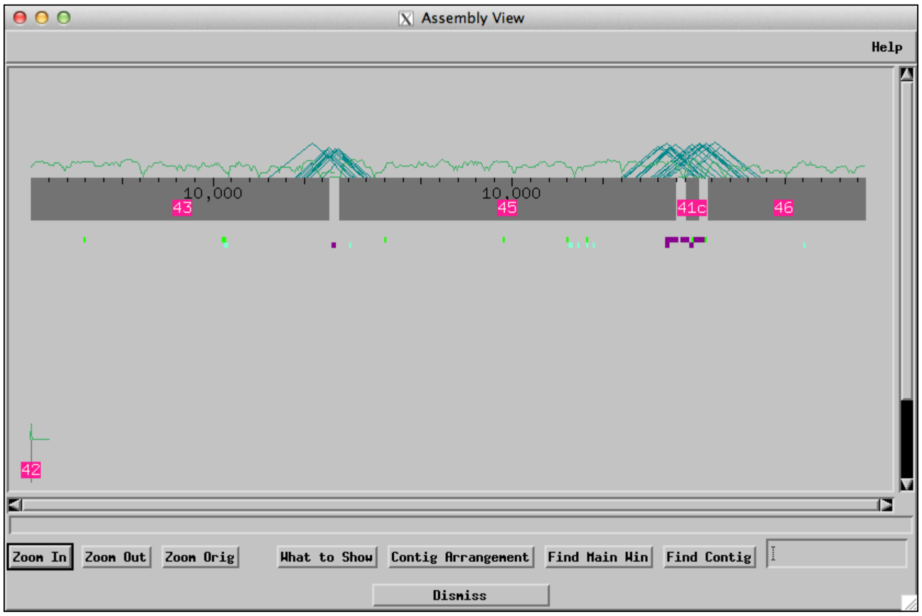


Figure 33. Assembly View after tear.

The ‘c’ next to contig 41 means that Assembly View has to reverse complement this contig for its mate pairs to be consistent with the other member of the mate pair in the other contigs. To correct this go to the Aligned Reads window for contig 41 and click on the ‘Compl Cont’ option. Open Assembly View again and all of the contigs are now oriented in the same direction (Figure 34).

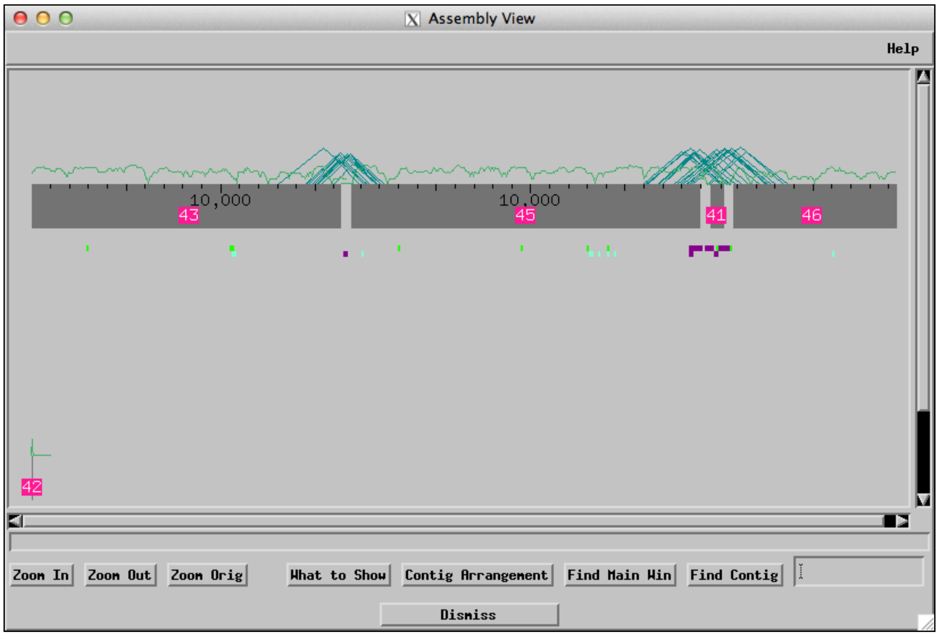
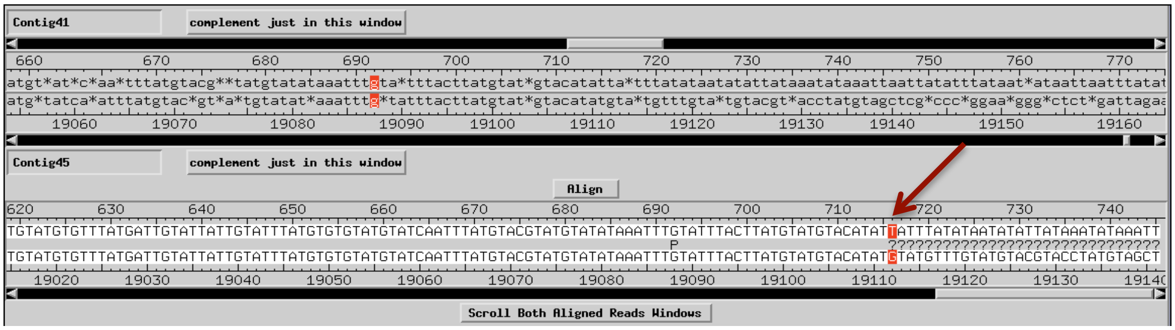


Figure 34. All the major contigs are in the same orientation

Now we need to join contigs 41, 45, and 46 together. We will start with joining 41 and 45. The method is the same as described earlier in resolving gaps. This time, however, when we look at the alignment between the two contigs we can see that there are high quality mismatches on the far right between the two sequences (Figure 35). Click on the first mismatch and then hit the 'Scroll both aligned reads windows'.



```
atgt*at*c*aa*tttatgtacg**tatgtatataaatttgta*tttacttatgtat*gtacatatta*tttatataatatattataaatataaattaattatatttataat*ataattaatttatat
atg*tatca*atttatgtac*gt*a*tgtatat*aaatttg*tatttacttatgtat*gtacatatgta*tgtttgta*tgtacgt*acctatgtagctcg*ccc*ggaa*ggg*ctct*gattaga
```

```
TGTATGTGTTTATGATTGTATTATTGTATTTATGTGTGTATGTATCAATTTATGTACGTATGTATATAAATTTGTATTTACTTATGTATGTACATATTATTTATATAATATATTATAAATATAAATT
                                                                                          P                         ????????????????????????????????
TGTATGTGTTTATGATTGTATTATTGTATTTATGTGTGTATGTATCAATTTATGTACGTATGTATATAAATTTGTATTTACTTATGTATGTACATATGTATGTTTGTATGTACGTACCTATGTAGCT
```

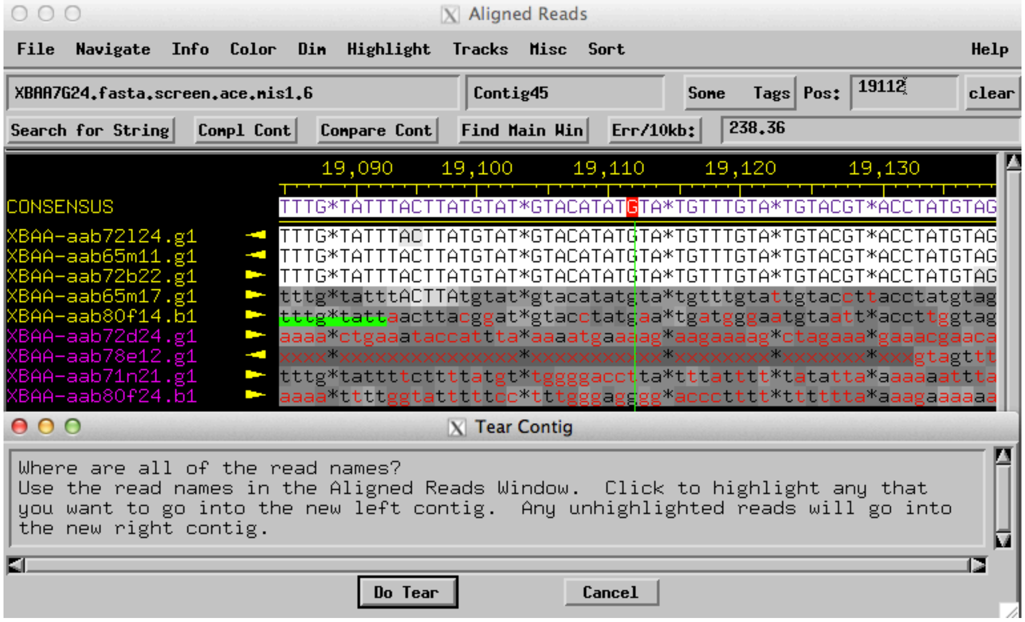
Figure 35. Compare Contigs window.

The Aligned Reads windows will now move to the position of the base difference. In contig 45 it appears that the reads containing the mismatching bases are misplaced in the assembly and so we will pull them out in order to make the join between 41 and 45.

To remove the misplaced reads, first click on the 'Highlight' option at the top of the Aligned Reads window for contig 45 and select 'Unhighlight All Reads in All Contigs.' Now we need to determine how to split this contig. We can see from the alignment in figure 35 that we have sequence that is shared between both contigs followed by two different sequences that are unique. There is a misassembly here and we can see from figure 35 that the sequences diverge at 19,112. Consequently, to resolve this misassembly, we must make a tear at that position.

When we tear the contig at position 19,112, we will use the discrepant bases starting from base 19,112 to partition the reads into two groups (To see these base differences more clearly, we will change the Dim options to "Dim nothing".) Looking at the discrepancies in this region, we find that one group of reads has very AT rich sequence while the other group has GC rich sequence. Look closely downstream of 19,112 to decide which reads belong together when we make the tear.

Now we can make a tear at the G in position 19112 of the contig by right clicking on the consensus and select 'tear contig at this consensus position.' Because the discrepancy is caused by the GC rich sequence, reads that contains a G at this consensus position (19112) should be in the new right contig. This means we should highlight (in purple) all the reads that do not contain a G at this consensus position so that they will go to the new left contig. Note that while the read XBAA-aab80f24.b1 has a G in the consensus position 19112, this G and the bases surrounding this region are very low quality. When we scroll to the left, we find that there are large number of additional discrepancies between this read and the consensus sequence. Consequently we will also highlight this read so that it will go to the new left contig. Hit the "Do Tear" button (Figure 36) and save the assembly.



Aligned Reads

XBAA7G24.fasta.screen.ace.mis1.6 | Contig45 | Some Tags | Pos: 19112 | clear

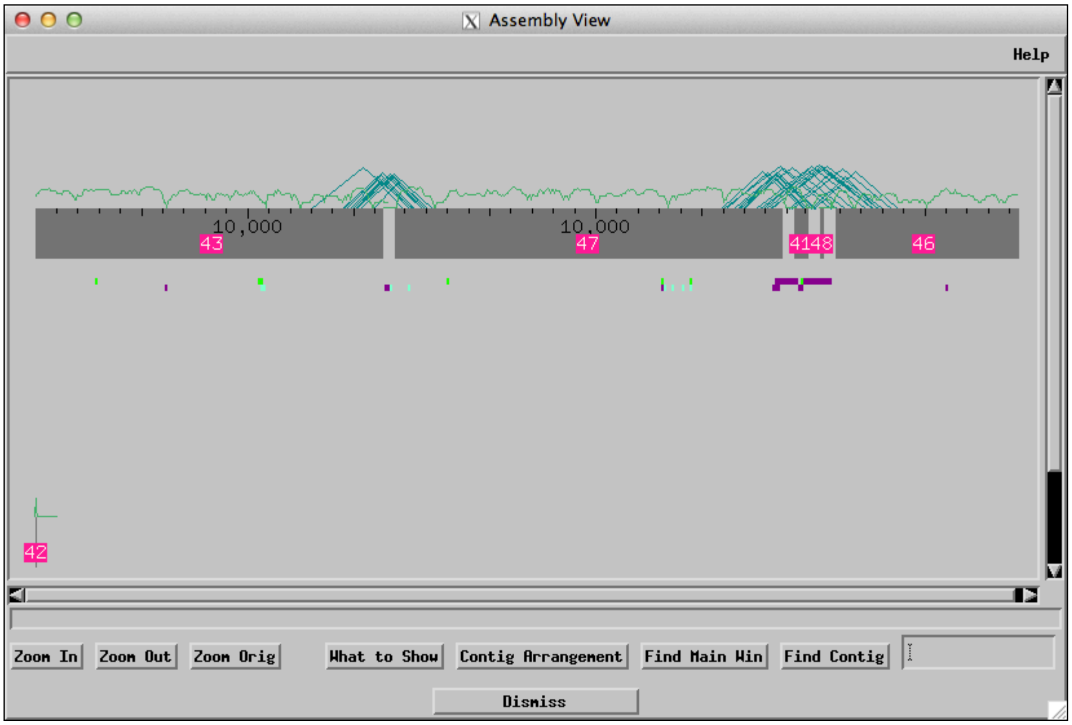
Search for String | Compl Cont | Compare Cont | Find Main Win | Err/10kb: 238.36

```
                      19,090    19,100     19,110     19,120     19,130
CONSENSUS             TTTG*TATTTACTTATGTAT*GTACATATGTA*TGTTTGTA*TGTACGT*ACCTATGTAG
XBAA-aab72124.g1   ◄  TTTG*TATTTACTTATGTAT*GTACATATGTA*TGTTTGTA*TGTACGT*ACCTATGTAG
XBAA-aab65m11.g1   ◄  TTTG*TATTTACTTATGTAT*GTACATATGTA*TGTTTGTA*TGTACGT*ACCTATGTAG
XBAA-aab72b22.g1   ►  TTTG*TATTTACTTATGTAT*GTACATATGTA*TGTTTGTA*TGTACGT*ACCTATGTAG
XBAA-aab65m17.g1   ►  tttg*tatttACTTAtgtat*gtacatatgta*tgtttgtattgtaccttacctatgtag
XBAA-aab80f14.b1   ►  tttg*tattaacttacggat*gtacctatgaa*tgatgggaatgtaatt*accttggtag
XBAA-aab72d24.g1   ►  aaaa*ctgaaataccattta*aaaatgaaaag*aagaaaag*ctagaaa*gaaacgaaca
XBAA-aab78e12.g1   ◄  xxxx*xxxxxxxxxxxxxxx*xxxxxxxxxxx*xxxxxxxx*xxxxxxx*xxxgtagttt
XBAA-aab71n21.g1   ►  tttg*tattttcttttatgt*tggggacctta*tttatttt*tatatta*aaaaaattta
XBAA-aab80f24.b1   ►  aaaa*tttggtattttttcc*tttgggagggg*accctttt*tttttta*aaagaaaaaa
```

Tear Contig

```
Where are all of the read names?
Use the read names in the Aligned Reads Window.  Click to highlight any that
you want to go into the new left contig.  Any unhighlighted reads will go into
the new right contig.
```

Figure 36. Partition reads into two groups using the discrepant base at 19,112.



Assembly View

Figure 37. Assembly View after tear

After the tear, contig 45 splits into contigs 47 and 48. The Assembly View shows the relative order and orientation of the contigs. Now we can make a join between the left end of contig 41 and the right end of contig 47 (Figure 37).

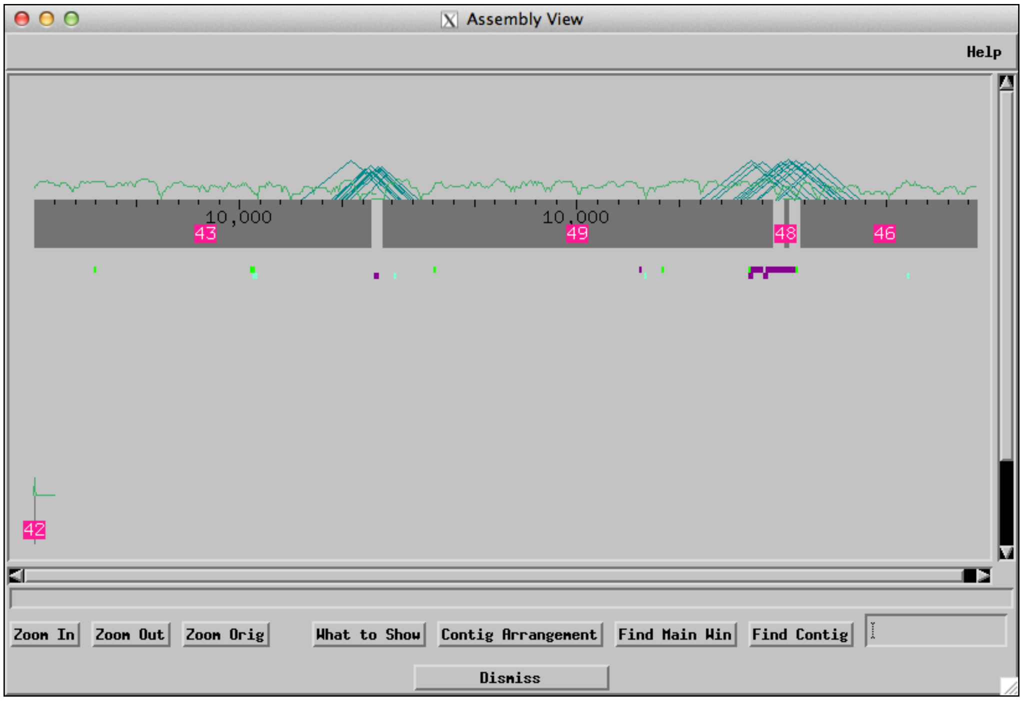


Figure 38. Assembly View after joining 41 to 47

Next, we will join the beginning of contig 48 to the end of contig 49 (Figure 38).

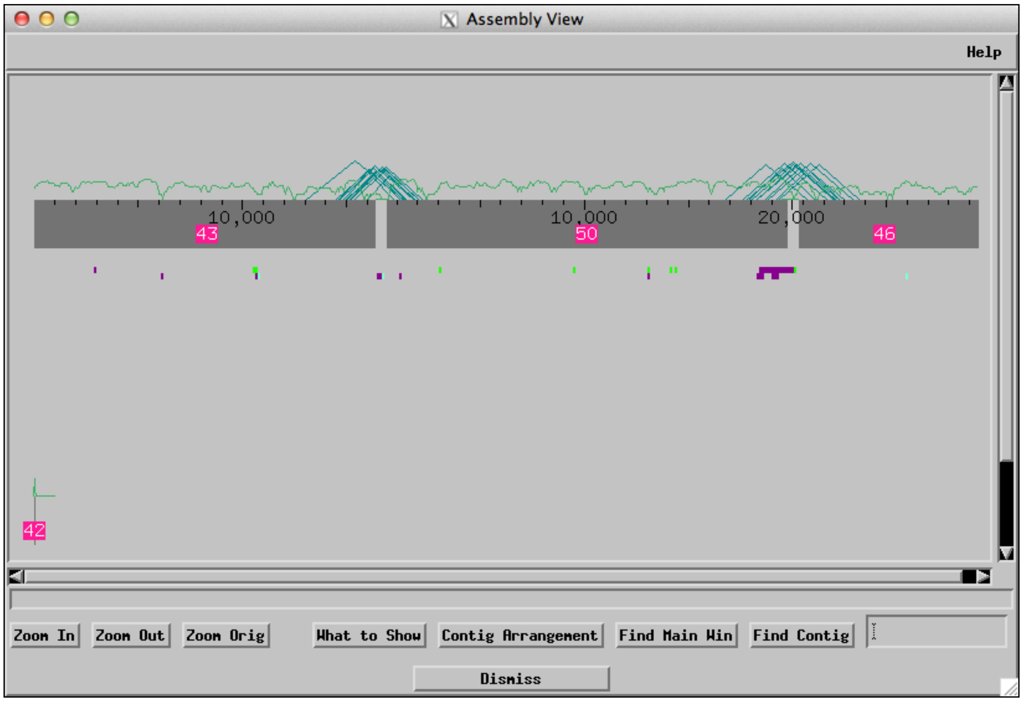


Figure 39. Assembly view after joining 48 and 49

Now, we join 50 and 46 together (Figure 39). The gap between Contigs 43 and 51 can be resolved using the additional data obtained earlier in this tutorial in the resolving gaps section (Figures 40, 41).

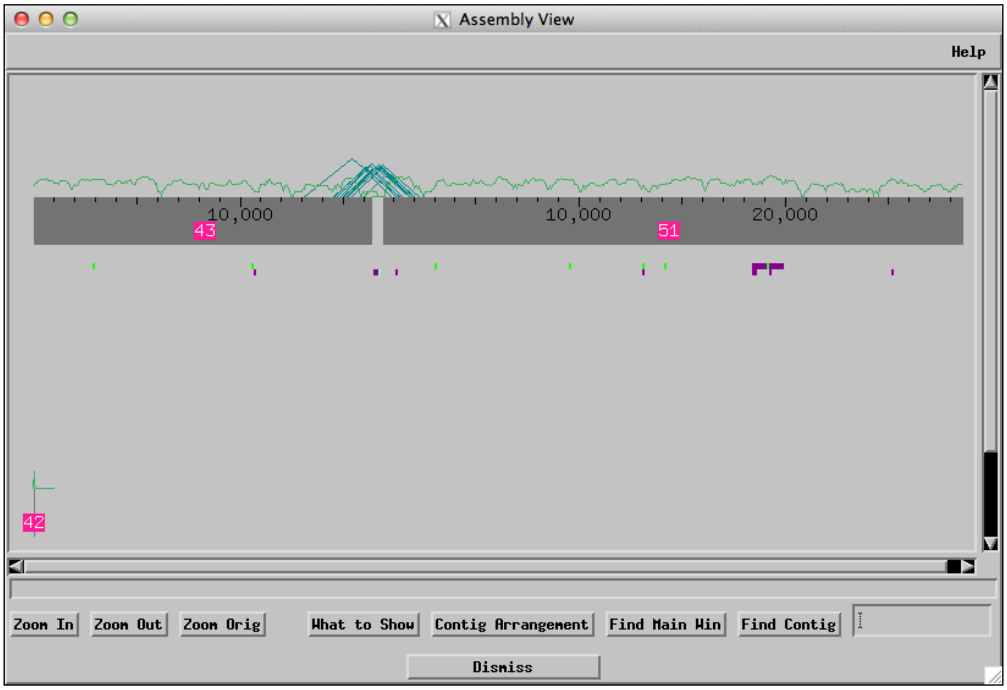


Figure 40. Assembly View after joining 50 and 46

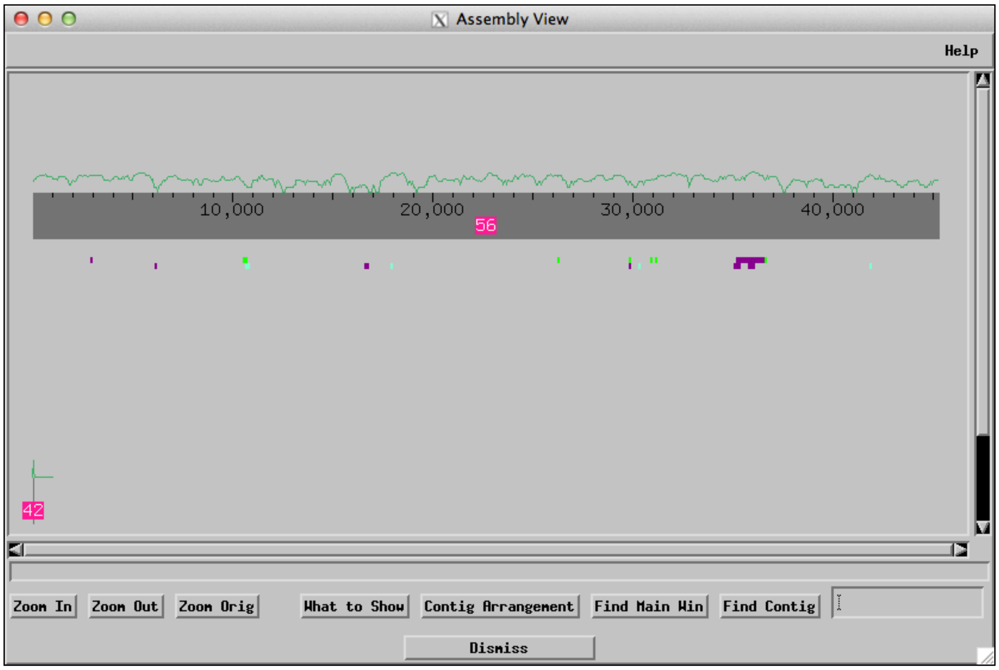


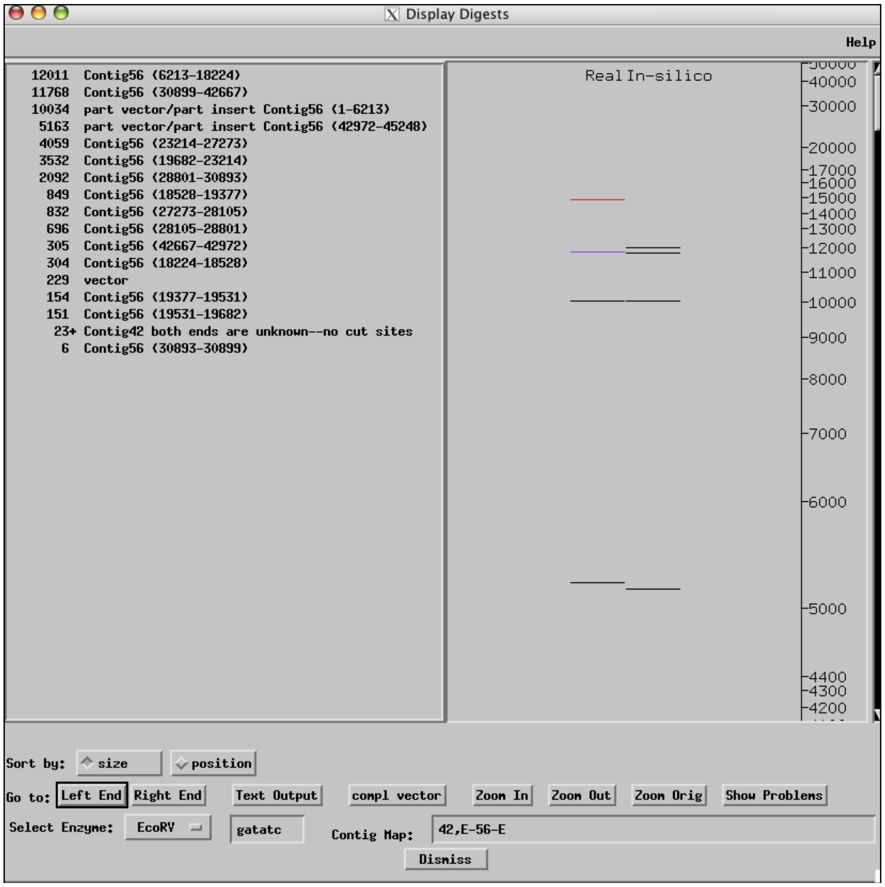
Figure 41. Final Assembly View

Using Restriction Digests to Verify the Final Assembly:

*Note: if you have not resolved all the gaps and misassemblies, you can quit Consed and open the ace file "**XBAA7G24.fasta.screen.ace.digest**" for this part of the tutorial.*

Finally, we can check our assembly by looking at the restriction digests obtained using *Eco*RV and *Hind*III, comparing the *in-vivo* results with the *in-silico* (computer generated) restriction fragments. If our final assembly is correct, the *in-silico* bands should match the bands in the real digest. To examine the digests, from the Consed Main window click on the "Digests" button. Verify that *Eco*RV, *Eco*RI, *Sac*I, and *Hind*III have been selected under the "Commonly Used Restriction Enzymes" section. At the bottom of the screen, verify that the 'Pathname of the Vector Sequence' points to the pfos1.seq vector sequence file. Change the "Full Pathname of File of Vector Sequence:" box to pfos1.seq if necessary (Figure 42).

Note: Because the vector sequence file (pfos1.seq) is in the edit_dir of the project directory, you do not need to specify the full path to the file.



```
12011  Contig56 (6213-18224)
11768  Contig56 (30899-42667)
10034  part vector/part insert Contig56 (1-6213)
 5163  part vector/part insert Contig56 (42972-45248)
 4059  Contig56 (23214-27273)
 3532  Contig56 (19682-23214)
 2092  Contig56 (28801-30893)
  849  Contig56 (18528-19377)
  832  Contig56 (27273-28105)
  696  Contig56 (28105-28801)
  305  Contig56 (42667-42972)
  304  Contig56 (18224-18528)
  229  vector
  154  Contig56 (19377-19531)
  151  Contig56 (19531-19682)
   23+ Contig42 both ends are unknown--no cut sites
    6  Contig56 (30893-30899)
```

Select Enzyme: EcoRV gatatc

Contig Map: 42,E-56-E

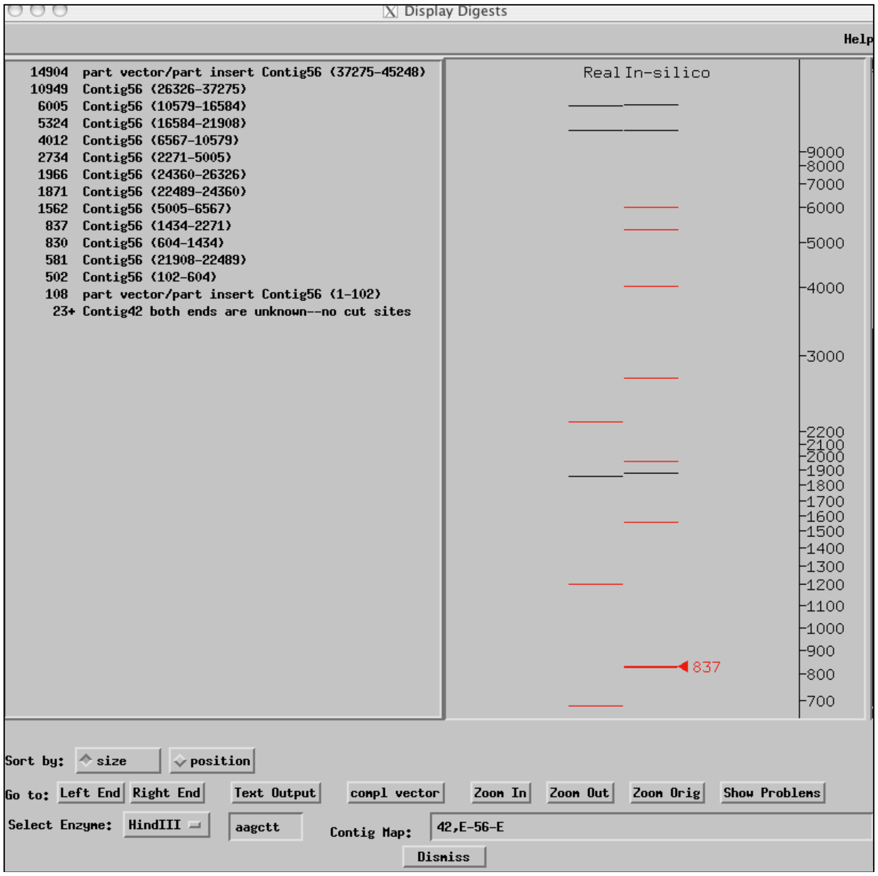
Figure 42. Real vs. *in-silico* digest comparison for EcoRV

Note: To see the entire digest window you may need to click the "Window" menu at the top of your screen then select "Zoom."

The real and *in-silico* *Eco*RV digests mostly agree with each other. The purple line on the real digests signifies a doublet, which corresponds to the two bands in the *in-silico* digest. (It is not unusual for the computer reading the gels to have difficulty discriminating between thick bands and doublets). The red line at the top indicates a fragment in the real digest that is missing in the *in-silico* digest. However, when we add up the sizes of all the *in-silico* fragments, the total size is approximately 46,500 bases, which is close to the estimated size of the fosmid. Hence the top red band in the real digest is likely to be spurious. Because the gel imaging software is optimized to detect bands between 1kb to 10kb, we will focus primarily on fragments in this size range when we compare the *in-silico* with the real digest.

The real *Hind*III digest has many discrepancies compared to the *in-silico* digest (Figure 43). However, the total fragment size from the real *Hind*III digest is substantially less than the expected total size of the fosmid. Hence the real *Hind*III digest may be unreliable. We can check the original gel image to see if there are additional bands in the real digest that were not called by the imaging software.

In addition to *Eco*RV and *Hind*III, we can also examine the real restriction digests for *Sac*I to *Eco*RI to see if our assembly is correct.

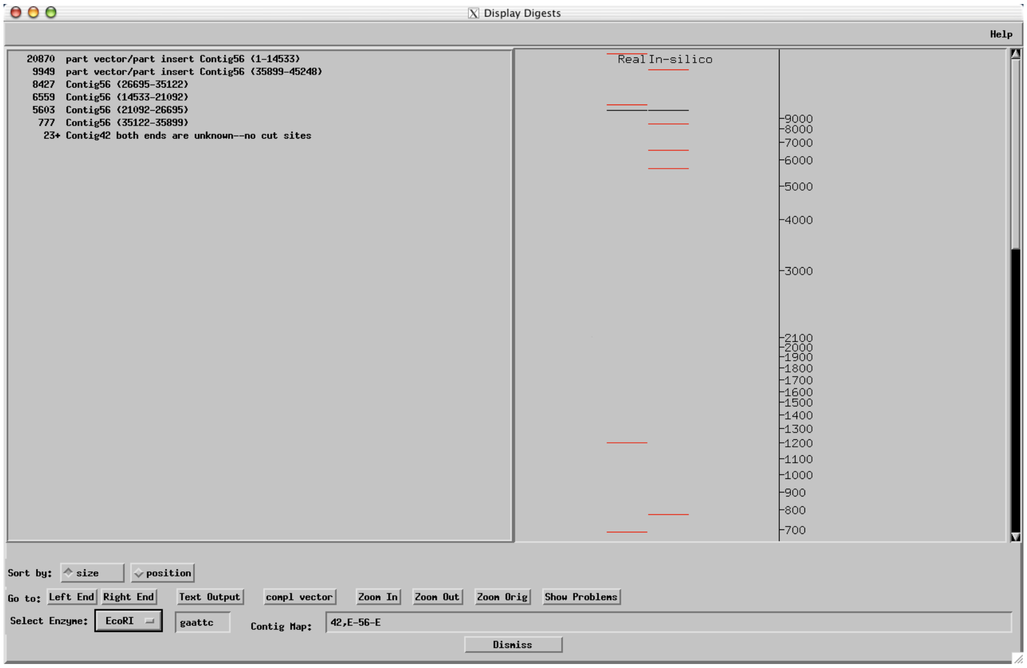


```
14904  part vector/part insert Contig56 (37275-45248)
10949  Contig56 (26326-37275)
 6005  Contig56 (10579-16584)
 5324  Contig56 (16584-21908)
 4012  Contig56 (6567-10579)
 2734  Contig56 (2271-5005)
 1966  Contig56 (24360-26326)
 1871  Contig56 (22489-24360)
 1562  Contig56 (5005-6567)
  837  Contig56 (1434-2271)
  830  Contig56 (604-1434)
  581  Contig56 (21908-22489)
  502  Contig56 (102-604)
  108  part vector/part insert Contig56 (1-102)
   23+ Contig42 both ends are unknown--no cut sites
```

Sort by: size position

Select Enzyme: HindIII aagctt Contig Map: 42,E-56-E

Figure 43. Real vs. *in-silico* digest comparison for HindIII

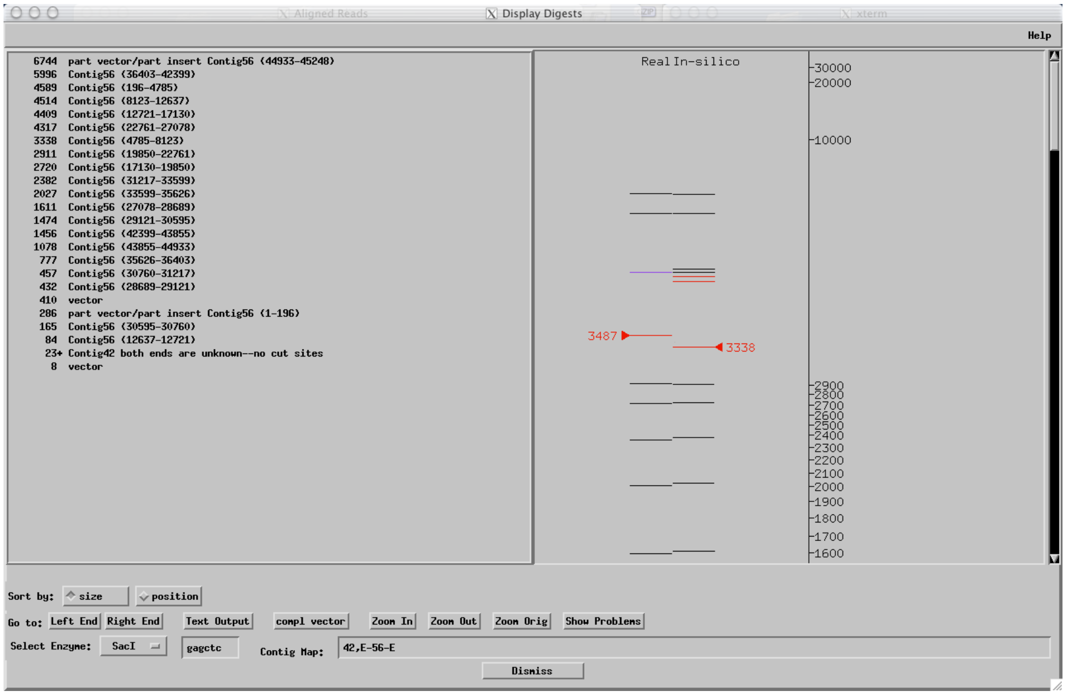


```
  20870  part vector/part insert Contig56 (1-14533)
   9949  part vector/part insert Contig56 (35899-45248)
   8427  Contig56 (26695-35122)
   6559  Contig56 (14533-21092)
   5603  Contig56 (21092-26695)
    777  Contig56 (35122-35899)
     23+ Contig42 both ends are unknown--no cut sites
```

Select Enzyme: EcoRI gaattc Contig Map: 42,E-56-E

Figure 44. EcoRI digest

There are many inconsistencies between the real and *in-silico* digests for *Eco*RI (Figure 44). The *in-silico* digests contains a set of smaller fragments that range from 5.6kb to 8.4 kb in size that do not have corresponding fragments in the real digest. Hence we cannot use this digest to verify our assembly.



```
   6744  part vector/part insert Contig56 (44933-45248)
   5996  Contig56 (36403-42399)
   4589  Contig56 (196-4785)
   4514  Contig56 (8123-12637)
   4409  Contig56 (12721-17130)
   4317  Contig56 (22761-27078)
   3338  Contig56 (4785-8123)
   2911  Contig56 (19850-22761)
   2720  Contig56 (17130-19850)
   2382  Contig56 (31217-33599)
   2027  Contig56 (33599-35626)
   1611  Contig56 (27078-28689)
   1474  Contig56 (29121-30595)
   1456  Contig56 (42399-43855)
   1078  Contig56 (43855-44933)
    777  Contig56 (35626-36403)
    457  Contig56 (30760-31217)
    432  Contig56 (28689-29121)
    410  vector
    286  part vector/part insert Contig56 (1-196)
    165  Contig56 (30595-30760)
     84  Contig56 (12637-12721)
     23+ Contig42 both ends are unknown--no cut sites
      8  vector
```

Select Enzyme: SacI gagctc Contig Map: 42,E-56-E

Figure 45. SacI digest

There are a few inconsistencies in the *Sac*I digest (Figure 45). The two purple bands (around 4.3 to 4.5 kb) in the real digest next to the four *in-silico* bands are actually both doublets. Because of the close proximity of these bands in the *in-silico* digest, the imaging software could have miscalled these bands. We can examine the actual gel image to confirm this hypothesis.

However, in addition to the discrepancies around 4.5 kb, there is another discrepancy at around 3.3 kb. This discrepancy is a cause for concern because we are missing approximately 100 bases compared to the real digest. From the fragment table on the left panel, we can see that the discrepant fragment 3338 is derived from the region ~4700 to ~8000 of Contig 56. Because the misassembly that we have resolved in this tutorial is located outside of this region, the analysis of the real restriction digests is consistent with the hypothesis that we have correctly resolved the misassembly. However, the digests also indicate that there are still a few remaining problem areas that must be resolved before the project can be considered to be finished.

Appendix A

‘Make High Quality’- changes the quality value of the selected bases to 99 and marks them with an orange tag.
‘Change Consensus’- changes the quality value of the bases to 99 and then changes the consensus to the highlighted region of the read.
‘Make Low Quality’- makes the selected bases low quality.
‘Make Low Quality to Left End’- makes all bases to the left of the highlighted region low quality.
‘Make Low Quality to Right End’- makes all bases to the right of the highlighted region low quality.
‘Change to n’s’- Changes the highlighted bases to n’s, which is an indication that are unknown bases.
‘Change to n’s to Left’- Changes the bases to the left of the tagged region to n’s.
‘Change to n’s to Right’- Changes the bases to the right of the tagged region to n’s.
‘Change to x’s to Left’- Changes the highlighted bases and the bases to the left of this region to x’s which means they are vector sequence.
‘Change to x’s to Right’- Changes the highlighted bases and the bases to the right of this region to x’s.
‘Add Tag’- allows you to add any tag to a stretch of read bases.

Last Update: 07/29/2013