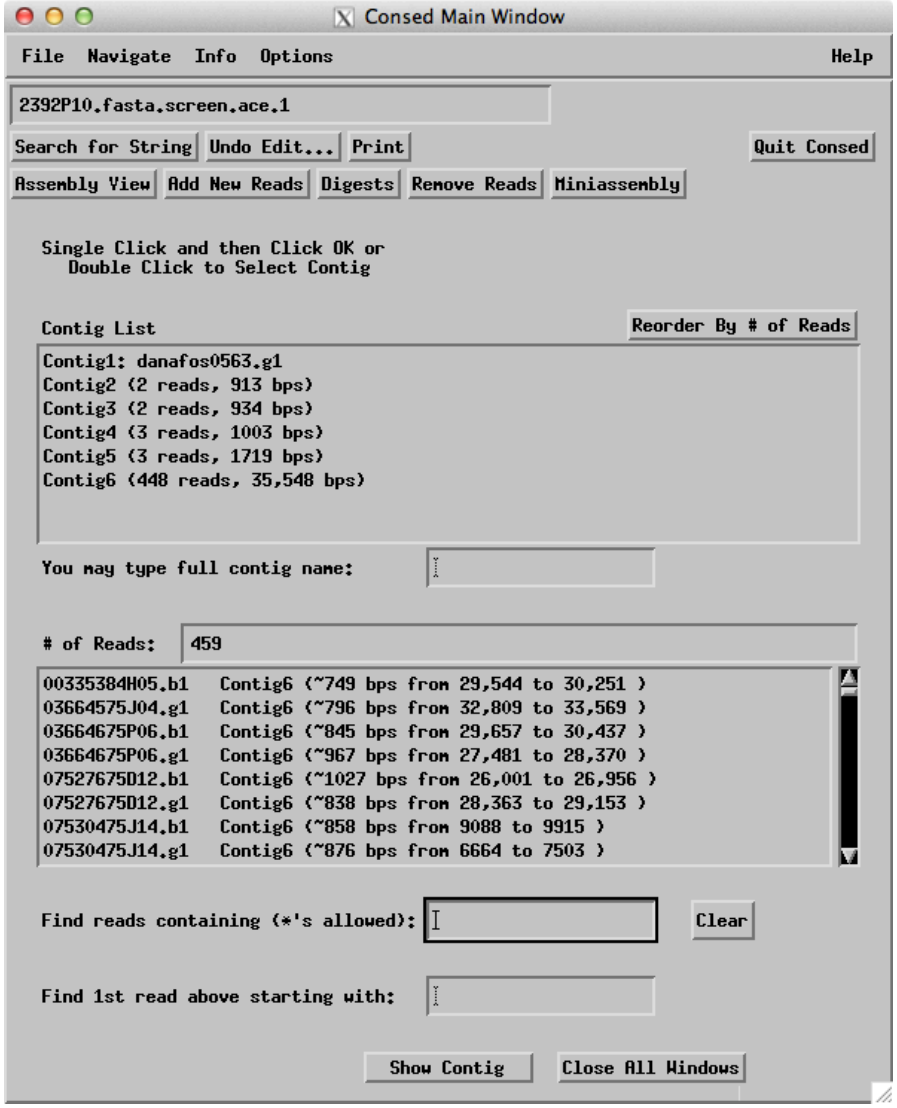
A Guide to Consed

Michelle Itano, Carolyn Cain, Tien Chusak, Justin Richner, and SCR Elgin.

Main Window



```
2392P10.fasta.screen.ace.1

Contig1: danafos0563.g1
Contig2 (2 reads, 913 bps)
Contig3 (2 reads, 934 bps)
Contig4 (3 reads, 1003 bps)
Contig5 (3 reads, 1719 bps)
Contig6 (448 reads, 35,548 bps)

# of Reads: 459

00335384H05.b1   Contig6 (~749 bps from 29,544 to 30,251 )
03664575J04.g1   Contig6 (~796 bps from 32,809 to 33,569 )
03664675P06.b1   Contig6 (~845 bps from 29,657 to 30,437 )
03664675P06.g1   Contig6 (~967 bps from 27,481 to 28,370 )
07527675D12.b1   Contig6 (~1027 bps from 26,001 to 26,956 )
07527675D12.g1   Contig6 (~838 bps from 28,363 to 29,153 )
07530475J14.b1   Contig6 (~858 bps from 9088 to 9915 )
07530475J14.g1   Contig6 (~876 bps from 6664 to 7503 )
```

Figure 1. The ‘Main Window’ is the starting point when Consed is opened. From here, you can access the ‘Assembly View’ and all contigs and reads.

How to access: Start Consed.

Main features:

- Clicking on the ‘Assembly View’ button leads to a visual representation of your contigs (see Figure 2).

- All contigs are shown in the ‘Contig List’ box. Contigs significant to your overall assembly are generally larger than 2000 bps. Double clicking on a contig in the ‘Contig List’ opens the Aligned Reads Window for that contig.
- The ‘Read List’ displays all reads that were placed into a contig. The name of the read is to the far left, followed by the name of its contig and location in the contig. The chemistry of the reaction that created each read is indicated in the read’s name. A read that ends in .b1 or .g1 was produced with Big Dye chemistry. Reads ending in _g.b1 or _g.g1 were created with dGTP chemistry. A read with _t.b1 or _t.g1 at the end was produced with 4:1 chemistry. Double clicking on a read in the ‘Read List’ opens an ‘Aligned Reads Window’ scrolled to the location of the selected read in its contig (see Figure 7).

Assembly View

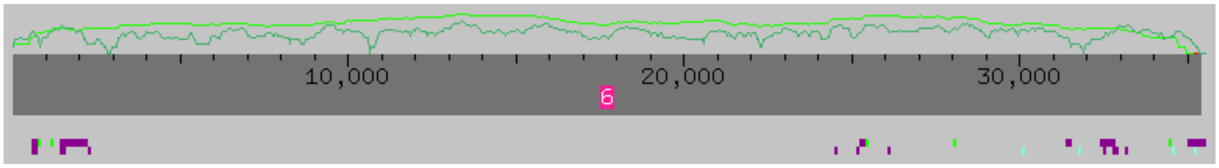


Figure 2. Assembly View: a clone that assembled into one contig, with no gaps.

The “Assembly View” button can be found on the main menu of Consed. Assembly View compiles a graphical representation of the major contigs and puts them in spatial relationship to one another using the initial forward/reverse pairs. Assembly View does not show contigs shorter than a set number of base pairs, and this number can be changed. Assembly View is a valuable tool used to quickly scan for problems within your assembly.

In Figure 2, we see that the user has only one contig of about 35,000 bases. The ultimate goal of a finisher, among other things, is an Assembly View with one contig. Above the contigs there are two different lines indicating data quality. The dark green line represents the number of high quality reads at that point in the contig. The higher the line, the more high quality reads there are. In Figure 2 at about 11,000 bases, the finisher has almost zero high quality reads. The light green line above the contig represents the total number of consistent forward and reverse mate pairs at that position, both high and low quality. Another goal of the finisher is to make sure there are high quality reads at every base in the contig.

Below the contig are purple and light blue boxes that represent tagged data. The tags shown in Assembly View can be changed by clicking the “What to Show” button.

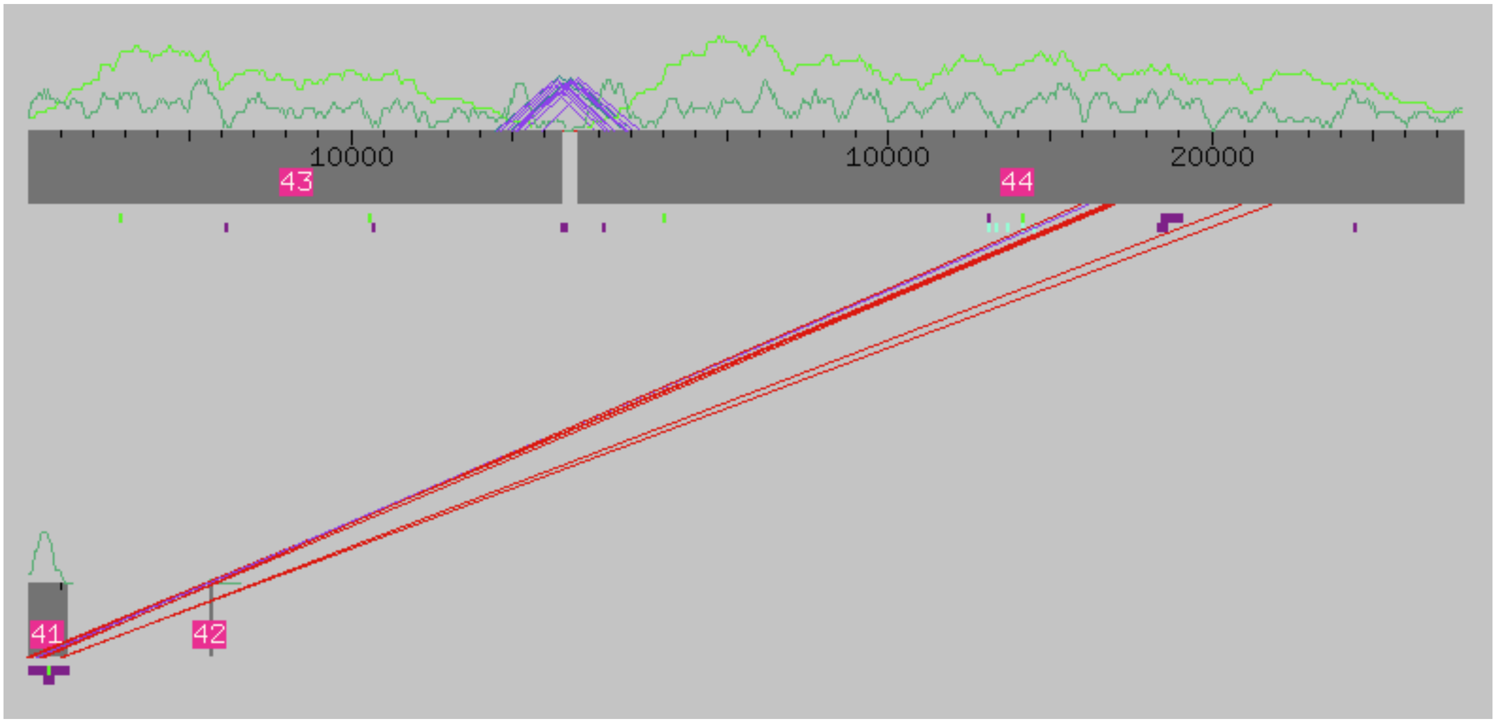


Figure 3. Assembly View for a clone assembled into four contigs.

Figure 3 shows a more complicated Assembly View. The assembly has four different contigs. The purple triangular lines connecting contigs 43 and 44 represent individual subclones that span the gap. Each subclone with a purple line contains a pair of reads (forward/reverse pair) with one on contig 43 and the other on contig 44. Gaps that are covered in this way are easy to close, because the finisher just has to call reactions starting from a primer further within the subclone that spans the gap.

The straight red lines running from contig 44 to contig 41 are also forward/reverse pairs from individual subclones. The red color indicates that the spacing of these pairs is not correct. Each subclone is roughly 3-5kb in size, so the forward/reverse pairs from one subclone must be roughly this far apart. If the spacing is incorrect, Assembly View connects the subclone's reads with a red line instead of purple. The incorrect forward/reverse pairs are called "inconsistent." In the above Assembly View, contig 41 probably belongs in the middle of contig 44, which would make the forward/reverse pairs consistent.

Figure 3 also shows contig 42 to be small and not connected to any other contigs by subclones. Both of these traits are signs that contig 42 is vector sequence, contamination from the *E. coli* strain used as a host for our sequence. Vector sequence is usually too short to be included into Assembly View. To make sure that any errant contig is vector sequence, the finisher can use the web-tool Blast. When the finisher uses Blast to check the contamination, he/she should find very high quality matches to known host sequence.

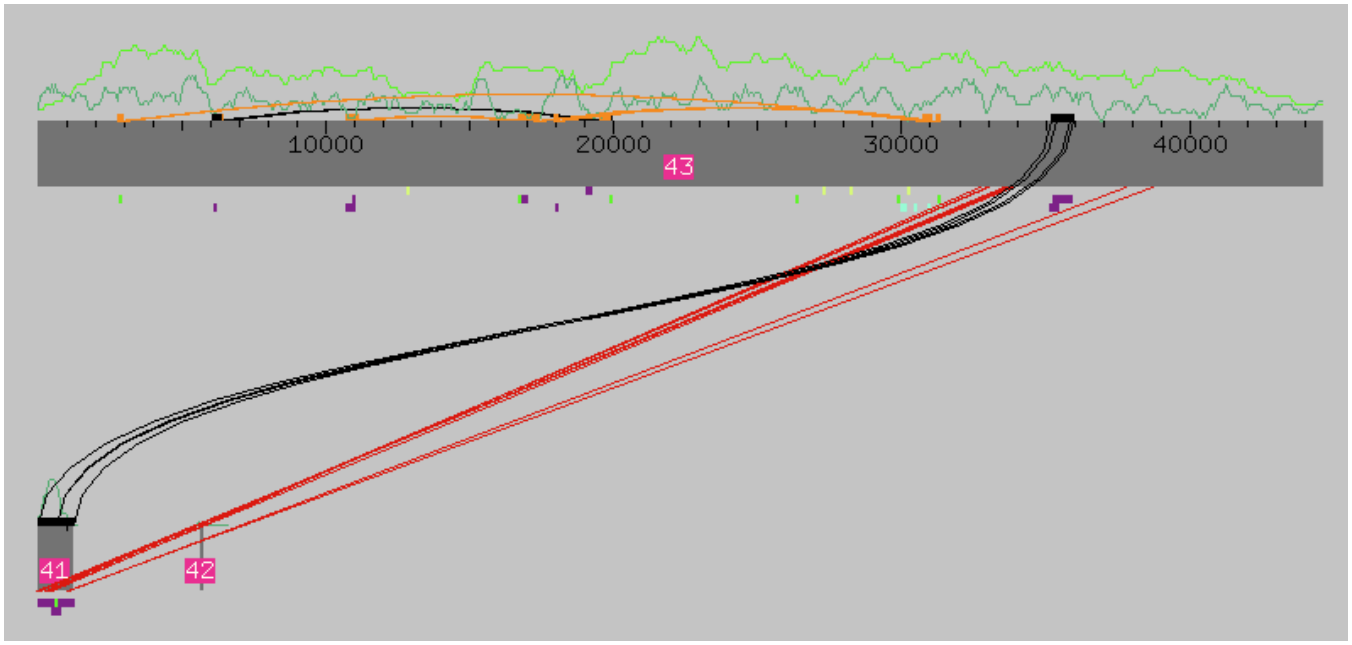


Figure 4. Assembly View showing the results of Crossmatch

Figure 4 shows Assembly View after Crossmatch was run. Crossmatch is a utility within Assembly View that shows areas of repeats within the assembly. Crossmatch can be turned on by navigating to “Sequence Matches” under the “What to Show” button. The region under the orange or black boxes represents the region in the assembly that is repeated. The two repeated regions are connected by a orange or black line. Orange boxes and lines represent repeats that are on the same strand of DNA. Black boxes and lines represent repeats on the two different strands of DNA. Crossmatch can be used to diagnose misassemblies. In Figure 4 the repeated sections flanking contig 41 indicate that contig 41 most likely should be inserted within the black box region of contig 43.

ASSEMBLY VIEW IS NOT ALWAYS RIGHT!!!!

This is the most important thing to remember about Assembly View. If the entire assembly is low quality or contains a lot of repeats, then the computer is likely to misassemble the data and have an errant Assembly View. Assembly View is a good tool to browse through the entire assembly. But the data is in the sequence, and this is where the finisher should do most of his/her work.

Navigation Windows

Consed allows the user to navigate through the assembly by identifying possible problem areas for finishers. It is possible to look for the following types of problems:

-high quality discrepancies (sites where different bases have been called in different reads, in areas that otherwise appear to match with high quality data in each read)
-which may indicate a misassembly or may indicate miscalled bases

-regions covered by only one single subclone
-which may indicate unreliable base calling given the lack of depth of coverage

-low consensus quality
-which may indicate miscalled bases or unreliable base calling

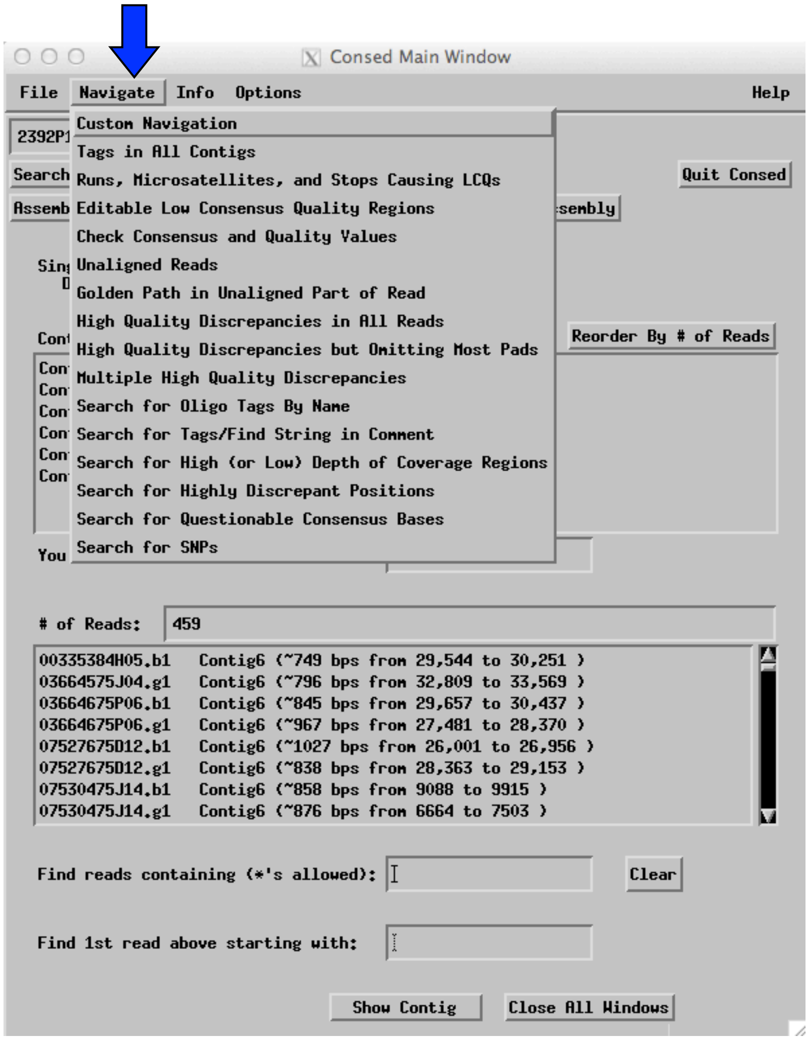
-regions covered by only one strand and one chemistry
-which may indicate unreliable base calling given that the same type of error could occur, without a check from the second strand or a different chemistry

-unaligned high quality regions
-which may indicate a misassembly or miscalled bases

Consed allows navigation to look at subcategories within the above categories (omitting pads, etc.). It is also possible to navigate to edits or tags, which makes it easier for the finisher to return to problem areas and check the effects of previously-made edits.

These windows are convenient for finishers because they allow the finisher to focus on a particular problem to improve the assembly. If desired, the finisher can systematically observe and resolve all of the regions containing a particular type of problem in one session. All of these characteristics indicate possible problems with the assembly that should be resolved before the sequence is considered finished to high quality standards.

It is possible to reach the Navigation windows either by selecting the “Navigate” tab on the Main Window (Figure 5a) or by selecting the “Navigate” tab from an Aligned Reads Window (Figure 5b). The navigation options available from the Main Window allow the user to navigate through the entire assembly, while those available from the Aligned Reads Window only navigate through the selected contig.

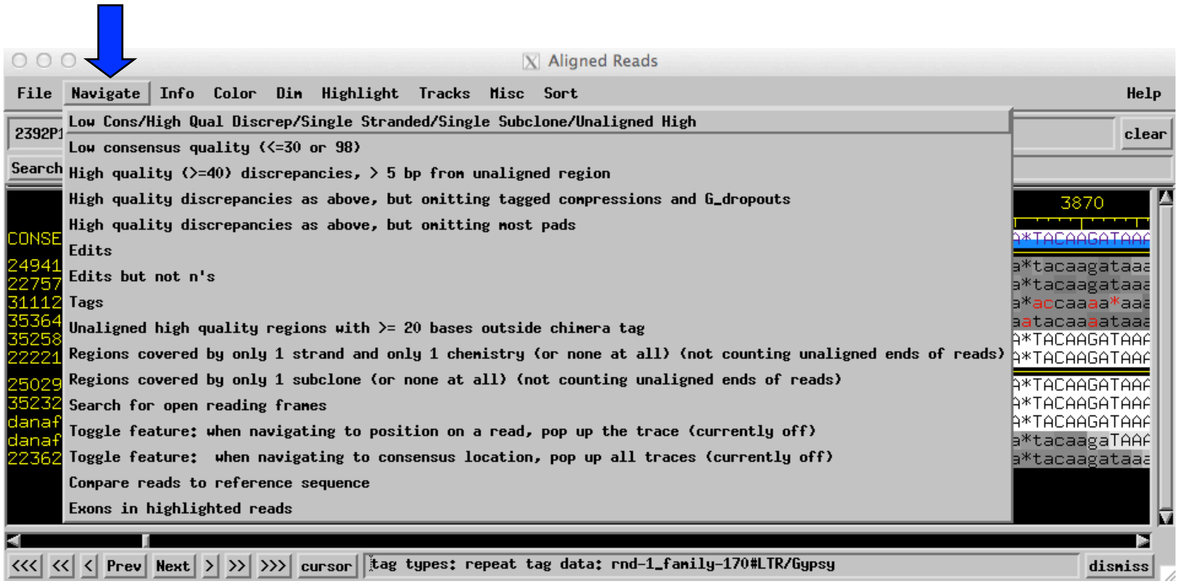

Consed Main Window
File
Navigate
Info
Options
Help
Custom Navigation
Tags in All Contigs
Runs, Microsatellites, and Stops Causing LCQs
Editable Low Consensus Quality Regions
Check Consensus and Quality Values
Unaligned Reads
Golden Path in Unaligned Part of Read
High Quality Discrepancies in All Reads
High Quality Discrepancies but Omitting Most Pads
Multiple High Quality Discrepancies
Search for Oligo Tags By Name
Search for Tags/Find String in Comment
Search for High (or Low) Depth of Coverage Regions
Search for Highly Discrepant Positions
Search for Questionable Consensus Bases
Search for SNPs
Quit Consed
Reorder By # of Reads
Clear
Show Contig
Close All Windows

```
# of Reads:  459

00335384H05.b1   Contig6 (~749 bps from 29,544 to 30,251 )
03664575J04.g1   Contig6 (~796 bps from 32,809 to 33,569 )
03664675P06.b1   Contig6 (~845 bps from 29,657 to 30,437 )
03664675P06.g1   Contig6 (~967 bps from 27,481 to 28,370 )
07527675D12.b1   Contig6 (~1027 bps from 26,001 to 26,956 )
07527675D12.g1   Contig6 (~838 bps from 28,363 to 29,153 )
07530475J14.b1   Contig6 (~858 bps from 9088 to 9915 )
07530475J14.g1   Contig6 (~876 bps from 6664 to 7503 )

Find reads containing (*'s allowed):
Find 1st read above starting with:
```

Figure 5a. Navigation options located in the Main Window

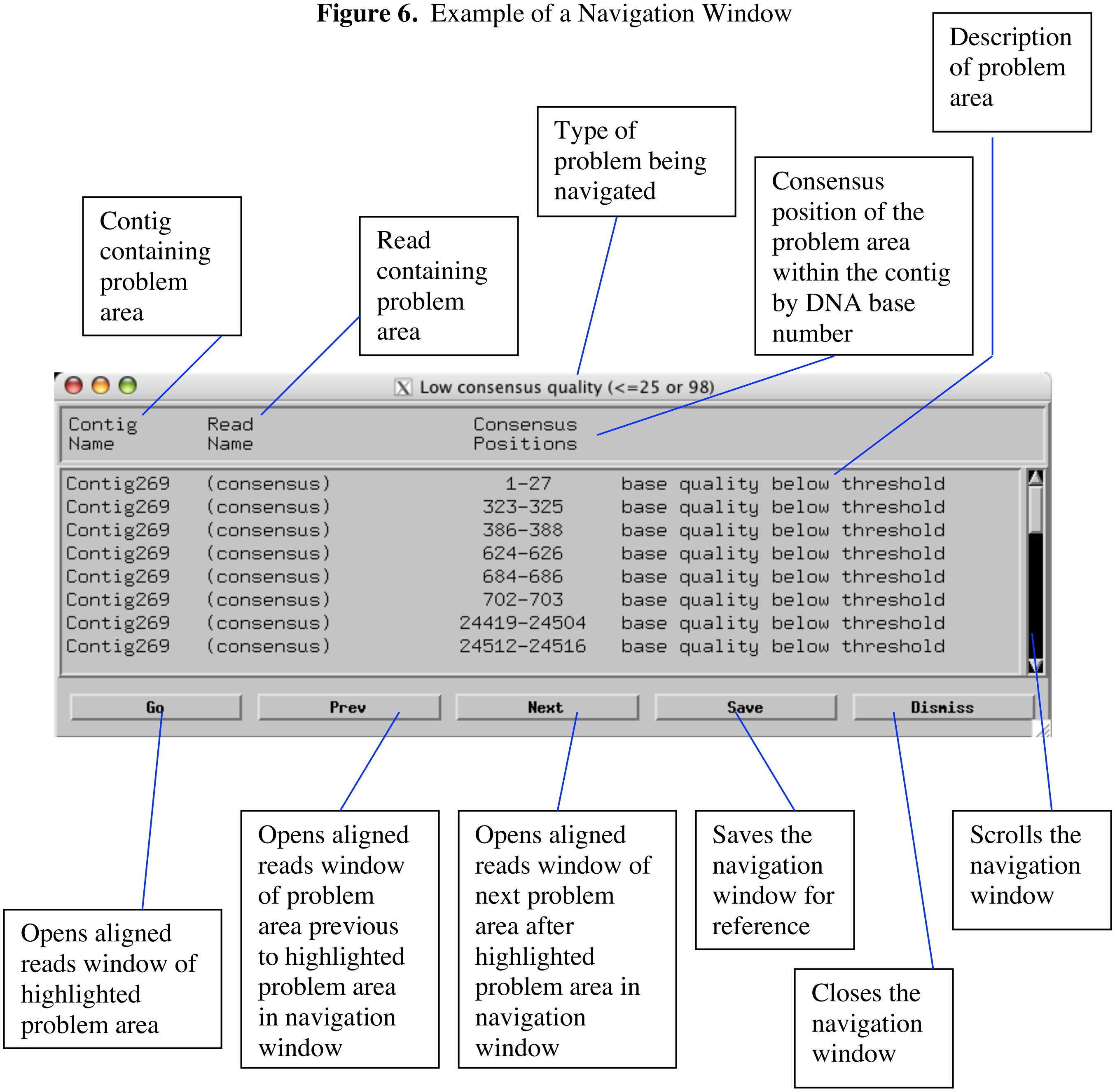

Aligned Reads
File
Navigate
Info
Color
Dim
Highlight
Tracks
Misc
Sort
Help
Low Cons/High Qual Discrep/Single Stranded/Single Subclone/Unaligned High
Low consensus quality (<=30 or 98)
High quality (>=40) discrepancies, > 5 bp from unaligned region
High quality discrepancies as above, but omitting tagged compressions and G_dropouts
High quality discrepancies as above, but omitting most pads
Edits
Edits but not n's
Tags
Unaligned high quality regions with >= 20 bases outside chimera tag
Regions covered by only 1 strand and only 1 chemistry (or none at all) (not counting unaligned ends of reads)
Regions covered by only 1 subclone (or none at all) (not counting unaligned ends of reads)
Search for open reading frames
Toggle feature: when navigating to position on a read, pop up the trace (currently off)
Toggle feature: when navigating to consensus location, pop up all traces (currently off)
Compare reads to reference sequence
Exons in highlighted reads
clear
3870
<<<
<<
<
Prev
Next
>
>>
>>>
cursor
dismiss

```
tag types: repeat tag data: rnd-1_family-170#LTR/Gypsy
```

Figure 5b. Navigation options located in the Aligned Reads Window

All of the windows are similar in design to the one shown below (Figure 6):

Figure 6. Example of a Navigation Window



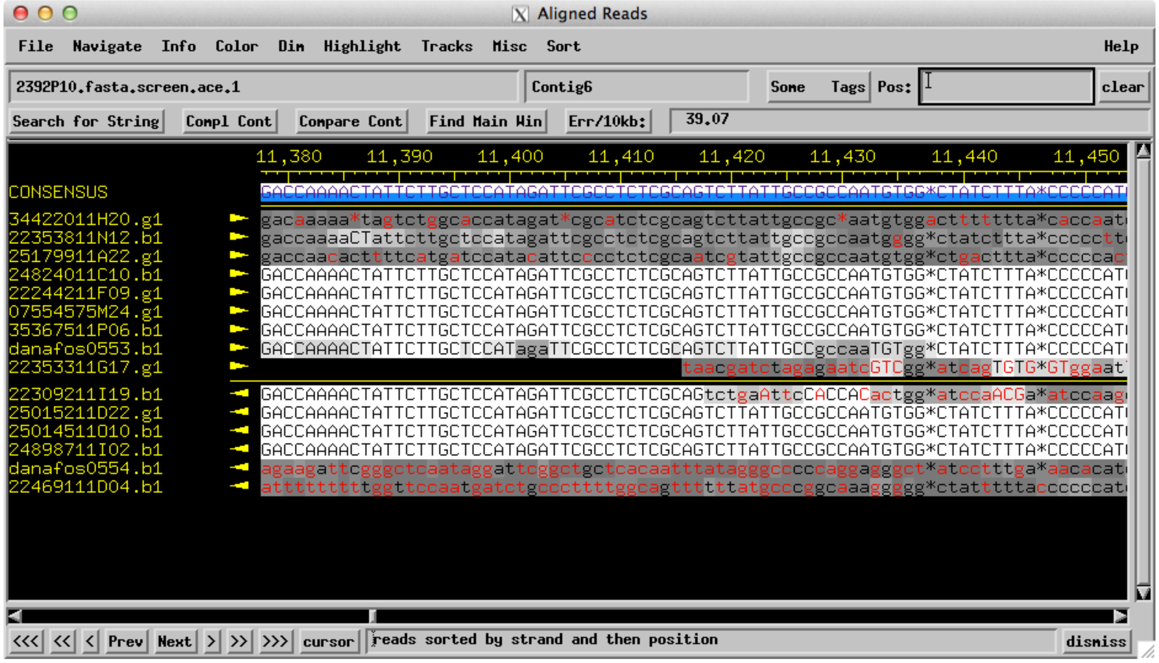
```
X Low consensus quality (<=25 or 98)

Contig        Read                    Consensus
Name          Name                    Positions

Contig269     (consensus)              1-27        base quality below threshold
Contig269     (consensus)            323-325       base quality below threshold
Contig269     (consensus)            386-388       base quality below threshold
Contig269     (consensus)            624-626       base quality below threshold
Contig269     (consensus)            684-686       base quality below threshold
Contig269     (consensus)            702-703       base quality below threshold
Contig269     (consensus)          24419-24504     base quality below threshold
Contig269     (consensus)          24512-24516     base quality below threshold
```

Description of problem area
Type of problem being navigated
Consensus position of the problem area within the contig by DNA base number
Contig containing problem area
Read containing problem area
Go
Prev
Next
Save
Dismiss
Opens aligned reads window of highlighted problem area
Opens aligned reads window of problem area previous to highlighted problem area in navigation window
Opens aligned reads window of next problem area after highlighted problem area in navigation window
Saves the navigation window for reference
Closes the navigation window
Scrolls the navigation window

Figure 7. Aligned Reads Window



2392P10.fasta.screen.ace.1

Contig6

Err/10kb: 39.07

```
                            11,380    11,390    11,400    11,410    11,420    11,430     11,440     11,450
CONSENSUS           GACCAAAACTATTCTTGCTCCATAGATTCGCCTCTCGCAGTCTTATTGCCGCCAATGTGG*CTATCTTTA*CCCCCAT
34422011H20.g1   ►  gacaaaaa*tagtctggcaccatagat*cgcatctcgcagtcttattgccgc*aatgtggacttttttta*caccaat
22353811N12.b1   ►  gaccaaaaCTattcttgctccatagattcgcctctcgcagtcttattgccgccaatgggg*ctatcttta*ccccctt
25179911A22.g1   ►  gaccaacactttttcatgatccatacattcccctctcgcaatcgtattgccgccaatgtgg*ctgacttta*cccccac
24824011C10.b1   ►  GACCAAAACTATTCTTGCTCCATAGATTCGCCTCTCGCAGTCTTATTGCCGCCAATGTGG*CTATCTTTA*CCCCCAT
22244211F09.g1   ►  GACCAAAACTATTCTTGCTCCATAGATTCGCCTCTCGCAGTCTTATTGCCGCCAATGTGG*CTATCTTTA*CCCCCAT
07554575M24.g1   ►  GACCAAAACTATTCTTGCTCCATAGATTCGCCTCTCGCAGTCTTATTGCCGCCAATGTGG*CTATCTTTA*CCCCCAT
35367511P06.b1   ►  GACCAAAACTATTCTTGCTCCATAGATTCGCCTCTCGCAGTCTTATTGCCGCCAATGTGG*CTATCTTTA*CCCCCAT
danafos0553.b1   ►  GACCAAAACTATTCTTGCTCCATagaTTCGCCTCTCGCAGTCTTATTGCCgccaaTGTgg*CTATCTTTA*CCCCCAT
22353311G17.g1   ►                                        taacgatctagagaatcGTCgg*atcagTGTG*GTggaat
22309211I19.b1   ◄  GACCAAAACTATTCTTGCTCCATAGATTCGCCTCTCGCAGtctgaAttcCACCACactgg*atccaACGa*atccaag
25015211D22.g1   ◄  GACCAAAACTATTCTTGCTCCATAGATTCGCCTCTCGCAGTCTTATTGCCGCCAATGTGG*CTATCTTTA*CCCCCAT
25014511O10.b1   ◄  GACCAAAACTATTCTTGCTCCATAGATTCGCCTCTCGCAGTCTTATTGCCGCCAATGTGG*CTATCTTTA*CCCCCAT
24898711I02.b1   ◄  GACCAAAACTATTCTTGCTCCATAGATTCGCCTCTCGCAGTCTTATTGCCGCCAATGTGG*CTATCTTTA*CCCCCAT
danafos0554.b1   ◄  agaagattcgggctcaataggattcggctgctcacaatttatagggcccccaggagggct*atcctttga*aacacat
22469111D04.b1   ◄  atttttttttggttccaatgatctgcccttttggcagtttttttatgcccggcaaaggggg*ctattttta*cccccat
```

reads sorted by strand and then position

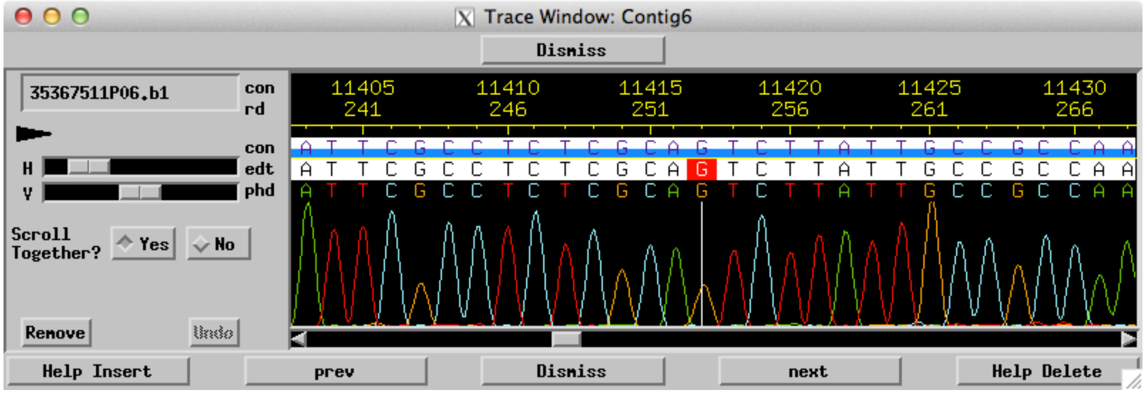
The 'Aligned Reads Window' shows the alignment of all reads in a contig, and the consensus sequence for a contig.

How to access: Double click on a contig in the 'Main Window.'

Main Features:

- Click on a read sequence with your middle mouse button (or holding the 'alt+option' key and clicking with a one-button mouse) to view the 'Trace Window'
- Click on 'Search for String' to search for an exact or approximate sequence.
- Click the 'Comp Cont' button to display the complement of all sequences in the contig (all 'a's change to 't,' all 'g's change to 'c,' etc.).
- Click on 'Compare Cont' to compare this consensus sequence to another contig sequence. (This is used when considering a forced join.)
- Clicking 'Find Main Window' brings the 'Main Window' to the front of all other windows on your screen.
- Click on the arrow buttons on the bottom to go forward or backward in the sequence. To scroll by a small amount, click the '<' or '>' button. To scroll by a large amount, click the '<<' or '>>' button. The '<<<' and '>>>' buttons scroll to the beginning and end (respectively) of the contig.
- Click on the name of a trace (e.g. 35367511P06.b1) to highlight it.
- Enter a number in the 'Pos: ' box to jump to that position in the contig.

Figure 8: Trace Window


Trace Window: Contig6
Dismiss
35367511P06.b1
con
rd
con
edt
phd
H
V
Scroll Together?
Yes
No
Remove
Undo
11405
241
11410
246
11415
251
11420
256
11425
261
11430
266
A T T C G C C T C T C G C A G T C T T A T T G C C G C C A A
A T T C G C C T C T C G C A G T C T T A T T G C C G C C A A
A T T C G C C T C T C G C A G T C T T A T T G C C G C C A A
Help Insert
prev
Dismiss
next
Help Delete

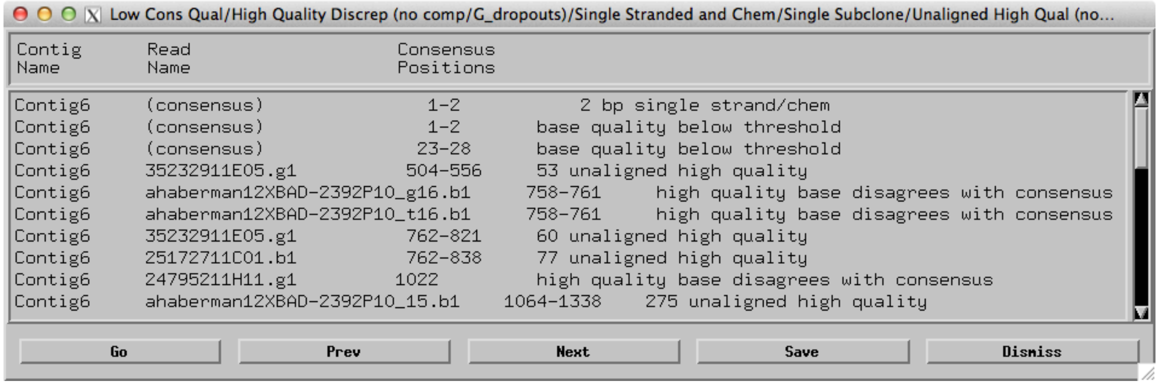
The 'Trace Window' displays the raw output of a sequencing reaction; the different color peaks correspond to the four nucleotides.

How to access: Click with the middle mouse button (or hold the 'alt+option' key and click with a one-button mouse) on a read sequence in the Aligned Reads Window.

Main Features:

- Swipe over part of the sequence (letters) while holding down the middle mouse button (or while holding the 'alt+option' key and clicking with a one-button mouse) to edit or tag the sequence.
- Slide 'H' and 'V' bars to change the size of the peaks horizontally and vertically.
- Click on the arrows at the bottom to see sequence to the left or right.

Figure 9. Navigator Window


Low Cons Qual/High Quality Discrep (no comp/G_dropouts)/Single Stranded and Chem/Single Subclone/Unaligned High Qual (no...
Go
Prev
Next
Save
Dismiss

```
Contig      Read                       Consensus
Name        Name                       Positions

Contig6     (consensus)                   1-2          2 bp single strand/chem
Contig6     (consensus)                   1-2       base quality below threshold
Contig6     (consensus)                  23-28      base quality below threshold
Contig6     35232911E05.g1              504-556     53 unaligned high quality
Contig6     ahaberman12XBAD-2392P10_g16.b1     758-761     high quality base disagrees with consensus
Contig6     ahaberman12XBAD-2392P10_t16.b1     758-761     high quality base disagrees with consensus
Contig6     35232911E05.g1              762-821     60 unaligned high quality
Contig6     25172711C01.b1              762-838     77 unaligned high quality
Contig6     24795211H11.g1             1022         high quality base disagrees with consensus
Contig6     ahaberman12XBAD-2392P10_15.b1    1064-1338    275 unaligned high quality
```

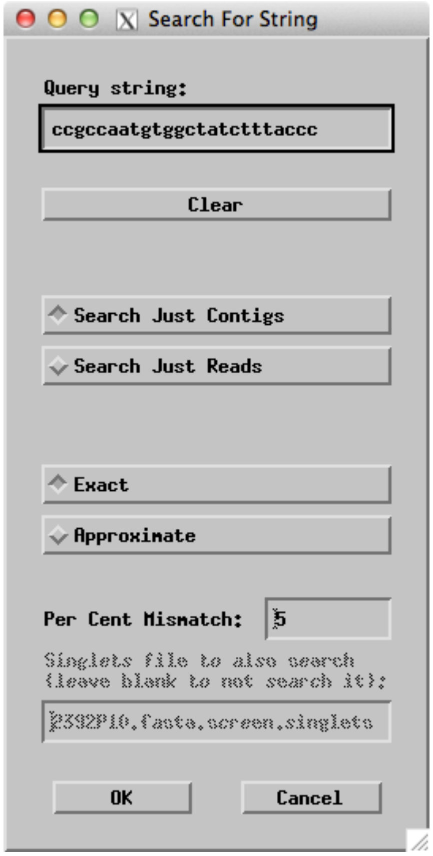
The 'Navigator Window' displays the position and description of possible problems in a contig.

How to access: In the 'Aligned Reads Window,' click on 'Navigator' in the top menu bar and choose a particular navigator from the drop-down menu.

Main Features:

- Double click on a line describing a problem to jump to the problem location in the contig.

Figure 10. Search for String Window



Query string: ccgccaatgtggctatctttaccc

Per Cent Mismatch: 5

Singlets file to also search (leave blank to not search it): P392P10.fasta.screen.singlets

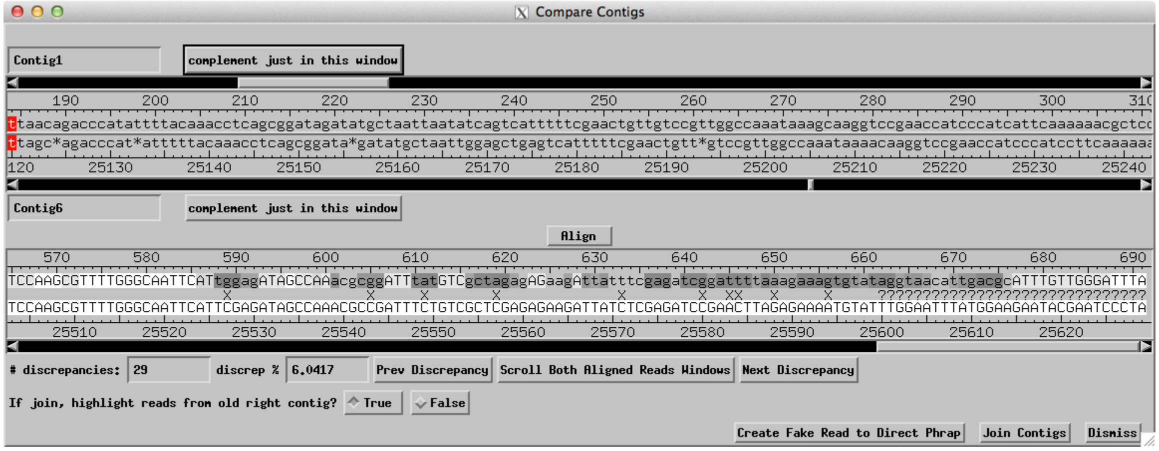
The ‘Search for String Window’ is the interface through which you can locate a string of bases in the assembly.

How to access: Click on the ‘Search for String’ box in either the ‘Main Window’ or the ‘Aligned Reads Window.’

Main Features:

- Enter a sequence to search for in the ‘Query string:' box.
- Click ‘Exact’ to search for an exact match. Click ‘Approximate’ and enter a number in ‘Per Cent Mismatch’ box to search for an approximate match.
- Use this to make sure that selected primers are unique.

Figure 11. Compare Contigs Window


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

Contig1

```
    190       200       210       220       230       240       250       260       270       280       290       300       310
ttaacagacccatattttacaaacctcagcggatagatatgctaattaatatcagtcattttcgaactgttgtccgttggccaaataaagcaaggtccgaaccatcccatcattcaaaaaacgctcc
ttagc*agacccat*attttacaaacctcagcggata*gatatgctaattggagctgagtcattttcgaactgtt*gtccgttggccaaataaaacaaggtccgaaccatcccatccttcaaaaaa
120      25130     25140     25150     25160     25170     25180     25190     25200     25210     25220     25230     25240
```

Contig6

```
   570       580       590       600       610       620       630       640       650       660       670       680       690
TCCAAGCGTTTTGGGCAATTCATtggagATAGCCAAacgcggATTtatGTCgctagagAGaagAttatttcgagatcggattttaaagaaagtgtataggtaacattgacgcATTTGTTGGGATTTA
                        X               X     X       X             X        X  XX   X     X     ??????????????????????????????
TCCAAGCGTTTTGGGCAATTCATTCGAGATAGCCAAACGCCGATTTCTGTCGCTCGAGAGAAGATTATCTCGAGATCCGAACTTAGAGAAAATGTATTTGGAATTTATGGAAGAATACGAATCCCTA
    25510     25520     25530     25540     25550     25560     25570     25580     25590     25600     25610     25620
```

discrepancies: 29   discrep % 6.0417

If join, highlight reads from old right contig?

The ‘Compare Contigs Window’ allows you to align, compare, and join two overlapping sequences from different contigs.

How to access: Click on ‘Compare Cont’ in the ‘Aligned Reads Windows’ for both of the contigs you would like to compare.

Main Features:

- The top portion of the window shows the initial alignment of the two sequences as imported from the contigs you are comparing. In this example, the top sequence is from Contig1 and the bottom sequence is from Contig6. At this point, you may scroll each sequence if they are not aligned properly.
- Click on ‘Align’ to see how well the sequences match up.
- The bottom portion of the window displays information after ‘Align’ is clicked. This section evaluates the alignment of the two sequences. Poor matching is indicated by symbols (such as ‘X’ and ‘?’) in between the two sequences.
- Click on ‘Join Contigs’ if you have confidence in the alignment, and would like to join the two contigs together.

GLOSSARY

.ace file – A computer file containing all of the sequence reads for a given project, often a fosmid or BAC clone.

Aligned reads – Consed window that shows all of the reads in a particular region. This window shows low quality sequence in shades of gray, and high quality sequence in white.

Assembly View – Consed window that shows the length and number of contigs based on the most recent Phred/Phrap compilation of sequence data. Forward/reverse pairs with reads in different contigs, and the amount of coverage are shown.

Assembly – the ordered compilation of read data by Phred/Phrap that uses sequence matches, forward/reverse read pairs, etc. to determine read location and orientation in the contig.

Autofinish – a computer program used to call reactions prior to a human finisher working on the sequence. It is fast, calls very efficient reactions, and tags editable regions for the human finisher to work on.

BAC – Bacterial Artificial Chromosome. A large segment of DNA ranging in size from 100,000 to 200,000 bp in length that is propagated in bacteria.

Chemistry – different mixtures of nucleotides used in sequencing reactions which yield better or worse results when dealing with problem regions. Commonly used chemistries include dGTP, Big Dye, and a 4:1 ratio of the two.

Clone – *E. coli* or other host containing a single recombinant DNA molecule to be sequenced

Complemented sequence match – A match reported in the Search for String window that finds the complementary sequence (sequence from the opposite strand) from that used to search in the consensus sequence of the assembly. Contrast with an **Uncomplemented sequence match** – A match reported in the Search for String window that finds the exact sequence used to search in the consensus sequence in the assembly.

Consed – program used to view and edit the data assembled by Phred/Phrap.

Consensus – this is the sequence generated by Phred/Phrap that represents the algorithm's best guess at the correct sequence of the DNA.

Contig – an assembly of overlapping clones, based on shared sequence or shared restriction fragments

Crossmatch – a program that finds similar sequences in Assembly View and marks them with orange bars (uncomplemented matches) or black bars (complemented matches) and corresponding lines connecting the matching areas.

Editing – the process of visually examining traces in Consed and deciding whether or not the consensus accurately reflects the sequence data for that region. Edited regions are tagged to alert people to the fact that the sequence was edited by a human and is not simply the product of the base-calling algorithm.

Finishing – the process of taking raw sequence data and creating a high quality sequence by editing and calling reactions to gather more data where needed.

Forced joins – the result of combining two contigs together in response to directions from the finisher to make one combined contig during the assembly process.

Forward/reverse pairs– see Paired end reads.

Fosmid – a cloning vector capable of containing an insert of 35-40 kb.

Gap – region with no sequence data. A gap may be spanned by forward/reverse pairs. Reactions are called by the finisher to add enough data to the assembly to close the gap.

Insert – DNA of interest that is ligated into the cloning vector. Genomic or cloned DNA is fragmented into 1-2 kb pieces that are then cloned and sequenced.

In-silico digest – a restriction digest done using the assembly created by Phred/Phrap and Consed to compare the assembly to actual restriction digest data. If the same fragments result from the in-silico and real digests, the finisher can be reasonably sure that the assembly is correct.

Navigator – the items in this menu allow a finisher to rapidly locate and view regions meeting the criteria specified by the navigator that is selected. This is particularly useful for locating regions that need to be worked on in order to complete the pre-submit checklist.

Oligo – short for oligonucleotide. Denotes a short, single strand sequence used as a primer in the sequencing process. Typically it is better to choose oligos ending in guanine or cytosine residues since they form more hydrogen bonds than adenine and thymine residues, thus binding the oligo more strongly to the template.

Pad (*) – a symbol that can be inserted in Consed that serves as a placeholder; it represents a space, not a base call. Often used to overwrite erroneously called bases or to correct minor alignment problems.

Paired end reads – the forward- and reverse-primed sequence reads that define the clone ends. Because the insert size is known to be ca. 2 kb, these sequences must occur within 2 kb of each other in the final assembly.

Phred 30 – sequences with Phred scores greater than 30 quality are deemed acceptable for the consensus sequence, even if the region is only covered by a single chemistry, strand, or subclone.

Phred/Phrap – base calling and assembly algorithm used to align reads prior to viewing in Consed. Its alignments are based on sequence similarity, and the position of forward/reverse read pairs.

Primer: Top-Stand Primer – A primer which is identical in sequence to the consensus sequence (top line in Aligned Reads Window), located to the left of the area of interest, which will produce sequence to the right of the indicated position in the Aligned Reads Window. The suggested sequence will be written in the primer box 5'-3'; it will be colored on the consensus sequence with the 3' base tagged in red.

Primer: Bottom-Strand Primer – A primer which is identical to sequence to the complementary strand of the consensus sequence, located to the right of the area of interest, which will produce sequence to the left of the indicated position in the Aligned Reads Window. The suggested sequence will be written in the primer box 5'-3', and will be the complement of the consensus sequence; it will be colored on the consensus sequence with the 3' base tagged in red.

Qstat file – a graphical representation of reaction statistics that allows for easy viewing of reaction failure rates and average read length. This allows one to calculate depth of coverage.

Read – the data from a single sequencing reaction is called a read. Reads are compiled into an assembly and can be viewed individually in Consed's Aligned Reads window.

Single-stranded region – a region that has sequence data for only one strand of the DNA (one direction as viewed in Consed).

SSR – simple sequence repeat, such as ACACACACAC.

Stealing reads – the act taking reads from an overlapping project to resolve gaps located in a similar area, or the act of removing reads from known repeat areas of an assembled fosmid, possibly for use elsewhere.

Subclone – similar to a clone, but instead of containing the entire length of the DNA of interest, it contains only a fragment; for sequencing we use subclones with approximately 2 kb of subject DNA. This fragment is sequenced from both ends, creating forward/reverse reads.

Supercontig – contigs that are lined up in order, often by using forward/reverse read pairs or other mapping information.

Tag – a label applied to regions that the finisher needs to mark for subsequent navigation. Tags commonly indicate locations of oligos, repeat regions, or edited bases.

Tear – the act of breaking one contig into two new contigs during the assembly process.

Trace – view of actual data collected by the sequencing machine. Traces are particularly helpful when editing problem regions, as they allow finishers to see base calls and to correct mistakes.

Vector – contains the fragment of genomic DNA that will be sequenced. Vectors are most often a type of plasmid, and have all the traits of a plasmid.

Vector sequence – any contaminating sequence from the cloning system.

Specialized glossary:

Drosophila melanogaster – the common fruit fly. It serves as an excellent model organism for studies in genetics and developmental biology. Its genome was finished in March of 2000, and has been updated several times since. This high quality sequence provides a reference for use in analyzing the genomes of other Drosophila species. The fourth, or 'dot' chromosome of *D. melanogaster* is generally considered heterochromatic, although it includes 82 genes.

Drosophila virilis – a relative of the well-known *D. melanogaster*; particularly interesting due to the euchromatic state of the dot chromosome, a region that is heterochromatic is *D. melanogaster*.

Euchromatin – less condensed regions of the genome containing the bulk of the genes.

Heterochromatin – the more condensed portion of genomic DNA, generally including highly repetitious sequences.