

RESTRICTION DIGESTS

Based on a handout originally available at

http://genome.wustl.edu/Overview/rst_digest_handout_20050127/restrictiondigest_Jan2005.html

What is a restriction digests?

Cloned DNA is cut with a restriction endonuclease (ex: *HindIII*, *EcoRV*, *EcoRI*, etc.). This generates fragments of different lengths, depending on the number of nucleotides between consecutive recognition sites. The fragments are then size-separated on an agarose gel and visualized with a DNA staining agent. The specific pattern of bands generated by digestion of a clone is referred to as a “fingerprint”.

The recognition sites at which these endonucleases cut the DNA are dictated by a specific base sequence which is different for each enzyme. These sites are usually palindromic. For example, *EcoRV* cuts any time 5' GATATC 3' is present in the sequence. Notice that the reverse complement of this sequence is exactly the same { 5' GATATC 3' }; hence the sequence is called palindromic.

When are restriction digests used?

-In mapping, digests are compared to identify overlapping clones, allowing one to create a physical map and to then choose a tiling path (a selected set of clones covering the region to be sequenced in which each clone has only a small amount of overlap between adjacent clones).

-To ensure that a finished project has been assembled correctly, its hypothetical (*in silico*) digest is compared with the real digest generated from purified cloned DNA. A matching set of fragments indicates that an assembly is correct, while discrepancies can indicate mis-assemblies or problematic regions.

-Digests can be used in analyzing unfinished projects to estimate gap sizes and/or orient contigs.

-Digest information is often required for areas that will be annotated in Genbank [to document unresolvable tandem repeats, SSRs (simple sequence repeats) , etc.].

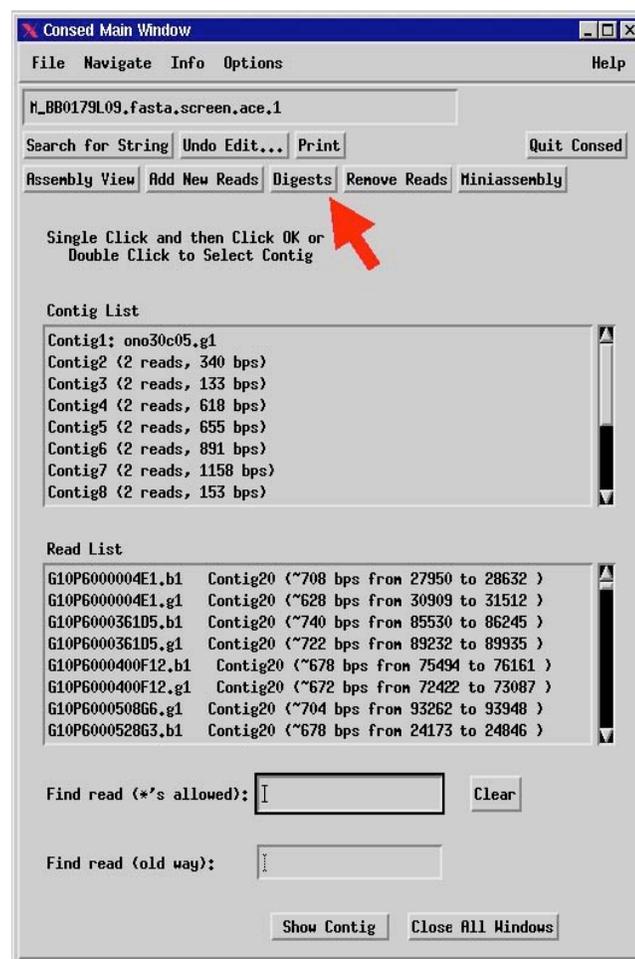
Checking the restriction digest using CONSED :

-A file, named fragSizes.txt, should be in the edit_dir of the project. If this file is not in your edit_dir you must obtain a copy before you can check your assembly. This file contains band calls (size estimates) from the real digests. The format of this file is shown below. It is a fasta format using enzyme headings with corresponding fragment sizes listed below that heading; the list ends with -1 to separate the results obtained using different enzymes. Thus, a file for a clone might look like this:

```
>EcoRV
150
3184
4599
8575
22571
-1
>HindIII
785
893
```

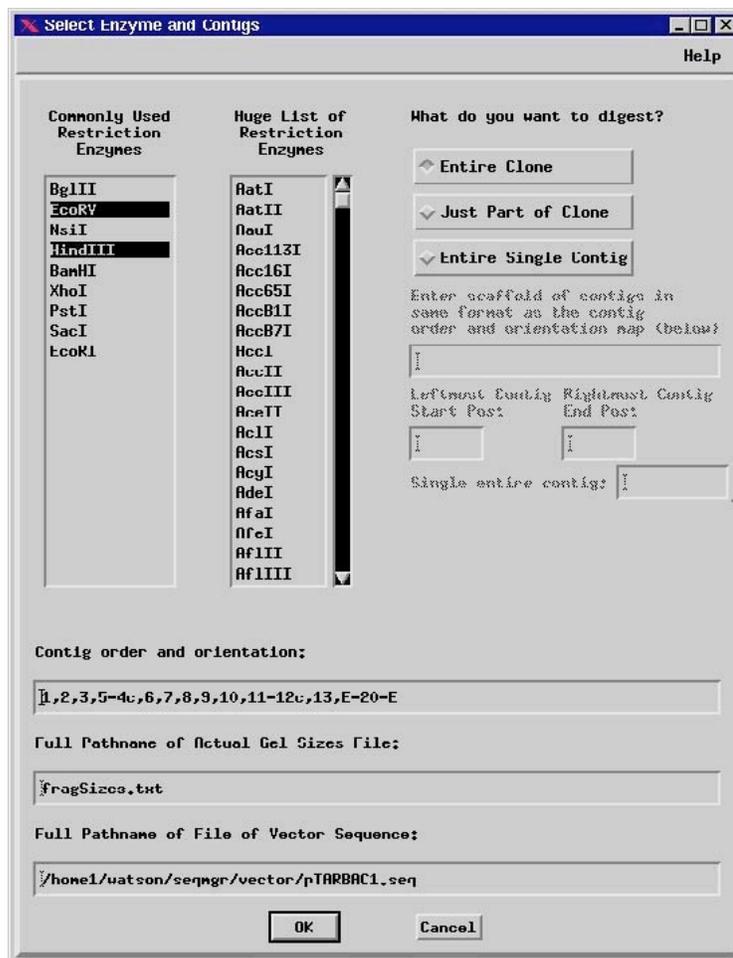
1516
2098
3402
3402
5466
7999
23402
-1

-To check whether your assembly matches the real digest, Consed should be opened, showing the ace file that needs to be checked. Once Consed is open, click on the **Digests** button on the Main Consed Window :



A new window will appear which allows the finisher to select several different options. The options include: which enzymes to use for the digest, which fragSizes.txt file to use, which contig/contigs to examine, and which vector sequence to include. Finishers can choose from an extensive list of enzymes by simply highlighting the ones they want. You must choose an enzyme for which there is data in the fragSizes.txt file. The standard set of enzymes includes *EcoRV*, *HindIII*, *SacI*, and *EcoRI*. The appropriate vector must also be indicated by putting in

the full pathname of the fasta file containing the vector sequence of your clone. The exact location of this file will depend on your local installation.



With this information, Consed selects a contig or orientation of contigs to be used to check the digest. C stands for “complemented” and E stands for “end of project.” See “Contig order and orientation” box. This is Consed’s best guess as to the order and orientation of the contigs. If the finisher is checking the digest after having completed an assembly, she would check the “Entire Clone” box. This technique will fail if Consed has difficulty detecting similarity between the assembly and the vector sequence. This failure will result in two extra “cut” sites being generated where the insert and vector join. In these cases select the “Entire Single Contig” and enter the contig number in the “Single entire contig” box.

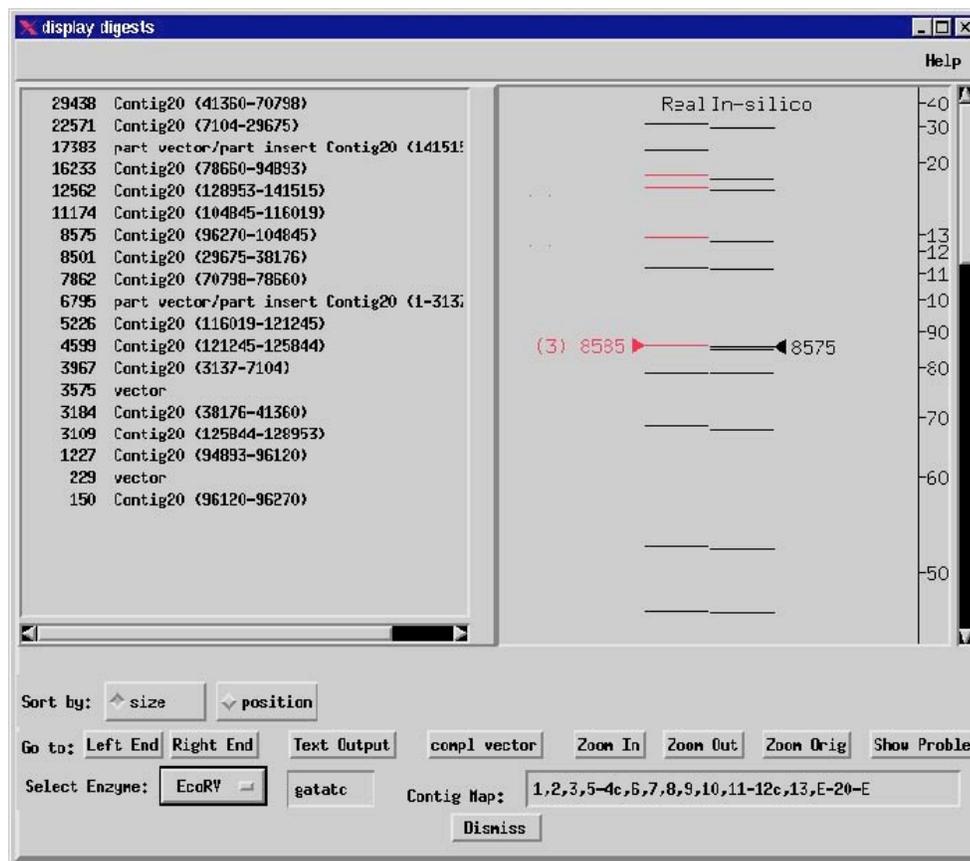
If the finisher wants to check a small region instead of the entire clone, the finisher can change the desired contigs under the “What do you want to digest?” section (top right of window). This could be the case if the finisher wanted to compare the real size of a band to a predicted band that contains an SSR to check for mis-assemblies, for example. If entering contig numbers under “What do you want to digest?,” the finisher need only enter contig numbers with dashes between them; the E which is listed under “Contig order and orientation” is not needed. Note that such specification is only possible if the appropriate option has been checked, and is not available under “entire clone.”

In all cases once the DNA sequence to be digested is selected the finisher would choose the enzymes from the two lists on the left. Most projects are digested with 4 enzymes, EcoRV, HindIII, EcoRI and SacI.

Choose “OK” once all the desired options have been selected.

If consed has difficulty determining the orientation of the insert with respect to the vector it will simply guess one orientation and give an error message that consed can't determine which end of vector is connected to the right end of contig Contig1 so you might need to complement the vector. If this happens click “OK” but be aware that you may need to have consed recalculate the fingerprint if it guessed incorrectly. If consed guessed incorrectly one or two bands will be wrong. These are the bands with one cut site in the vector and one cut site in the insert. To have consed calculate the *in silico* digest pattern for the opposite orientation click the “compl vecor” button.

A new window will appear showing the real digest compared to the *in silico* digest, with both text and graphical representations. The text portion shows the band sizes of the *in silico* digest. The graphical portion shows these bands compared to those of the real digest. Discrepancies (as identified by Consed) are shown with a red line. Purple bands indicate more than one band in the same general location and a “v” denotes a vector band. This window also has several useful options to choose from. “Sort by” allows the finisher to choose how they want to view the digest bands. Sorting by position makes it easier to find a specific location in the sequence that a finisher may want to have a closer look at. Other options on the window are self-explanatory.



There is some error in estimating the size of bands on the gels. This error can be quite substantial for very small and very large fragments. For bands between 1 kb and 10 kb in size you can consider an *in silico* band to match a real band if their sizes differ by less than 2%. For bands larger than this range larger differences are acceptable; this is left to the judgment of the finisher. (The error in size estimation increases as you get farther away from the optimal 1-10 kb range.) For bands smaller than 1 kb the size error can increase as well so values slightly

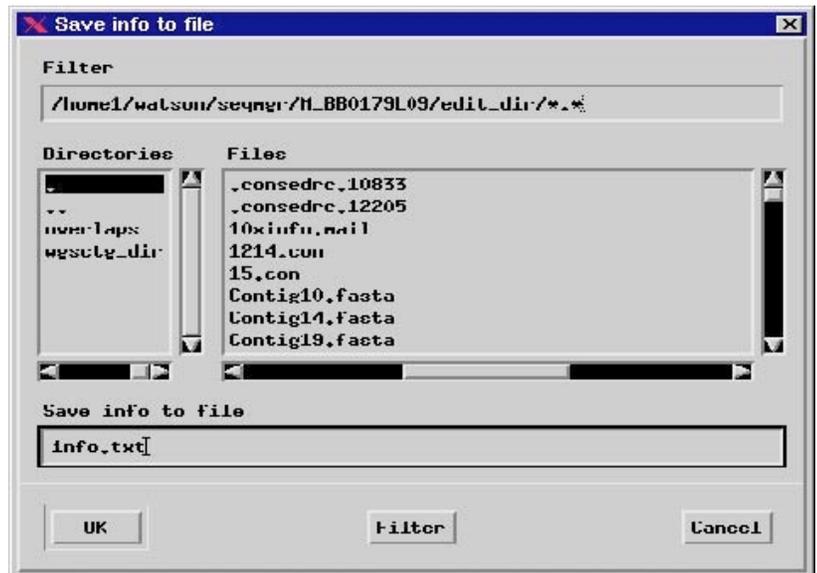
larger than 2% are acceptable. In addition, as bands get smaller than 1 kb there is increasing likelihood that the automatic band calling software will miss the band entirely (i.e. it will be too faint) thus it is acceptable for bands smaller than 1 kb to be missing from the “Real” lane. It is also possible that the real digest may miscall the number of bands that all migrate at the same rate. For example, in the figure above, in the analysis of the real digest the band at 8585 bp was estimated to be a triplet. The *in silico* digest suggests that it is a doublet. One would assume that the difference is due to a mis-estimate of the number of bands in the real digest and not that there was a mis-assembly.

Another common problem is the appearance of extra bands in the real fragment list. Some restriction enzymes do not completely digest the DNA and have a tendency to leave behind faint bands of incomplete digestion. Given that the calling of bands is automated it is not uncommon for the computer to mis-identify a faint band as a real band. You may be able to gather evidence that you have this problem if you know the total size of your clone (e.g. fosmids are always around 50 kb including vector). Simply add up the size of all the bands in the real digest and compare to the known total size of the clone. If the total size of the real digest is larger than the known size of the clone, you have uncovered evidence that you have spurious bands in your real digest.

Choose “Text Output”; this will bring up a new window that allows the finisher to save the digest information as a text file. Once the desired output has been achieved, choose “save to file.”

Real Frag Size	In Silico Size	In Silico Position
31224	29438	Contig20 (41360-70798)
22952	22571	Contig20 (7104-29675)
17901	17383	part vector/part insert Contig20 (141515-153710)
17901		
16471	16233	Contig20 (78660-94893)
16471		
12805	12562	Contig20 (128953 141515)
12805		
11260	11174	Contig20 (104845-116019)
8585	8575	Contig20 (96270-104845)
8585	8501	Contig20 (29675 38176)
8585		
7880	7862	Contig20 (70798-78660)
6874	6795	part vector/part insert Contig20 (1-3137)
5255	5226	Contig20 (116019-121245)
4610	4599	Contig20 (121245-125844)
4007	3967	Contig20 (3137-7104)
3592	3575	vector

This will bring up another window that allows the finisher to save this output in the desired directory.



The above settings would save the digest information in a file called “info.txt” in the edit_dir of the current project.