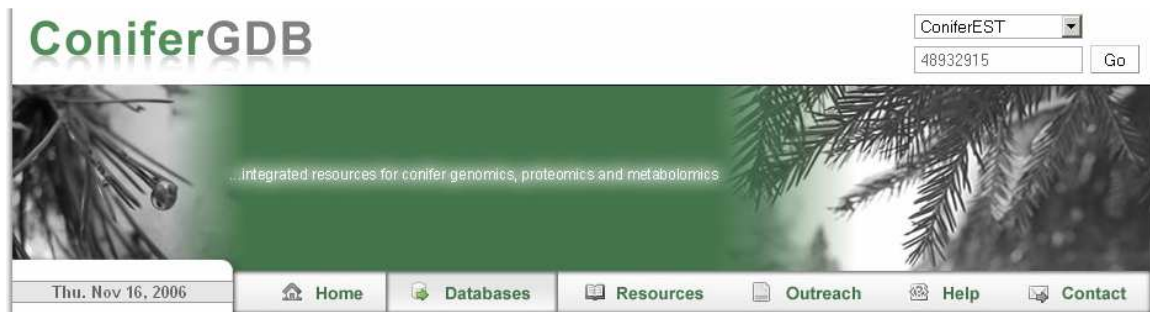


ConiferGDB Frequently Asked Questions

[1] How can I retrieve a single sequence with sequence name, GenBank accession or gi number?

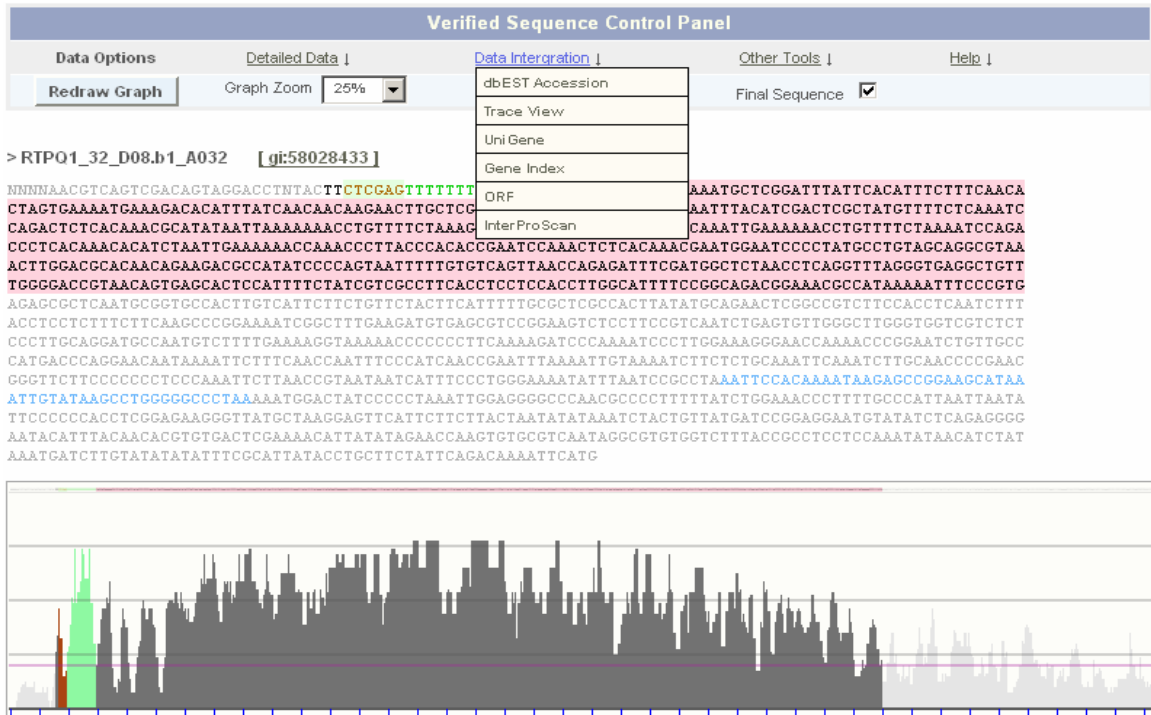
In ConiferGDB homepage, users first select the ConiferEST option within the pull-down menu shown in the top portion of the website. Users then enter either the specific sequence name (e.g. FLD1_34_H08.g1_A029), GenBank Accession (e.g., CO162374), or GenBank gi number (e.g., 48932915) and click the Go button.



[2] How can I retrieve a group of sequences with partial sequence name, GenBank accession or gi number?

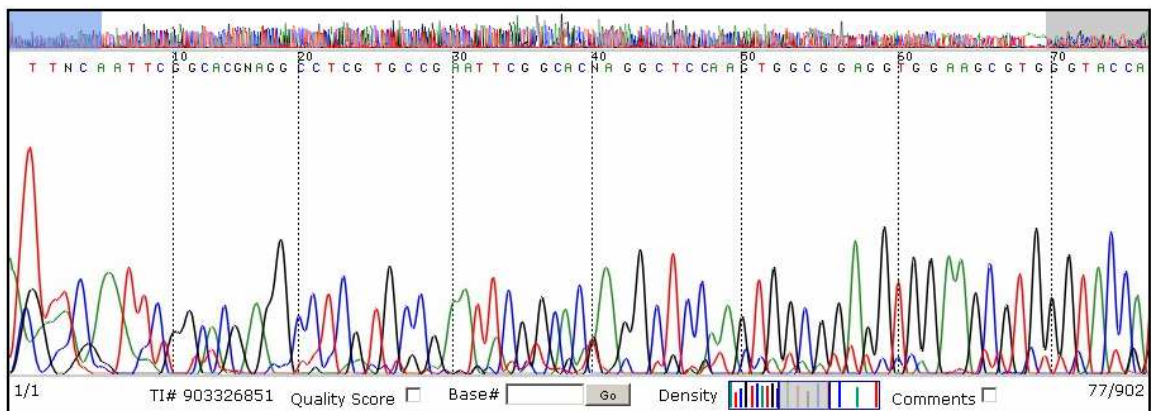
In above example, you can type “FLD1”, “FLD1_34” or any other partial sequence name to get a group of sequences which share the common partial name. You can also use partial GenBank accession (e.g. CO1623 instead of CO162374) or gi number (e.g. 4893 instead of 48932915) to get a group of sequences.

[3] How can I obtain ORF and EST cluster information for a specific sequence?



ConiferEST seamlessly integrates with other public EST resources such as NCBI ORF Finder and UniGene, as well as Gene Index. After users obtain a sequence with *in-silico* verified features (e.g. the 3' termini of a cDNA insert), they can click “Data Integration” within the *Verified Sequence Control Panel* and then select relevant menu items such as “ORF”, “UniGene” and “Gene Index” to detect open reading frames and retrieve related EST cluster information.

[4] How can I view the original chromatograph file for a given sequence?



ConiferEST seamlessly integrates with other public EST resources such as NCBI Trace Archive. After users obtain a sequence with *in-silico* verified features (e.g. the 3' termini of a cDNA insert), then users can see the original trace files deposited in NCBI Trace

Archive by clicking “Trace View” shown in the figure for the question [3]. ConiferEST is aimed to be a bridge between NCBI dbEST and NCBI Trace Archive. Obviously, your targeted sequences need to be deposited already in GenBank Trace Archive before viewing.

[5] [Have you assigned any protein domain signatures for your sequences?](#)

We have developed a high-throughput Perl program that identifies the best ORF for each sequence read having the expected *in-silico* verified sequence features. The filtration criteria include: (1) for all ESTs whose ORF start position is greater than 3, the ORFs must start with a start codon, and this requirement is waived for all ESTs whose ORF start position is less than or equal to 3; (2) for all ESTs, the ORF stop position should end with a stop codon; and (3) if there are multiple ORFs in a given sequence read that pass the first two criteria, the program will automatically pick the one with the maximum length. Subsequently, each ORF is scanned by InterProScan (version 4.2) to obtain protein domain signatures. In order to gather more information, we included all InterPro member databases - UniProt, PROSITE, Pfam, PRINTS, ProDom, SMART, TIGRFAMS, PIRSF, SUPERFAMILY, Gene3D and PANTHER for scanning. In current ConiferEST release (Build 1.0), we have conducted InterProScan for 43,857 sequences with *in-silico* verified features. Among all the 43847 peptides we obtained, 27,104 retrieved InterProScan annotation results, 19,218 retrieved InterPro entries, 17,290 retrieved Pfam domains, and 14,001 retrieved GO term annotations.

[6] [Is there more integration available in ConiferEST in the future?](#)

As a public community resource, we strive to provide various data integration for biologists. If you have any suggestion or comment, please send your email to Dr. Chun Liang (liangc@muohio.edu).