

GenomeBuilder

Integrating external applications and databases using CORBA
Juha Muilu, Patricia Rodriguez-Tomé, Alan Robinson
EMBL-EBI, Hinxton, Cambridge CB10 1SD, UK
muilu@ebi.ac.uk

Expressed sequence tags (ESTs) are short (200-500bp) DNA sequences generated from the 3' and 5' ends of cDNA clones selected at random. The purpose of EST sequencing is to scan rapidly all the expressed mRNA and hence to provide a "tag" for an expressed gene. Due to the fast sequencing process, the number of sequenced EST's is increasing rapidly. Approximately a thousand new entries are submitted each day to the publicly available EST sequence databases.

The EST databases contain a potential wealth of valuable information about expressed genes, e.g. their different splice variants and expression profiles. A problem with collections of EST sequences is that these data are highly redundant and often low in quality. The redundancy can be exploited to correct possible sequencing errors, but it can also make homology searches difficult. The informatics of this large volume of data is also not trivial.

In this poster we present a Java tool, called GenomeBuilder, which is designed for the visualization of EST and other high-throughput sequence clusters and assemblies. The tool has the following features:

- 1) Properties of the sequences can be colour-coded or shown as a text on the sequences.
- 2) Regions of similarity between the sequences can be highlighted
- 3) Linked windows mean an overall view of the assembly can be maintained in one window while working at a higher resolution in another.
- 4) Sequences in an assembly can be edited and moved

The EBI is investigating the use of CORBA (Common Object Request Broker Architecture) to resolve some of the current IT problems in bioinformatics. Distributed computing using CORBA provides a standard means to connect computational resources on different machines over the Internet. Computer applications can access these resources if they implement the programming interface to a CORBA server (specified using the Interface Definition Language (IDL)).

Currently, CORBA interfaces to EST sequence [1], EST cluster (JESAM [2]) and radiation hybrid databases (RHdb [3]) are implemented and available in GenomeBuilder. New databases can be added dynamically to GenomeBuilder if they provide an IDL interface and a CORBA server.

Within GenomeBuilder, the AppLab [4] program is used to generate client and server side components of external applications. Currently, two applications have been added to GenomeBuilder using AppLab: CAP3 [5] for making multiple alignments and CLEANUP [6] for removing redundant sequences.

Version 1.0 of the GenomeBuilder code with documentation is available for download at <http://industry.ebi.ac.uk/~muilu/GBuilder/>.

References:

1. Muilu J - <http://industry.ebi.ac.uk/~muilu/EST>
2. Parsons J - <http://corba.ebi.ac.uk/EST>
3. Rodriguez-Tomé P, Lijnzaad P *Nucleic Acids Research* 27(1):115-8(1999) -<http://corba.ebi.ac.uk/RHdb>
4. Senger M - <http://industry.ebi.ac.uk/applab/>
5. Huang X, CAP3: A DNA Sequence Assembly Program, *submitted* - <http://genome.cs.mtu.edu/cap/cap3.html>
6. Pesole G, Grillo G, Liuni S *Comput Chem* 20(1), 141-4 (1996) - <ftp://area.ba.cnr.it/pub/software/Cleanup>