



Genomic Arabidopsis Resource Network

Issue 4

26th march 2001

The first closing date for applications in the BBSRC initiative Exploiting Genomics is 30 April 2001. If you are considering to submit an application which will be making use of GARNet services, you need a feasibility statement from the GARNet office. Please mail Ottoline Leyser at hm011@york.ac.uk to obtain this statement.

An enormous amount of data for functional genomics will be created from the different resources and services provided by GARNet. NASC will be handling most of the bioinformatics for GARNet. It is crucial for the success of GARNet and all of your research projects that these data are stored and made available in a way that allows easy access to them. Availability of only raw data is of limited use to the research community. Therefore, the intention of the NASC databases is to give you the possibility to carry out data analysis and to make the right data analysis tools available giving you the information you require. NASC appreciates your feedback on what you require from these databases.

NASC aims to link its databases with those of other functional genomics projects, and the combination of data created by different projects will make the available information much more valuable for worldwide research progress. This puts further demands on database design: the different databases need to be compatible. If databases are linked, the transition between the different databases needs to be seamless.

This month the NASC database designers, Jamie Kincaid and Keith Bradnam, are giving you an overview of the bioinformatics activities that are going on at NASC with the aim of providing you with the information you need.

[Karin van de Sande](#)

Current bioinformatics activities at NASC

GARNet microarray database

A development version of the NASC transcriptomics project database (NTP) has been set up using MySQL database software running on a redhat 7 linux server. The database uses the same schema as the EBI's proposed public repository for microarray data - [ArrayExpress](#), achieved by using [maxdSQL](#) - a relational database implementation of ArrayExpress designed by the [Microarray Bioinformatics group](#) at Manchester University.

A java based data loading tool - [maxdLoad](#) has also been implemented and is being tested using data from AFGC microarray experiments held in the [Stanford Microarray database](#). The software provides a simple and intuitive user interface and will also provide us with an idea of the time / disk space / processor power needed for loading experimental data from an 11K array. This will then help us avoid bottlenecks when loading data from a 25K array.

Sequenced insert tools

NASC holds stocks and sequence information for several sequence tagged populations such as the SLAT lines. Several tools have been developed both as part of the GARNet project and in existing programmes in order to enable researchers to get the most out of the information available.

1. [Insert BLAST server](#)

Allows you to blast search a sequence against all Arabidopsis inserts and order from the catalogue the appropriate insert-containing lines.

2. [InsertWatch](#)

A service where users submit sequences to NASC that as well as being searched against all currently available Arabidopsis inserts will be held in confidence and searched against new insert lines when they become available. Users receive email notification of new insert stocks matching their sequence of interest and are provided direct links into the catalogue to order the relevant lines.

3. [Database of insert sequence homologies](#)

Insert sequences are routinely searched against all Arabidopsis sequences as part of the regular Blast analysis underpinning the Arabidopsis Genome Resource (AGR) - a part of UK CropNet. The resulting database of matches can be searched by a phrase, i.e. 'auxin', 'expansin' etc or by an EMBL/Genbank accession number, and the appropriate stocks ordered directly from the catalogue.

New UK CropNet database

[UK CropNet](#) have recently launched a new database (to complement the five [existing CropNet databases](#)). This new database is [CropSeqDB](#) - a sequence database for UK crop plants, and has been created for two main reasons:

1. To act as a repository of sequence information for UK crop species
2. To complement our existing Arabidopsis database ([AGR](#)) by linking sequences in CropSeqDB to their likely Arabidopsis homologue in AGR.

As a UK crop plant database, CropSeqDB only contains information for species detailed in the MAFF National Listing of UK crop species (which includes cereals, vegetables, oil & fibre plants, and fodder crops).

CropSeqDB currently contains nearly half a million sequences (taken from the EMBL database) representing over 150 species. Likely Arabidopsis homologues have been identified for about 70% of these sequences.

CropSeqDB [can be searched](#) by EMBL accession, gene name, gene product, etc. and all sequences with homologies are linked directly to protein entries in the AGR database. A [BLAST server](#) is also provided.

[Jamie Kincaid](#) and Keith Bradnam.

[back to welcome page](#) | [back to top of page](#)

Website maintained by [Karin van de Sande](#). | Please mail me your comments or suggestions.