

GeneGram User Guide

Martin Goodfellow, martin.goodfellow@cis.strath.ac.uk,
<http://genegram.cis.strath.ac.uk>

GeneGram provides a basic user interface for mapping microarray probes to genomes. It offers the mapping of probes to the rat, mouse and human genome. The mapping process consists of building the index; storing the probes and implementing and running the searches. GeneGram is distributed under the GNU General Public License and completely written in Java. It depends on an Oracle 10 or 11g database. SQL was used via JDBC to communicate with the database as well as PL/SQL which was run as a command line argument from within the application. Configuration details for connecting to the database as well as queries specific to the genome being used are defined in XML configuration files.

1 Configuration

Before starting the software the XML configuration files must be edited to include the URL of the Oracle 10 or 11g database the software has to be used with. Additionally a username and password has to be included for an Oracle account with sufficient privileges to create and modify the created tables.

```
<database>
  <url> url </url>
  <database-name> database-name </database-name>
  <username> username </username>
  <password> password </password>
</database>
```

On starting the application a genome must be selected: rat, mouse or human from the drop down menu provided. This can be seen in Figure 1

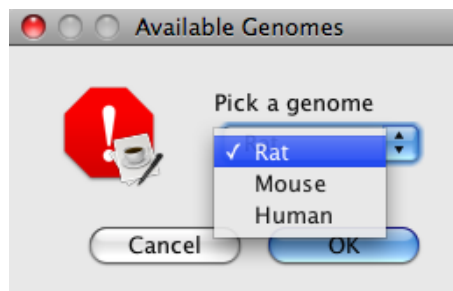


Figure 1: Selecting a Genome

2 Building the Index

To build the index the location of the chromosomes must be specified i.e. the directory that contains them all. These chromosomes must be in their uncompressed form. The destination of the control and .dat files for loading the data into the database must also be specified. This can be seen in Figure 2.

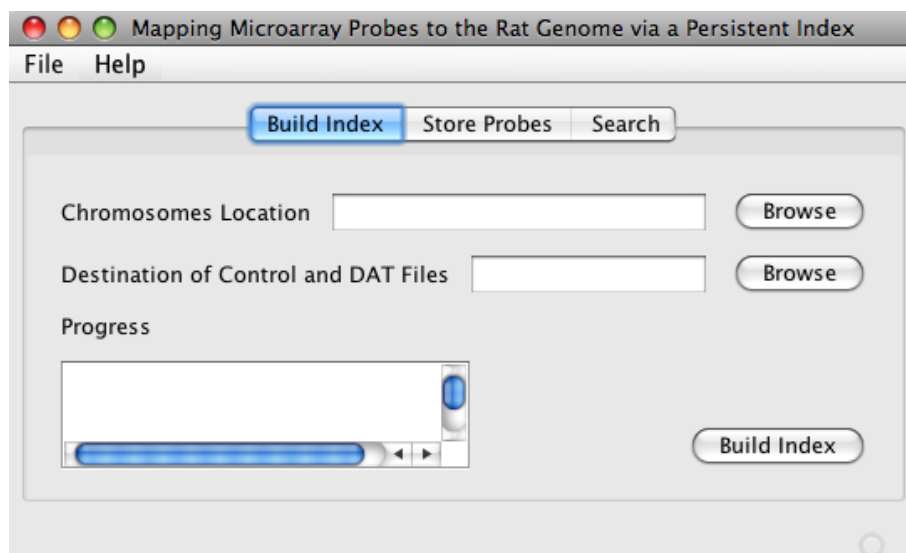


Figure 2: Building the Index

3 Storing the Probes

To store the probes the location of the probes must be specified i.e. the directory that contains them all. These probes must be in their uncompressed form. The destination of the control and .dat files for loading the data into the database must also be specified. This can be seen in Figure 3.

4 Implementing and Running the Searches

Users can select a search to perform from one of four options: 19ex, 19k1, 19k3, 16ex. This can be seen in Figure 4. These searches correspond to the match definitions we use in relation to the 25 base probe sequences. Our four scenarios are as follows. The first three scenarios focus on the central 19 bases, allowing the three bases at either end of the probe to mismatch. In the first definition, 19ex, the central 19 bases are matched exactly. In 19k1 we allow at most one mismatch in central 19 bases, and in 19k3 we allow at most three mismatches. The fourth definition, 16ex, differs from the others, and requires that 16 contiguous bases match exactly anywhere within the 25 bp probe. Multiple 16ex matches within one 25bp probe are treated as one match. BLAST matches to the probe can be either contained in the probe, and called a ‘contained match’ or partially overlap the probe, which constitutes a ‘discounted match’. A visual representation of these searches can be seen in Figure 5.

The results of the searches are stored within the database. They are stored in the following tables:

- 19ex - Perfect
- 19k1 - Onemismatch
- 19k3 - Mismatches
- 16ex - Sixteen

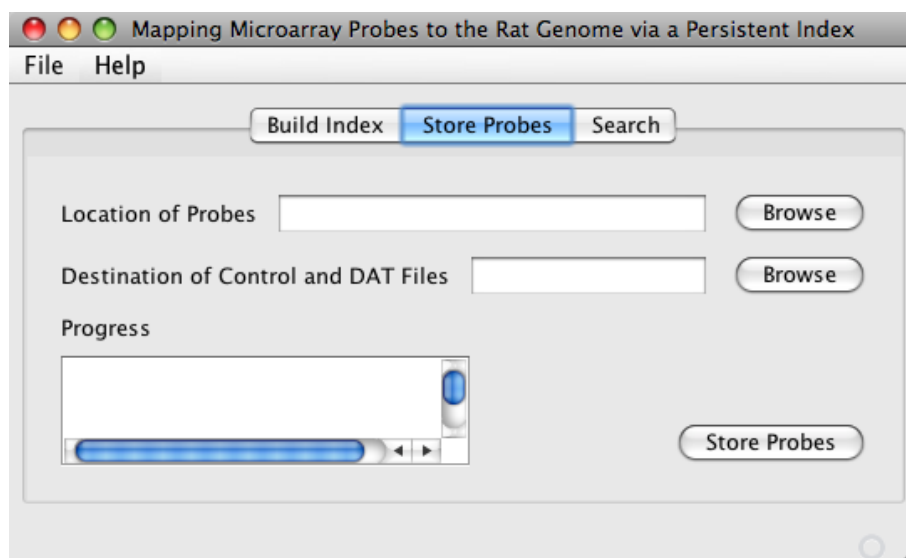


Figure 3: Storing the Probes

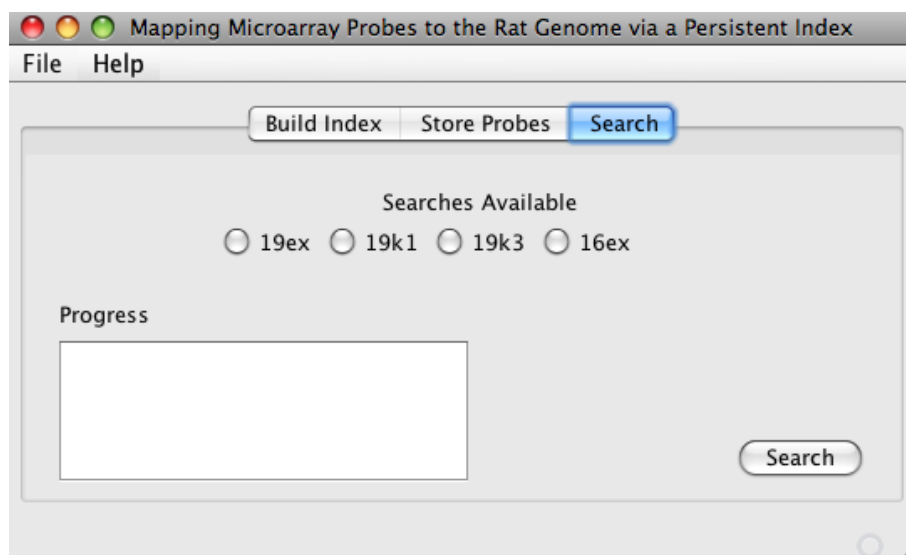


Figure 4: Implementing and Running the Searches

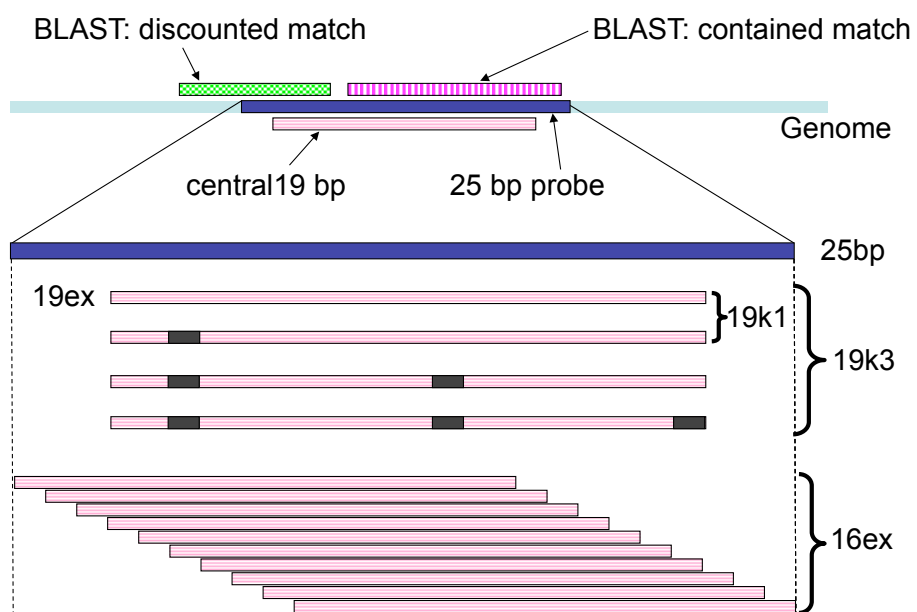


Figure 5: Search Definitions