HOWTO - Running a pipeline from the command line using `bigou.pl`

XML vs. Database

Once a pipeline has been designed and EGene's configuration file is saved, the pipelines are ready to be run. To run a pipeline from the command line, we use the Perl program `bigou.pl`. Pipelines can be run in XML mode or database mode. In XML mode all information is transmitted from one step of the pipeline to the other in XML format. This format can also be used to store all the processing performed into a set of sequences, generating in fact a flatfile database. In database mode, all information produced by the pipeline is stored in a database. This database has to be created before the pipeline is run.

To run a pipeline in XML mode, we should type the following command:

```
bigou.pl -c <configuration_file_name>
```

To run a pipeline in DB mode we should type the command

```
bigou.pl -c <configuration_file_name> -d <database_name> -h
<host_name> -u <database_user> -p <database_password>
```


Where:
- `database_name:` is the name of the database.
- `host_name:` is the computer hosting the database.
- `database_user:` is the user name under which information should be stored in the database.
- `database_password:` is the user's password to access to the database.

Running in parallel or sequential mode

EGene allows the use to run pipelines in two distinct modes. In parallel mode, `bigou.pl` runs the steps of the pipeline as independent processes communicating using Unix pipes. This means that in a multiprocessor machine we will have parallel execution of some of the components. The degree of parallelism will depend on how many processors the machine contains and on the scheduling policy of the operating system.

Pipelines in EGene can also be run in sequential mode, one component at a time, generating intermediary files to store the XML of the pipeline after each step. The file will have the same name as the one specified for that step in the "PHASE" parameter. If the pipeline is running in database mode, these intermediary files will contain only the identifiers of the sequences being processed. To run pipelines in sequential mode, we have to add the command line option –m (many files option):

```
bigou.pl -c <configuration_file_name> -m
```

Additional parameters for `bigou.pl`:

- `-?` – prints a help screen.
- `-v` – prints the version of `bigou.pl`.