Database Generation & Primer Searching Tool Documentation

Pre-requisites

1. *Python3* - This MUST be installed in the computer/machine.

2. *primer-searching-tool.zip* package provided.

   > NOTE: This package conatins two files listed below.

3. *generate_database.py* - A python program for generating database.

4. *search_primers.py* - A python program for searching primers.

Steps to set python environment

Launch terminal in Linux PC (#05-01)

1. Open terminal and type the following command to set python environment.

   ```
   > source ~/Documents/python-env/ml-env/bin/activate
   ```

2. Next, navigate to the folder where `generate_database.py/search_primers.py` file is present.

   > NOTE: These files are present in the downloaded package.

Database generation tool

Files used for database generation

1. `*.xls` - An excel file containing gene annotations.

   > NOTE: If annotations are available in *.gff file format, it can be converted to `*.xls` file manually.
   > The conversion just requires opening the `*.gff` file in Microsoft Excel and saving it in `*.xls`
   > format.
   >
   > For example, take a look at `Hamer_annotations_v1-final.gff` which is converted to
   > `Hamer_annotations_v1-final.xls` file.

2. `*.fa` - A FASTA file containing genome sequence.

   > NOTE: An example of FASTA file is `Hamericanus_final_cds_2021.fa`.

Steps to follow

1. In the terminal type the following command and press enter.

```
> python generate_database.py <path to xls file> <path to fa file>
```

> NOTE: An example of the same is as shown below:

```
> python generate_database.py Hamer_annotations_v1-final.xls
Hamericanus_final_cds_2021.fa
```

2. A file named `output.json` will be generated which needs be used in the next primer searching tool.

Primer Searching tool

Files used for primer searching tool

1. `output.json` (database file generated from above explained tool).

Steps to follow

1. In the terminal type the following command and press enter.

```
> python search_primers.py <path to json file>
```

> NOTE: An example of the same is as shown below:

```
> python search_primers.py output.json
```

2. Enter gene ID in next step and press enter.

> NOTE: Gene ID can be found from `output.json` or `*.xls` file.

3. Enter gene name and press enter.

> NOTE: Gene name can be found from `output.json` or `*.xls` file.

4. Three `*csv` files will be generated named as allgc, onegc and nogc which contains primer sequences.
   - `allgc.csv` primers with GC at 3' end in both forward and reverse primers.
   - `onegc.csv` primers with GC at 3'end in one of the forward or reverse primers.
   - `nog.csv` primers with no GC at 3'end.