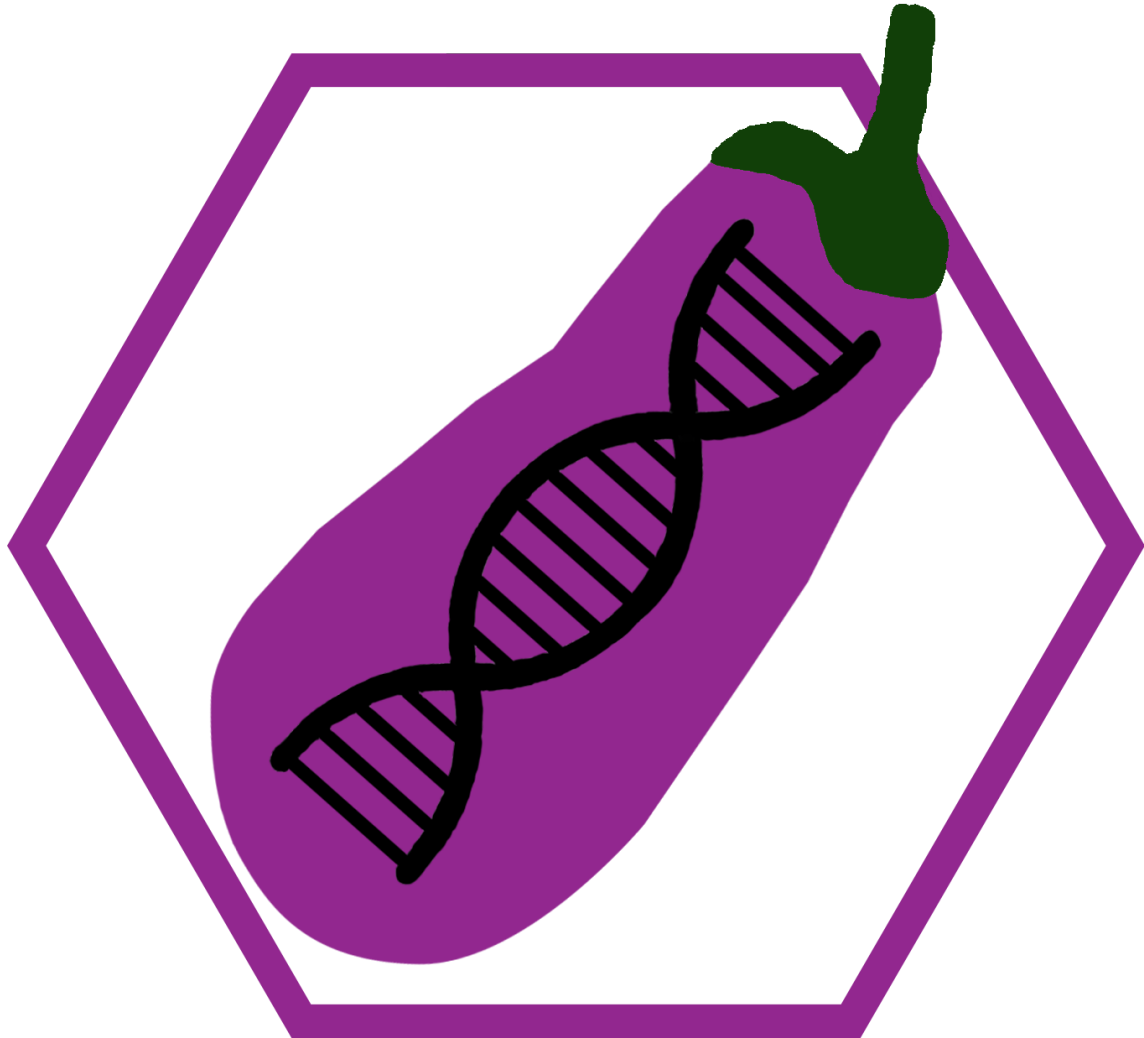


EGgPLant V1.0

User manual



EGgPLant has been developed by Morgan Ellis and Owen Holland.

For any issues please contact either of the above. Please note: For the most up to date readme see github.com/mrellis96/EGgPLant

Requirements:

Linux OS (including WSL on Windows) or Mac OS

IMPORTANT: EGgPLant will download and install the needed programs (R when using EGPL or EGPLQ; NCBI Blast+, NCBI Entrez Direct and Vim when using EGDB) on both systems if it is not present. For Linux OS this will involve adding to and updating the apt list and then installing the programs so the sudoer password will be required. For Mac OS this will involve installing homebrewer if it is not already present and installing the programs through homebrewer

Contents:

| | |
|---------------------------------------|---|
| Installation: | 3 |
| eggqual - Quality Plot Generator..... | 3 |
| eggplant - Main Pipeline:..... | 4 |
| eggdb - Blast DB Creator: | 5 |
| Updates: | 5 |
| DISCLAIMER | 6 |

Installation:

EGgPLant is available at <https://github.com/mrellis96/EGgPLant>.

Download the tarball (EGgPLant_VXX.tar.gz) and the install script install.eggplant.sh

run `bash install.eggplant.sh` in the same directory as the EGgPLant tarball. The install script will now install all the necessary programs and move the EGgPLant script to /usr/bin/ (/usr/local/bin/ on mac)

Manual install:

- 1) Extract files
- 2) Go to the folder containing the script file and run:

```
export PATH=$PATH:~/Path/To/Scripts
```

eggqual - Quality Plot Generator:

- 1) Go to the directory where you would like to run the pipeline
- 2) Ensure your fastq.gz are in a subdirectory called "RawFastq" with in the current directory
- 3)run:

```
eggqual -d [path/to/reads]
```

-d: Directory, where the raw reads are stored - REQUIRED

NOTE: `eggqual -h` will display the help file

eggplant - Main Pipeline:

1) Go to the directory where you would like to run the pipeline

2) run:

```
eggplant -d [] -f [] -r [] -l [] -p [] -t [] -y [] -n [] -m [] -b [] -v[] -o []
```

- d Directory - The directory where the raw reads are stored
- f Forward primer - The sequence of the forward primer for cutadapt to remove
- r Reverse primer - The sequence of the reverse primer for cutadapt to remove
- l Minimum overlap - The minimum number of base pair overlap between the primer and the returned sequence before cutadapt will recognise the primer
- p Poly tails - Tells cutadapt to remove poly tails. Add letter for base pair to remove
- t Truncate forward reads - Where you would like filterAndTrim to truncate the sequences (default = no truncation). Must be positive integer
- y Truncate reverse reads - IF PAIRED END, where you would like filterAndTrim to truncate the reverse reads (default = no truncation). Must be positive integer
- n Minimum read length - Remove reads with length less than [n] (default = 20)
- m Minimum number of reads - Minimum number of reads per sequences allowed after chimera removal (default = 10). Must be positive integer
- b BLAST database - Database to BLAST output sequences against. Must be BLASTn formatted database (see eggdb)
- v Taxonomic map - Path to taxonomic mapping file for Blast Database to pass to LCA script
- o Cluster ASVs to OTUs based on similarity (0-1) (default = no clustering)

3) Check number of reads after each step in terminal for optimising these variables.

NOTE: `eggplant -h` will display the help file

`eggplant -c` will display the citation file

eggdb - Blast DB Creator:

1) Go to the directory where you would like to run the database created

2)run

```
eggdb -q [ ] -d [ ]
```

-q: Query, Enter the desired query to search. To get the correct syntax for the query, go to the NCBI site and perform the desired search. Then copy the query from the "Search details" box. - It is recommended to search broadly for your gene region of your marker. The query must be enclosed with quotation marks. If you have double quotation marks WITHIN the query the query MUST be enclosed with SINGLE quotation marks - REQUIRED

-d: Database Name, The name for the database - NOTE: Spaces are not accepted. - REQUIRED

3) For blast to find the database the path must be exported

(export `BLASTDB=/path/to/database/`). This can be added to your bashrc automatically by the script or run at the start of each session.

NOTE: `eggdb -h` will display the help file

Updates:

V0.1.1: Added cancel option to Paired End Question; Changed README to .txt file; Added disclaimer; Added EGPLQ to run qPlots before Main Pipeline; Script now checks for Forward Directory, if present will skip sorting step; Fixed loop to Paired End Question; Script will now check if needed programs are installed and if not, install them.

V0.1.1.1: Bug Fixes, Script will check for reverse reads before making reverse/.

V0.1.1.2: Set repository to <https://cran.csiro.au/>, Added text for transparency and debugging.

V0.1.2: Added support for Mac OS, Quality plots are now an aggregate of the whole data set rather than the first 9 sequences.

V0.2.0: Added EGDB to create databases from queries, altered the way the script searches for packages so it is more reliable on both OS systems.

V0.2.1: Added Help messages to EGPL, EGPLQ, EGDB. Added Install script to automatically install script.

V0.2.2: Added -d option for EGPL and EGPLQ to specify the directory the raw reads are stored in, fixed various spelling and grammatical errors.

V0.2.3: Added -o option to EGPL to cluster ASVs into OTUs. Minor Fixes to install script.

V0.2.3.1: Fixed errors in README. Change repository to ubuntu version used

V0.2.3.2: Moved ubuntu version check so it no longer interferes with the installation of R, Changed Vim script to awk script in EGDB, Added '--relabel OTU_' to EGPL, Moved updates to end of README, added '(Y/N/C)' to EGPL.Inst.sh.

V0.2.3.3: Fixed errors with EGPL.inst.sh.

V0.2.4: Fixed issues with R library paths, removed redundancies in EGPL, EGPLQ, EGDB created from EGPL.Inst.sh.

V0.2.5 Added option [-n] to EGPL, Fixed bug that caused mac pipelines to error if option [-o] is added, Made 10 the default for option [-m].

V0.3.0: EGPL now uses 'Cutadapt' to remove primers instead of dada2:filterAndTrim, ASV's now added to Pipeline_results.csv instead of sequences. ASV's automatically converted to OTU's and stored in OTU.csv.

V0.3.1: Readded truncation to pipeline, added Poly tail remover, added vSearch citations

V0.3.1.1: Fixed numerous errors preventing r from running, added "then" to cutadapt script, fixed minor errors in cutadapt script, fixed errors in README

V0.4.0: Added Interactive Pipeline (EGIP)

V0.4.0.1: Improved the way EGIP and EGPL pass options to RScript to simplify R code and reduce chance of errors

V0.4.1: Added EGDB and BlastN to EGIP, R now outputs fasta file as 'Results.fa'.

V0.4.1.1: Name changed to EGgPLant

V0.4.2: Changed install script to install to a local dir. not /usr/bin, Added LCA script, renamed EGPL to eggplant.

V0.5.0: Integrated blast and LCA script in to main eggplant pipeline, depreciated EGIP, improved speed through awk handling.

V1.0: Set public release

DISCLAIMER:

While all effort was taken to ensure this script is free from errors, Morgan Ellis, Owen Holland and the EcoGenetics Lab take no responsibility to any Damage, corruption or issue that may arise from running this script or pipeline. Please use at own risk.