#### \*\*\*\*\*

IDXL is a python based tool for computing species tree from a set of incongruent gene trees with Incomplete Lineage Sorting (ILS). Following measures between individual couplets are compurted for species tree estimation.

A) Internode count between individual couplets (proposed in NJst (Liu et. al. 2011))B) Excess gene leaves count between individual couplets, computed for all the input gene trees.

These measures are used to form respective distance matrices, which are subsequently used for NJ based species tree construction.

Description

-----

Input

-----

A collection of gene trees (formed by sampling individual genes from a group of taxa) with overlapping taxa set, having topological incongruence due to Incomplete Lineage Sorting (ILS). Gene trees may or may not be weighted; our species tree estimation method does not consider the branch lengths of the tree.

Input gene trees can be either in NEWICK format or in NEXUS format. However, all the gene trees should have identical input formats. They should be placed in a standard tree list file, according to the syntax of NEXUS or NEWICK formats. Such a tree list text file is to be provided as an input of this executable.

Output

-----

A species tree covering all the taxa of the gene trees. Output species tree is generated in the NEWICK format.

#### Method implemented

IDXL: Here, relative internode count and relative excess gene count measures are checked for all the couplets. A couplet showing minimum value of the product of these two measures is selected for NJ like agglomeration.

#### \*\*\*\*\*

Dependencies

\*

These methods are developed in Linux Systems (Ubuntu 16.04), using Python 2.7. It is tested and meant for systems having linux OS (Fedora / Ubuntu).

Note: Ubuntu with lower OS version may have problem in executing the binary executable. So, users are adviced to execute it in latest Ubuntu or Fedora versions.

\*\*\*\*\*

Execution of the package \*\*\*\*\*\*\*\*

Step 1: Unpack the IDXL.zip file.

Step 2: User should check inside the directory IDXL, a stand alone executable of the same name will be found.

Step 3: User should provide executable permission of the package, with the following command: chmod +x IDXL

(assuming that user is within the folder 'IDXL')

Step 4: Now the executable can be run with the following command line options:

./IDXL [options]

\_\_\_\_\_

Details of the options are mentioned below:

-h, --help show this help message and exit

-I INP FILENAME, --INPFILE=INP FILENAME

Name of the input file containing gene trees. The file should be in either newick or nexus format. This is a mandatory option.

-O OUT\_FILENAME, --OUTFILE=OUT\_FILENAME

Name of the output file to contain target species tree. This option is not mandatory. If the output filename is not specified, the output species tree will be stored in a folder named 'IDXL\_OUT' which will be placed at the same directory containing the input treelist file.

-p INP\_FILE\_FORMAT, --inpform=INP\_FILE\_FORMAT

1 - input file format is NEWICK (default)

2 - input file format is NEXUS

-r TAXON\_NAME, --ROOT=TAXON\_NAME

User can specify a taxon name which would be used to root the output species tree (outgroup based rooting).

Example of a command (followed for the results published in the manuscript)

### A) ./IDXL -I input.txt -p1

command descriptions:

1) -I specifies the input filename

2) input.txt : contains the input gene treelist

3) -p option is for specifying the input treelist file format. Here it is NEWICK since 1 is provided as the input.

## B) ./IDXL -I inp\_nexus.txt -p2 -O 'outtree.tre'

Here, a nexus file named 'inp\_nexus.txt' contains the input gene trees. The option '-p2' stands for the nexus file input. In addition, output species tree will be stored in the filename 'outtree.tre'.

# \*\*\*\*\*

Sourya Bhattacharyya Department of Computer Science and Engineering Indian Institute of Technology Kharagpur <sourya.bhatta@gmail.com>