Catfish User Reference

Mingfu Shao[*1] and Carl Kingsford[†1]

[1]Computational Biology Department, School of Computer Science, Carnegie Mellon University

October 21, 2016

1   Installation

To install Catfish, you need to first download Boost library, and then compile the source code of Catfish.

Download Boost from `http://www.boost.org`. Uncompress it somewhere (compiling and installing are not necessary). Set environment variable `BOOST_HOME` to indicate the directory of Boost. For example, for Unix platforms, add the following statement to the file `~/.bash_profile`:

```
export BOOST_HOME="/directory/to/your/boost/boost_1_60_0"
```

Get the source code of Catfish through `git`:

```
$git clone git@github.com:shaomingfu/catfish.git .
```

Execute the following commands to generate `Makefile` and compile:

```
$cd src
$aclocal
$autoconf
$autoheader
$automake -a
$./configure
$make
```

The executable file `catfish` will be present at `src/src`. You might want to link it into `bin` through

```
$cd bin
$ln -sf ../src/src/catfish .
```

2   Command line

The usage of Catfish is as follows:

```
$./catfish -i input.sgr/input.gtf -o output.out [-a algo]
```

`-i` parameter specifies the input file. Catfish accepts two types of input file formats. The first one is `.sgr`, which specifies a directed acyclic graph. The first line of the file gives n, indicating the number of vertices

---

[*]mingfu.shao@cs.cmu.edu

[†]carlk@cs.cmu.edu

in the graph. The vertices are named from 0 to $n-1$, where vertex 0 has to be the source vertex and vertex $n-1$ has to be the sink vertex. Each of the following line specifies an edge, which consists of three integers: the in-vertex, out-vertex and the weight of this edge. The second input file format is `.gtf`. If it is this file format, Catfish will merge all transcripts for each gene into a splice graph, and then try to decompose it.

`-o` parameter specifies the output file, which will show the predicted paths and their associated abundances.

`-a` parameter specifies the algorithm. There are three options: `full`, `core`, and `greedy`. With option of `core`, the program will only run the core algorithm to partly decompose the given splice graph, which will predict fewer paths but with higher accuracy. With option of `full`, the program will completely decompose the given splice graph, using greedy algorithm following the core part of the algorithm. With option of `greedy`, the program will only use greedy algorithm to fully decompose the given splice graph. This parameter is optional, and its default value is `full`.