Shun's de Bruijn Graph Assembler

This is a de Bruijn graph genome assembler made for BIO 365. It is written for Python3.

Usage

Usage for the assembler can be found by using the `-h` or `--help` flag like the following:

```
$ python3 SHUN-dBGA.py --help
```

This will print out the basic usage to the terminal.

```
Usage: SHUN-dBGA.py [options]

Options:
  -h, --help            show this help message and exit
  -f FILENAME, --file=FILENAME
                        the fasta file to read from
  -k K, --kmer-length=K
                        the value of k for kmers; the default value is k = 10
  -e, --handle-errors   handles errors if this flag is on
  -p, --plot-histogram  will create an png image of the kmer frequencies if
                        this flag is on (the system running the script must
                        have matplotlib installed to be able to use this
                        flag); must be used in conjunction with the -e flag
  -v, --verbose         if this flag is on, will print out metrics to the
                        console
```

The only required option is `-f` of `--file` to specify the input.

Example

```
$ python3 SHUN-dBGA.py -f input/real.error.small.fasta -k 20 -e
```

Output

Running the script will output all generated contigs and metrics into the `results/` directory.

For example, running:

```
$ python3 SHUN-dBGA.py -f input/real.error.small.fasta -k 20 -e
```

will output the contigs in `results/real.error.small.contigs.k.20.fasta` and the metrics can be found in `results/real.error.small.metrics.tsv`.

If you run the assembler with the `-p` flag, the generated image will be outputted in the `coverage/` directory.

For example, running the assembler with the same flags as above will save the image as `coverage/real.error.small.coverage.k.20.png`.