Software installation and execution instructions

Installation:

1. Un-jar the jar file.
2. Create an Eclipse workspace.
3. Create a Java project called “operonsproj” in the workspace.
4. Into the project, import dirs “bin”, “data”, “src”, and “lib”. You can easily do this by selecting all 4 dirs and dragging them into the icon for the Java workspace.
5. Download the opencsv jar from <https://sourceforge.net/projects/opencsv/>, and import it into lib.
6. Download WEKA from <https://waikato.github.io/weka-wiki/downloading_weka/>. Import the WEKA jar into lib.
7. Download the Apache commons-math binaries jar from <https://commons.apache.org/proper/commons-math/download_math.cgi>. Import the jar into lib.
8. In Project 🡪 Properties 🡪 Java Build Path, add the OpenCSV, WEKA and Apache jars to the classpath as external jars.
9. Import the followingdata files into the “data” dir (apologies for not providing them, but your author is not authorized to redistribute them):
   1. The GenBank nucleotide record for your organism. The article uses *Crocosphaera watsonii* WH8501, available at https://www.ncbi.nlm.nih.gov/nuccore/AADV02000001.1/
   2. Operon predictions for your organism. Download from <http://www.microbesonline.org/operons/>
   3. Time-series expression data for your organism, in comma-separated values (CSV) format. After any header lines, each line should contain expression data for one gene. The article uses data from a study by Shi et al. (our reference 26).
10. Create a simple text file that describes the time-series file that you imported in (9c) above. This file, in tab-separated values (TSV) format, describes the structure of the time-series csv file. The descriptor for the *Crocosphaera* study is included in the release for reference (data/Croco\_columns.csv). Each line has 2 fields.
    * 1. The first line should say, “SKIP<tab>N”, where N is the number of header and blank lines to be skipped at the start of the time-series csv file.
      2. Each remaining line should describe one time point of the study. The format is “Column#<tab>hours”, where “hours” gives the number of elapsed hours at the time point. For example, in the *Crocosphaera*  study CSV file, column 7 gives expression measurement at the 3-hour time point, so its line is “7<tab>3”.

Execution:

* Find the main method at the bottom of source file Experiment.java in the operons package.
* Edit the 4 lines “File xxx = new File(…);”, substituting the pathnames of your 4 files. These files should be in the “data” folder as described above.
* To evaluate WEKA classifiers, call exper.rankClassifiers();
* After a classifier has been chosen, call
  + Classifier classifier = new ClassifierClassname();
  + File otsv = new File(“ … “); // report written to here
  + exper.evaluateMergeCandidatesWithLMT(classifier, otsv);