Installation instructions for the software described and used in

Hosseini S-R, Barve A, Wagner A. Exhaustive analysis of a genotype space comprising 1015 central carbon metabolisms reveals an organization conducive to metabolic innovation. PLoS Comput Biol. 2015;11:e1004329.

If you use this software, please cite the above publication.

The associated github repository (https://github.com/rzgar/EMETNET) contains required material to reproduce the exhaustive analysis of central carbon metabolism genotype space in [1]. Please note that we supply this software for free, and will regretfully not be able to provide installation support, technical support, or bug fixes.

You first need to install the software according to the instructions given in the README.md file (https://github.com/rzgar/EMETNET/blob/master/README.md). The folder CCM contains the file universe.net, which includes the relevant 72 central carbon metabolism reactions. The folder CCM/ENVS contains ten files that define the chemical environments used in [1]. Finally the folder CCM/BLOCKS contains five different "block files" that play an important role in the exhaustive enumeration. Their use is described in chapter two of [2], which is available from the author upon request, or electronically at (http://e-collection.library.ethz.ch/view/eth:7522?q=(keywords_en:PHENOTYPE)

The algorithm described in [2] proceeds in five steps described below. These steps require the three C++ programs contained in the CCM folder. The algorithm must be executed for each carbon source separately (Note: the file ENVS/C4.flx defines a glucose-containing environment, and is used as an example here). Before executing the code, it is useful to create a directory called VIABLES, and within it ten subdirectories, one for the genotypes viable on each carbon source. In the example below, VIABLES/C4/ will store the genotypes viable on glucose.

Step 1:
In this step, you will determine the set of viable genotypes from each of the 511 (2^9-1) genotypes corresponding to each block as follows:

```
for ((i=1;i<6;i++));do metnet-exgen ~/universe.net ~/BLOCKS/block$i.dat ~/ENVS/C4.flx ~/VIABLES/C4/viables_$i.dat;done
```

As a result, five output files storing genotypes viable on glucose will be generated.

Step 2:
In this step, you will merge pairs of files containing viable genotypes created in the previous step. You must use the program "metnet-split-merge-viability.cpp", which requires two files from step 1 and the number of lines (viable genotypes) in each of them as command-line input. Execute the following command for all 10 pairs of files from step 1:

```
metnet-split-merge-viability ~/universe.net ~/C4/viables_1.dat ~/C4/viables_2.dat  ~/C4/viables_1_2.dat ~/ENVS/C4.flx --s1 NUM1 --s2 NUM2
```

Note: In this example, NUM1 is the number of lines in ~/C4/viables_1.dat, and NUM2 is the number of lines in ~/C4/viables_2.dat.

Step 3:
This step creates a merged triplet of files containing viable genotypes (a single file generated in step 1 with a merged pair generated in step 2) as follows. Execute the following command for all 10 possible triplets of files:
metnet-split-merge-viability ~/universe.net ~/C4/viables_1.dat ~/C4/viables_2_3.dat ~/C4/viables_1_2_3.dat ~/ENVS/C4.flx --s1 NUM1 --s2 NUM2
Note: In this example, NUM1 is the number of lines in ~/C4/viables_1.dat, and NUM2 is the number of lines in ~/C4/viables_1_2.dat.

Step 4:
This step creates merged quadruplets of viable files (a single file generated in step 1 with a merged triplet generated in step 3).
Execute the following command for all 5 possible quadruplets of files:
metnet-split-merge-viability ~/universe.net ~/C4/viables_1.dat ~/C4/viables_2_3_4.dat ~/C4/viables_1_2_3_4.dat ~/ENVS/C4.flx --s1 NUM1 --s2 NUM2
Note: In this example, NUM1 is the number of lines in ~/C4/viables_1.dat, and NUM2 is the number of lines in ~/C4/viables_1_2_3.dat.

Step 5:
The number of viable genotypes resulting from step 4 is very large (more than 10^8), which can result in storage and performance problems when the output data is analyzed. To alleviate such problems, you can use the program "metnet-split-merge-memoryfriendly-viability.cpp". This step creates a merged quintuplet of files containing viable genotypes (using a single file generated in step 1 with a merged quadruplet generated in step 4). If the number of viable genotypes in ~/C4/viables_1.dat is given by NUM1, you need to execute the following command:

for ((i=1;i<NUM1;i++));do **metnet-split-memoryfriendly-merge-viability** ~/universe.net ~/C4/viables_1.dat ~/C4/viables_2_3_4_5.dat ~/C4/viables_1_2_3_4_5_$i.dat ~/ENVS/C4.flx --s1 $i;done

Note: Here, NUM1 is not passed to the program. Instead, in each of the (NUM1) iterations of the loop, a single genotype from ~/C4/viables_1.dat (indexed with $i) is merged with all viable genotypes in ~/C4/viables_2_3_4_5.dat.

Next, combine all viable genotypes in a single file:
cat ~/C4/viables*.dat>>~/C4/Final.dat

Subsequently, you can partition the resulting file based on the number of reactions as follows:
awk '(NF==10){print}' ~/C4/Final.dat>>~/C4/deletions_10.dat
so, ~/C4/deletions_10.dat will contain all the genotypes with 10 absent reactions that are still viable on glucose..

References:

1. Hosseini SR, Barve A, Wagner A. 2015 Exhaustive analysis of a genotype space comprising 1015 central carbon metabolisms reveals an organization conducive to metabolic innovation. Plos Comput Biol. 11, e1004329.
2. Hosseini S-R. Exhaustive genotype-phenotype mapping in metabolic genotype space. 2013, Zurich, Switzerland: Swiss Federal Instiute of Technology