User manual

Numericware i: Identical By State Matrix Calculator

Kim, Bongsong

1. For Windows users

Numericware i will be installed under "C:/Numericware/Numericware_i" by default. Thus, this folder is the default folder. You can change the path of the default folder during the installation.

2. For Linux users

(1) On Linux, the installation is not required. The directory including the following two files, Numericware_i and inverse, will be the default folder.

(2) Uncompress command:

tar –xvf Numericware_i.tar.gz

(3) Command to run Numericware i:

./Numericware_i

3. Example files

The default folder includes three example files:

(1) alphanumeric.txt
(2) snp.csv
(3) iupac.csv

For practice, you can use Numericware i with any of the files shown above.

4. Data format

Numericware i can read data sets based on one of three delimiters (comma, tab, space), and write the resulting IBS matrix using one of two delimiters (comma, tab).

Numericware i supports three types of data points: alphanumeric, a pair of SNPs, and IUPAC types.

(1) Alphanumeric type: users can define three alphanumeric scores for three allelic types (e.g. 0, 1, and 2 for AA, Aa, and aa), and missing value (e.g. NA).

(2) A pair of SNPs type: 16 SNP pairs (AA, AT, AG, AC, CA, CT, CG, CC, GA, GT, GG, GC, TA, TT, TG, TC) can be recognized. The remaining values will be considered as missing value(s).

(3) IUPAC (International Union of Pure and Applied Chemistry) type: 10 IUPAC codes (A, C, G, T, R, Y, S, W, K, M) can be recognized. The remaining values will be considered as missing value(s). Each code can match one or two pair(s) of SNPs as follows (http://www.bioinformatics.org/sms2/iupac.html):

IUPAC	A pair of SNPs
A	AA
C	CC
G	GG
T	TT
R	AG or GA
Y	CT or TC
S	CG or GC
W	AT or TA
K	GT or TG
M	AC or CA

Numericware i requires marker names and entity IDs for column and row labels, respectively. The correct format of data is as follows:

	m1	m2	m3	m4
ID1				
ID2				
ID3				
ID4				

What if the marker names and IDs are laid out the other way around? No worry. Numericware i supports transposing the data. At the upper-left hand corner, any value in the cell will be ignored. You can leave it blank.

5. How to use Numericware i

Numericware i is really easy to use. You need not learn any syntax to use it. Giving Numericware i your job is simply a series of answering the questions about:

(1) Type in the path of your data file

Step1> Enter the path of the input file: **C:\Numericware\Numericware_i\alphanumeric.txt**

(2) *Choose a delimiter type*

Step1> What is the delimiter?

Step1> 1. Comma
Step1> 2. Space
Step1> 3. Tab
Step1> Please choose a number: **1**

(3) In case that it is necessary to transpose the data, transpose the data set

Step1> Do you want to create a transposed data? (Y/N): **Y**
Step1> Please enter a transposed file name: **t_alphanumeric.txt**
Step1> You can choose a delimiter between tab (T) and comma (C). Which do you want (T/C): **T**
Step1> You chose the tab-delimiter.
Step1> The transposed data can be found at "C:\Numericware\Numericware_i\t_alphanumeric.txt".
Step1> Numericware i goes back to the start.

(4) *Choose the genotype score format*

Step2> What is the genotype format?
Step2> 1. Alphanumeric
Step2> 2. A pair of SNPs
Step2> 3. IUPAC
Step2> Please choose a number: **1**

(5) *In case that scores for your data is alphanumeric, type in score(s) for missing value(s)*

Step2> Is there any score(s) for missing value(s)? (Y/N): **y**
Step2> Please enter the score(s) for missing value(s). If you have multiple scores, please use a space or tab as a separator: **na**

(6) *In case that scores for your data is alphanumeric, type in identical by state coefficients for every pair of scores*
Assuming that valid scores = 0, 1, and 2; and an invalid score = "na":

Step2> Please enter identical by state coefficients for each pair of scores:
Step2> Missing value(s): NA
Step2> 0 0: **2**
Step2> 0 1: **1**
Step2> 0 2: **0**
Step2> 1 1: **1**
Step2> 1 2: **1**
Step2> 2 2: **2**

In case that scores for your data is either the nucleotide pair or IUPAC, there is no need to enter the identical by descent coefficients manually. It will be automatically calculated.

(7) *Set a working directory*

Assuming that Numericware_i.exe file is placed under "C:\Numericware\Numericware_i" directory:

Step3> The default directory is "C:\Numericware\Numericware_i".
Step3> Numericware i requires to create a working directory under the default directory.
Step3> The working directory is the place where all resulting files will be saved.
Step3> Please create a name of your working directory: **dir_1**

(8) *Type in a resulting file name*

Step3> Please enter an outcome filename: **result.csv**
Step3> The resulting IBS matrix will be found at "C:\Numericware\Numericware_i\dir_1\result.csv".
Step3> You can choose a delimiter between tab (T) and comma (C). Which do you want (T/C): **C**
Step3> You chose the comma-delimiter.

(9) *Choose a genomic type that you want to analyze between haplotype and global genomes*

In case that an analysis for the global genome is wanted:

Step3> If you want to compute a haplotype IBS matrix, please enter 'H'. If you want to compute a global IBS matrix, please enter 'G': **g**

In case that an analysis for the haplotype delimited between m1 and m1000 for two markers is wanted:

Step3> If you want to compute a haplotype IBS matrix, please enter 'H'. If you want to compute a global IBS matrix, please enter 'G': **h**
Step3> Please input the first marker name: **V1**
Step3> Please input the second marker name: **V1000**

(10) *Set the number that you want to chop your data into*
Step4> Please enter a number for how many pieces you want to chop the data into: **20**

(11) Set the number of CPU cores that you want to use

Step5> How many CPUs do you want to use?: **7**

(12) *Decide whether inverting the IBS matrix will be conducted.*

Step5> Do you want to have the inverse of the IBS matrix (Y/N): **Y**

6. How to go back up to the previous step

As shown above, in order to give your job to Numericware i, you need answer to multiple questions. Occasionally you may want to change your previous answer. In this case, you can simply go back up by typing the step you want to be.

Suppose that you are now in Step 5. If you want go back up to Step 4, type "step4" into the blank as follows:

Step5> How many CPUs do you want to use?: **step4**

Then, all your previous answers given after Step 4 will be cleaned up, and you will see the first line of Step 4 as follows:

Step4> If your data size is huge, your computer memory will not able to accommodate the whole data.

Now you can start from Step 4.