***FINAL EXAM for IBC 2016 course, April 27, 2016***

In the INTRON2 workstation in my home directory (/home/afedorov) find a file “GEMA\_HAPLOTYPES50”. It is a table with 50 rows and 2003 columns. It is a modification of the SNP modeling table we worked during last week. You need to ignore first two columns and the last column. The rest 2000 columns represent SNP data for 2000 modeling human individuals. Each cell contains 0/0, 0/1, 1/0, or 1/1. Zero stands for a reference allele of the SNP and one stands for the mutant allele. Each person has two alleles –maternal (first) and paternal (second) which are separated by backslash ‘\’ character. The file looks like:

43 0 1\1 1\1 1\0 1\0 0\1 1\1 1\1 0\0 … (first row)

111 0 0\0 0\0 0\1 0\1 0\0 0\0 0\0 1\1 …

117 0 1\1 1\0 1\0 0\0 1\0 1\0 1\1 0\0 …

133 0 1\1 1\0 1\1 0\1 1\1 0\1 0\1 1\1…

179 0 1\0 1\1 1\0 1\1 1\1 1\1 1\1 1\0 …

.

.

.

1726 0 0\0 1\0 0\0 0\0 0\0 0\0 0\0 0\0 …(last row)

TASK: You need to write a Perl script that performs two steps:

* Task 1: remove all information about paternal alleles (second digit after ‘\’) and keep only maternal alleles. The file should look like:

43 0 1 1 1 1 0 1 1 0 … (first row)

111 0 0 0 0 0 0 0 0 1 …

117 0 1 1 1 0 1 1 1 0 …

133 0 1 1 1 0 1 0 0 1…

179 0 1 1 1 1 1 1 1 1 …

.

.

.

1726 0 0 1 0 0 0 0 0 0 …(last row)

* Task 2: For each modeling individual (each column) concatenate all alleles into a single string starting from the first row and ending by the last row. The output result should look like:

10111111101101110111010000011000010100100111001010 individual 1

10111011100000001000010111100000010110100111001001 individual 2

10111111011101110100000100100111101011111001101100 individual 3

10001011101011001111101100011111101011111001101000 individual 4

00111111101011001100010100100000010110111001001110 individual 5

10101000101011001100010100100000010110111001001110 individual 6

10101000010000000100000000101100001010000011101000 individual 7

01011011010000001010000100011000010100101001101000 individual 2000

You can use Internet, books, notebooks. HOWEVER, no conversations between students are allowed. Also, you CANNOT look into files of other students!