

About program

**hrom_reconstruction_GL – «gradient descent»
from multiple random initial points for task 3, stated in
Chromo_manual: reconstruction of chromosome structures, when
they are given in the leaves and absent in the inner nodes**

Program "hrom_reconstruction" reconstructs chromosome structures in the inner nodes of the given tree (not necessarily binary) using the given structures in the leaves of this tree. The corresponding algorithm is described in [1] and below in this manual.

Input files:

- file "input.tre" contains one string, where the tree is written in the bracket format. Its leaves are either enumerated from left to right using numbers 1, 2, 3, ... or marked by the names of corresponding species.
- file "input.chrom" contains chromosome structures in the leaves in the same order, in which they appear in the bracket format, that is, according to their enumeration. Structures are separated by strings, containing multiple symbols "*". Thus, structure of each chromosome occupies starts at the new line.

Chromosome is represented as the sequence of the names of genes, separated by space symbol. If gene is being transcribed from complement (minus-) strand, it has symbol "-" in front of its name. Chromosome description is followed by the letter "C", if chromosome is circular, and "L", if it is linear.

The last structure is followed by two consecutive strings. The first string contains several star symbols, the second – single symbol "#", which indicates the end of file.

The program is executed by means of running hrom_reconstruction.exe module. The result is printed to "output " file. The inner tree nodes are enumerated in depth-first search traversal order, when traversal starts from the tree root and it visits leaves in order of their enumeration. Each inner node is followed by the chromosome structure, that has been found by the algorithm.

Parameter t indicates the number of randomly chosen initial points. The value of this parameter was set to $t=31$ in the above examples.

The description of «gradient descent» algorithm

Note, that algorithm is fast and effective, if the initial point is chosen correctly, which itself is a difficult problem (see task 3a in Chromo_manual). Let us describe the algorithm of descent from one point and for one fixed value of parameter p . Variation of the initial point and this parameter requires multiprocessor system.

On each step we run through all the inner nodes of the tree and all the operations, applicable to this structure. We choose the pair: node-operation, which delivers minimal arrangement to our tree; then this operation is applied to the node and new arrangement is obtained. We repeat this step till the total price of the arrangement decreases.

The examples of computations can be found on this site: http://lab6.iitp.ru/ru/hrom_reconstruction/.

Input files. For example, the following data is taken for the artificial example:

1) file [input.tre](#) contains one string ((1,2),(3,4),(5,6));

It is a tree with six leaves;

2) file [input.chrom](#) contains initial chromosome structures for six leaves:

+g2+g3+g4+g5+g6 (C)

+g7 (C)

+g8+g9 (C)

+g1+g3+g4+g5+g6 (C)

+g7 (C)

+g8-g9 (C)

+g1+g2+g4+g5+g6 (C)

+g8 (C)

+g7+g9 (C)

+g1+g2+g3+g5+g6 (C)

+g8 (C)

+g7-g9 (C)

+g1+g2+g3+g4+g6 (C)

+g9 (C)

+g7+g8 (C)

+g1+g2+g3+g4+g5 (C)

+g9 (C)

+g7-g8 (C)

###

Every structure in this example contains three cycles, including 5, 1 or 2 genes. Structures differ from each other by gene composition, gene order: the second, fourth and sixth leaves one gene is located on the complement strand.

Output file [output](#) for this example is:

#leaves 6, #species 6, #genes 9

assignment cost 9

assignment itself:

node 1 with the composition of 9 genes:

+g1+g2+g3+g4+g5+g6 (C)

+g9 (C)

+g8 (C)

+g7 (C)

node 2 with the composition of 9 genes:

+g1+g2+g3+g4+g5+g6 (C)

+g9-g8 (C)

+g7 (C)

node 3 with the composition of 9 genes:

+g1+g2+g3+g4+g5+g6 (C)

+g9 (C)

+g8 (C)

+g7 (C)

node 4 with the composition of 9 genes:

+g1+g2+g3+g4+g5+g6 (C)

+g9 (C)

+g8-g7 (C)

Node 1 is the root of the tree, the node

2 is the ancestor of leaves 1 and 2, the

node 2 is ancestor of leaves 3 and 4,
the node 4 is ancestor of leaves 5 and
6.

Biological examples

Two biological examples are attached to the program.

First example contains [tree](#), chromosome structures description [chromosomes](#) and optimal arrangement [arrangement](#).

Second example also contains [tree](#), chromosome structures description [chromosomes](#) and optimal arrangement [arrangement](#).

Reference

[1] K.Yu. Gorbunov, R.A. Gershgorin, V.A. Lyubetsky. Rearrangement and Inference of Chromosome Structures. *Molecular Biology*, 2015, Vol. 49, No. 3, pp. 327–338.