Mgenome 1.0

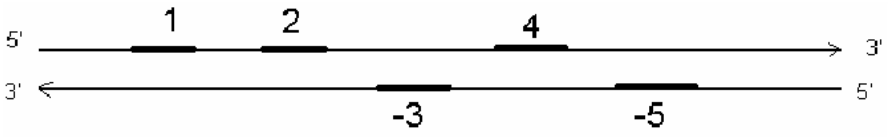
This program is designed to find the optimal trees for multiple genome rearrangement by signed reversals. The problem is modeled as: For a given collection of genomes represented by signed permutations on genes, find a tree to connect the given genomes by reversal paths such that the number of all signed reversals is minimized.

Definition A genome is represented by a signed permutation:

$$P=p_1\ p_2\ \cdots\ p_n$$

where each p_i stands for a gene located in a DNA. The sign (+/-) of pi denotes the strand of p_i , i.e., "+" indicates that p_i is in the forward strand of the DNA and "-" means that p_i is in the forward strand of the DNA.

Example Genome P= 1 2 -3 4 -5 in a DNA



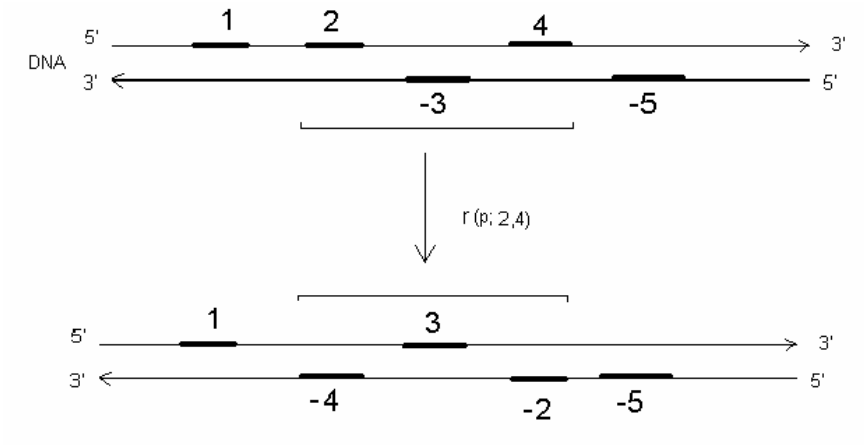
Definition A reversal is defined as an operation that transfers p into p(i,j)

$$p = p_1\ p_2\ \cdots\ p_{i-1}\ p_{i+1}\ p_{i+2}\ \cdots\ p_j\ p_{j+1}\ \cdots\ p_n$$

$$p(i,j) = p_1\ p_2\ \cdots\ p_{i-1}\ -p_j\ -p_{j-1}\ \cdots\ -p_i\ p_{j+1}\ \cdots\ p_n$$

The segment [i,j] of p changes its both orientation and signs.

Example The following signed reversal transfers P_1 = 2 -3 4 -5 into
P_2 = 1 -4 3 -2 -5:



Multiple Genome Rearrangement By Signed Reversal

Let $g_1, g_2, \ldots, g_m$ be a collection of genomes represented by signed permutations on genes. Find an optimal tree that connects the given genomes by reversal paths such that the number of all signed reversals is minimized.

Mgenome 1.0 is designed to find such an optimal tree for the given genomes.

The program accepts its input data stored in a file "nodes.dat", which is in FASTA format for all genomes.

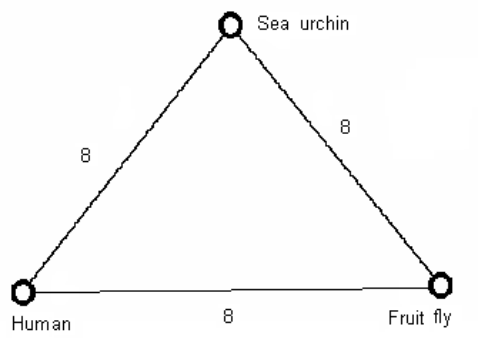
Example

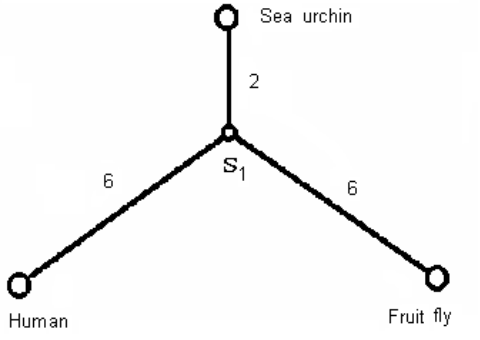
```
>Human
4  5  1  6  9  8  -2  3  -7  10
>Sea urchin
-2  -3  4  7  5  1  10  -8  6  9
>Fruit fly
-3  4  5  7  1  10  2  8  -9  6
```

Then the program find the optimal tree:

```
>a[0]: Human
4  5  1  6  9  8  -2  3  -7  10

>a[1]: Sea urchin
-2  -3  4  7  5  1  10  -8  6  9

>a[2]: Fruit fly
-3  4  5  7  1  10  2  8  -9  6

Distances:

   d[0,1]: 8
```

```
   d[0,2]: 8
   d[1,2]: 8

Internal node(s):

>s[1]:
-10  -1  -5  -7  -4  3  2  -8  -9  -6

  Reversal proccess:

a[0] - s[1]:

          4 5 1 6 9 8 -2 3 -7 10

R[ 1, 1]: -4 5 1 6 9 8 -2 3 -7 10

R[10, 1]: -10 7 -3 2 -8 -9 -6 -1 -5 4

R[ 3, 3]: -10 7 3 2 -8 -9 -6 -1 -5 4

R[ 9, 3]: -10 7 5 1 6 9 8 -2 -3 4

R[ 4, 2]: -10 -1 -5 -7 6 9 8 -2 -3 4

R[10, 5]: -10 -1 -5 -7 -4 3 2 -8 -9 -6

Reversal distance = 6

s[1] - a[1]:

          -10 -1 -5 -7 -4 3 2 -8 -9 -6

R[ 7, 1]: -2 -3 4 7 5 1 10 -8 -9 -6

R[10, 9]: -2 -3 4 7 5 1 10 -8 6 9

Reversal distance = 2

s[1] - a[2]:

          -10 -1 -5 -7 -4 3 2 -8 -9 -6

R[ 6, 1]: -3 4 7 5 1 10 2 -8 -9 -6

R[ 3, 3]: -3 4 -7 5 1 10 2 -8 -9 -6

R[ 4, 3]: -3 4 -5 7 1 10 2 -8 -9 -6

R[ 3, 3]: -3 4 5 7 1 10 2 -8 -9 -6

R[ 8, 8]: -3 4 5 7 1 10 2 8 -9 -6

R[10,10]: -3 4 5 7 1 10 2 8 -9 6

Reversal distance = 6


Total reversal distance = 14
```





Web-based Program The program is also designed as a web-based one and can be run on-line:

http://xgu.zool.iastate.edu

References

The program is implemented from the algorithms in the following papers:

1. S. Wu and X. Gu, Multiple genome rearrangement by reversals, Pacific Symposium on Biocomputing 7(2002)259-270.

2. S. Wu and X. Gu, Algorithms for multiple genome rearrangement by signed reversals, Pacific Symposium on Biocomputing, 8(2003), 363-374.