CTRD: a fast applet for computing signed translocation distance between genomes

WangSen Feng¹, Lusheng Wang²,* and Daming Zhu³

¹Department of Computer Science, Peking University, Beijing, 100871, P.R. China, ²Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong and ³School of Computer Science and Technology, Shandong University, Jinan, Shandong, P.R. China

Received on February 8, 2004; revised on April 12, 2004; accepted on May 27, 2004
Advance Access publication June 16, 2004

ABSTRACT
Summary: CTRD is a software for computing translocation distance between genomes. It takes two genomes as its input and tests whether one genome can be transformed into the other. If possible, it computes the translocation distance between two genomes, and gives the translocation operation serial. We adopt the fastest known $O(n^2 \log n)$ algorithm.

Our contributions include (1) give a necessary and sufficient condition to ensure that one genome can be transformed into the other for translocation operations, and (2) develop a software using the fastest known $O(n^2 \log n)$ algorithm.

Contact: cswangl@cityu.edu.hk

1 INTRODUCTION
A computational approach to evolutionary studies based on rearrangements was pioneered by Sankoff et al. (1990). A lot of work has been done in this field. For reversal distance, see Bafna and Pevzner (1995a); Berman and Hannenhalli (1996); Bader et al. (2001) and Hannenhalli and Pevzner (2004). For transposition distance problem, see Bafna and Pevzner (1995b, 1998). Hannenhalli (1997) first studied the signed translocation distance. Zhu and Ma (2002) gave the fastest algorithm that runs in $O(n^2 \log n)$ time. Given two signed genomes, it is not true that one genome can always be transformed into the other by using translocation operations. Previously, no one has seriously discussed the conditions that one signed genome can be transformed into the other for translocation operations. We give the necessary and sufficient conditions here and develop a software, CTRD, to compute the signed translocation distance between two genomes. We adopted the fastest known $O(n^2 \log n)$ algorithm of Zhu and Ma (2002). CTRD takes two genomes (expressed as signed integer sequences) as input, and tests whether the source genome A can be transformed into the destination genome B using translocation operations. If yes, it computes the translocation distance between the two given genomes and outputs the translocation serial.

2 PRELIMINARIES
A genome is a collection of chromosomes and a chromosome consists of a set of genes organized in a linear order. A gene is represented as a signed integer. A chromosome is a sequence of signed numbers. A genome is a set of sequences of signed numbers. For example, $A = [(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12), (13, 14, 15, 16, 17)]$ is a genome containing two chromosomes. The first chromosome has 12 genes: 1, 2, ..., 12, and the second has five genes: 13, 14, ..., 17.

Translocation is one of the most common rearrangement events in mammalian evolution, it exchanges genetic material between different chromosomes. A translocation acting on chromosomes $X = (X_1, X_2)$ and $Y = (Y_1, Y_2)$ swaps the segments in the chromosomes and results two new chromosomes. Here, we study the most common type of translocation, reciprocal translocation where each of the four segments, $X_1, X_2, Y_1$ and $Y_2$, is non-empty. A translocation is a prefix–prefix translocation, if the prefix of one chromosome is swapped with the prefix of the other chromosome and a translocation is a prefix–suffix translocation, if the prefix of one chromosome is swapped with the suffix of the other chromosome (Fig. 1).

Given two genomes, A and B, translocation distance between A and B is the minimum number of translocations required to transform A into B.

3 TESTING CONDITIONS
Here, we give the necessary and sufficient conditions that one genome can be transformed into the other.

THEOREM 1. Considering the source genome A and the destination genome B, A can be transformed into B by
translocation if and only if:

1. the two genomes contain the same set of genes;
2. the two genomes contain the same number (must be at least 2) of chromosomes;
3. the two genomes have the same set of ending (either head or tail) genes;
4. for any gene \( g \) that is an ending gene in \( A \), (1) if \( g \)'s sign in \( A \) is different from that in \( B \), then \( g \) must be a head in one genome and a tail in the other; (2) if \( g \) has the same sign in both \( A \) and \( B \), then \( g \) must be either a head in both genomes or a tail in both genomes.

The proof of the theorem can be found in http://www.cs.cityu.edu.hk/~lwang/software/Translocation/proof.pdf.

4 IMPLEMENTATION

CTRD is a Java Applet developed with JDK 1.4.2_02 and its graphical user interface (GUI) is created with JFC/Swing. When the user visit the Web page, it is downloaded onto user’s machine and executed in the user’s browser. We adopt the technique of Java Applet because it has two advantages: (1) compatible with multiple platforms (CTRD can run on most operating systems and Web browsers, including Unix and Windows, Netscape and IE) and (2) easily interact with users (GUI created with Swing is beautiful and user-friendly, all interactions are on client local computer without the need to communicate with the server).

CTRD accepts two genomes which are represented by integer sequences as its input. There are two input methods, the user can either directly input source genome A and destination genome B in the ‘Input Area’ by hand, (If the data is in a file, the user can copy it to the ‘Input Area’ as follows: (1) highlight the data and press ‘ctrl + C’ and (2) move the focus to the input area and press ‘ctrl V’) or click the ‘Examples’ button to get a dialog and choose one item in the list, and click ‘OK’ button to load the example.

After the two genomes are input, click the ‘Start’ button to begin the computation. Then CTRD checks the possibility on whether the source genome A can be transformed into the destination genome B by translocation operations based on Theorem 1. If yes, it will give the user the translocation distance between the two genomes and output the translocation operations. Users can copy the output result from the ‘Output Area’ to his or her own text editor. The intermediate steps of genome transformation can be shown in a separated window.

ACKNOWLEDGEMENT

This work is fully supported by a grant from the Research Grants Council of the Hong Kong SAR, China (Project No. CityU 1196/03E).

REFERENCES