JOURNAL OF COMPUTATIONAL BIOLOGY
Volume 17, Number 5, 2010

Pp. 685–705
DOI: 10.1089/cmb.2009.0025

Sorting by Reversals, Generalized Transpositions, and Translocations Using Permutation Groups

YEN-LIN HUANG[1] and CHIN LUNG LU[2]

ABSTRACT

In this article, we consider the problem of sorting a linear/circular, multi-chromosomal genome by reversals, block-interchanges (i.e., generalized transpositions), and translocations (including fusions and fissions) where the used operations can be weighted differently, which aims to find a sequence of reversal, block-interchange, and translocation operations such that the sum of these operation weights in the sequence is minimum. It is known that this sorting problem can be solved in polynomial time on the basis of breakpoint graphs, when block-interchanges are weighted 2 (or $\geq$3) and the others are weighted 1. In this study, we design a novel and easily implemented algorithm for this problem by utilizing the permutation group theory in algebra.

Key words: algebra, block-interchange, fission, fusion, generalized transposition, genome rearrangement, permutation group, reversal, translocation.

1. INTRODUCTION

GENOME REARRANGEMENT STUDIES based on genome-wide analysis of gene orders play an important role in the phylogenetic tree reconstruction (Sankoff et al., 1992; Hannenhalli and Pevzner, 1995, 1999; Pevzner and Tesler, 2003; Belda et al., 2005). In the studies of genome rearrangements, a gene is usually represented by a signed integer, where the associated sign indicates on which of the two complementary DNA strands the gene is located, a chromosome by a sequence of genes and a genome by a set of chromosomes. Given two genomes of the same set of genes, the *genome rearrangement problem* aims to compute a minimum sequence of rearrangement operations required to transform one genome into another. The operations used as rearrangement events within genomes with single chromosomes include reversals (Hannenhalli and Pevzner, 1999; Kaplan et al., 1999; Bader et al., 2001; Tannier et al., 2007), transpositions (Bafna and Pevzner, 1998; Elias and Hartman, 2005), and block-interchanges (Christie, 1996; Lin et al., 2005), where *reversals*, also called *inversions*, affect a block of consecutive integers in the chromosome by reversing the order and flipping the signs of the integers; *transpositions* affect two adjacent blocks in the chromosome by exchanging their positions; *block-interchanges* are *generalized transpositions* by allowing the exchanged blocks not being adjacent in the chromosome. In genomes with multiple chromosomes, the rearrangement operations include

[1]Department of Computer Science, National Tsing Hua University, Hsinchu, Taiwan.
[2]Institute of Bioinformatics and Systems Biology, Department of Biological Science and Technology, National Chiao Tung University, Hsinchu, Taiwan.

translocations (Hannenhalli and Pevzner, 1995; Hannenhalli, 1996; Bergeron et al., 2006a; Ozery-Flato and Shamir, 2006), fusions (Hannenhalli and Pevzner, 1995; Meidanis and Dias, 2001; Lu et al., 2006), and fissions (Hannenhalli and Pevzner, 1995; Meidanis and Dias, 2001; Lu et al., 2006), where *translocations* exchange the end segments between two chromosomes; *fusions* join two chromosomes into a bigger one; *fissions* break a chromosome into two smaller ones. Usually, the genome rearrangement problem is viewed as a problem of *sorting* one permutation into another using rearrangement operations, if the given genomes are represented by permutations with one having positive, sorted integers.

Meidanis and Dias (2000) were the first to relate the theory of permutation groups to the study of genome rearrangements, by demonstrating that many properties and simple operations related to permutation groups in algebra can be directly applied to model several commonly used rearrangement events that affect the permutations of representing genomes. Indeed, a few subsequent studies (Meidanis and Dias, 2001; Lin et al., 2005; Lu et al., 2006; Mira and Meidanis, 2007) have proved that the permutation groups seem to be a useful tool in the design of efficient algorithms for some genome rearrangement problems, especially involving reversals, transpositions, block-interchanges, fusions and/or fissions. To date, however, no algorithmic research has been done on how to apply permutation groups to the genome rearrangement problems with translocations being involved.

In this study, we focus on the problem of sorting a signed permutation by reversals, generalized transpositions (or, equivalently, block-interchanges here) and translocations (including fusions and fissions), and discuss the design of its efficient algorithms by the utilization of permutation groups. The complexity of this problem is still unknown so far, if all of the used rearrangement operations are assigned the same weight (hereinafter, designated as the *unweighted* case). In real biological data, however, transpositions, acting as a special case of block-interchanges on a genome, occur with about half the frequency of reversals (Blanchette et al., 1996). Moreover, Eriksen (2002) used simulations to find that the optimal weights for reversals and transpositions (including inverted transpositions) are 1 and 2, respectively, reflecting that the optimal reversal in the generic case will remove one breakpoint, while the optimal transposition removes two breakpoints. Therefore, it seems to be biologically meaningful to assign at least twice the weight to block-interchanges than to the others. In this differently weighted case, the problem consists in finding a sequence of rearrangement operations such that the sum of the operation weights in the sequence is minimum. If block-interchanges are at least three times the weight of reversals, this weighted genome rearrangement problem then becomes that of sorting a signed permutation by reversals and translocations (including fusions and fissions), which is now solvable in polynomial time (Hannenhalli and Pevzner, 1995), because a block-interchange can be mimicked by three reversals (e.g., three consecutive blocks [*A, B, C*] can be transformed into [*C*, *B*, *A*] by a block-interchange, as well as by three reversals with scenario of [*A*, $-C$, $-B$], [*C*, $-A$, $-B$], and [*C*, *B*, *A*]), and as a result, there is always an optimal solution for the problem that contains nothing but only reversals and translocations. When block-interchanges are weighted 2 and the others are weighted 1, the weighted genome rearrangement problem can still be solved in polynomial time on the basis of breakpoint graphs (Yancopoulos et al., 2005). In this article, we present a novel and easily implemented algorithm for this weighted genome rearrangement problem by the application of permutation groups. Notably, if the block-interchanges in this problem are restricted to only ordinary transpositions, Eriksen (2002) has proposed a $(1+\varepsilon)$-approximation algorithm, although the complexity of finding an exact solution for the problem is still unknown so far.

Recently, Alekseyev and Pevzner (2008; Alekseyev, 2008) introduced a more general rearrangement model by defining a new and more powerful operation called *multi-break operation* (or simply *k-break*) acting on a breakpoint graph. Given two genomes, say *P* (the initial genome) and *Q* (the target genome), for a genome rearrangement problem, their *breakpoint graph* is an edge-colored graph *G*(*P*, *Q*) defined as follows: (1) Each gene is represented by two vertices in *G*(*P*, *Q*) that denote the two ends of the gene and are labelled as tail and head, respectively, where the direction from tail to head corresponds to the sign (strand) of the gene. (2) There is a black (respectively, gray) edge to connect two vertices in *G*(*P*, *Q*) if their corresponding gene ends are adjacent in *P* (respectively, *Q*). Given *k* black edges, forming a matching on $2k$ vertices, in a breakpoint graph, a *k*-break is defined as replacement of these edges with a set of *k* black edges that form another matching on the same set of $2k$ vertices. Note that an *h* break is a special case of a *k* break for $h<k$, in which case only *h* edges are replaced and the others remain the same. Basically, reversals, translocations, fusions, and fissions can be modeled by 2-breaks, while transpositions and block interchanges can be modeled by 3-breaks and 4-breaks, respectively. Although the *k*-break rearrangements may be unlikely to occur for $k>3$ in chromosomal evolution, they can provide a unifying and simpler

model for studying genome rearrangements. In fact, the 2-breaks are equivalent to the so-called double-cut-and-join (DCJ) operations, first introduced by Yancopoulos et al. (2005). Particularly, two consecutive DCJs can still model transpositions and block-interchanges. Later, Bergeron et al. (2006b) further utilized the DCJ operations by introducing a dual graph of breakpoint graph, called *adjacent graphs*, by replacing every edge of the breakpoint graph by a vertex and every vertex by an edge. Basically, the breakpoint and adjacent graphs are dual to each other and any properties can be found from one can also be found from the other.

The rest of this article is organized as follows. Some basic concepts and properties of permutation groups in algebra are introduced in Section 2, and their relationships with genome rearrangements are further described in Section 3. In Section 4, based on the permutation groups, we present an exact algorithm for the problem of sorting a circular chromosome by reversals and block-interchanges with a weight proportion 1:2 and also show its applicability to linear chromosomes. In Section 5, we further consider additional translocations (including fusions and fissions), which are weighted 1, when dealing with multi-chromosomal genomes and propose an efficient algorithm for the problem with circular/linear genomes. Finally, in Section 6, we provide conclusions.

2. PRELIMINARIES

Here, we briefly review a few basics on permutations in group theory that we will use for the study of genome rearrangement, by following the definitions and notation as introduced in Meidanis and Dias (2000) and Mira and Meidanis (2007). These concepts can be found in any textbook of algebra (Fraleigh, 2003). Given a set E of some integers, a *permutation* is a one-to-one mapping from E into itself. For instance, we may define a permutation α for $E = \{1, 2, 3, 4, 5, 6\}$ by specifying $\alpha(1) = 3$, $\alpha(2) = 1$, $\alpha(3) = 2$, $\alpha(4) = 4$, $\alpha(5) = 6$ and $\alpha(6) = 5$. Furthermore, the above mapping for α can be expressed using a *cycle notation* as illustrated in Figure 1 and simply denoted by $\alpha = (1, 3, 2)(4)(5, 6)$ (i.e., a product of three cycles). A cycle of length k, say $\alpha' = (a_1, a_2, \ldots, a_k)$, is called *k-cycle* and its length is denoted by $|\alpha'|$. In addition, this cycle can be rewritten as $(a_i, a_{i+1}, \ldots, a_k, a_1, a_2, \ldots, a_{i-1})$ (i.e., indices are cyclic), where $1 \leq i \leq k$. For example, $(1, 3, 2) = (3, 2, 1) = (2, 1, 3)$. Any two cycles are said to be *disjoint* if they have no elements in common. Then a permutation can be written in a unique way as the product of disjoint cycles, which is called the *cycle decomposition* of this permutation, if the order of the cycles in the product is ignored. In the rest of this article, we say "cycle in a permutation" to mean "cycle in the cycle decomposition of this permutation," unless otherwise specified. Usually, a 1-cycle, in which its element is said to be *fixed*, in a permutation is not written explicitly. Hence, the permutation α exemplified above is usually written as $(1, 3, 2)(5, 6)$. The so-called *identity permutation*, denoted by $\mathbf{1}$, is the permutation whose elements are all fixed.

Given two permutations α and β of E, their *composition* (or *product*), denoted by $\alpha\beta$, is defined to be a permutation of E with $\alpha\beta(x) = \alpha(\beta(x))$ for all $x \in E$. Clearly, $\alpha\beta = \beta\alpha$, if α and β are disjoint cycles. The *inverse* of α is a permutation, denoted as α^{-1}, such that $\alpha\alpha^{-1} = \alpha^{-1}\alpha = \mathbf{1}$. Notably, if a permutation is expressed by the product of disjoint cycles, then its inverse can be obtained by just reversing the order of the elements in each cycle. For instance, the inverse of a permutation $(1, 3, 2)(5, 6)$ of $E = \{1, 2, \ldots, 6\}$ is $(2, 3, 1)(6, 5)$. The *conjugation* of β by α, denoted by $\alpha \cdot \beta$, is the permutation $\alpha\beta\alpha^{-1}$, which actually is a permutation with the same cycle structure of β but each element x is changed by $\alpha(x)$. More clearly, if $\beta = (b_1, b_2, \ldots, b_j)(b_{j+1}, b_{j+2}, \ldots b_k)$, then $\alpha \cdot \beta = \alpha\beta\alpha^{-1} = (\alpha(b_1), \alpha(b_2), \ldots, \alpha(b_j))(\alpha(b_{j+1}), \alpha(b_{j+2}), \ldots, \alpha(b_k))$.

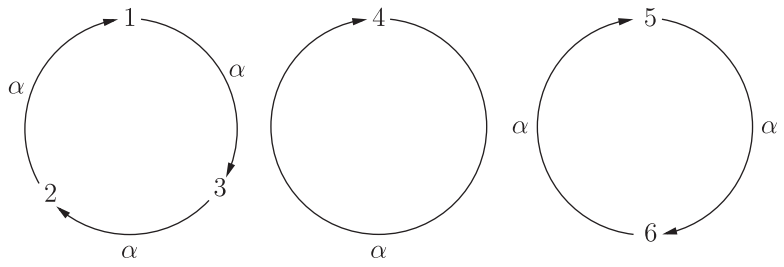

FIG. 1. Cycle diagram of a permutation $\alpha = (1, 3, 2)(4)(5, 6)$, meaning that $\alpha(1) = 3$, $\alpha(3) = 2$, $\alpha(2) = 1$, $\alpha(4) = 4$, $\alpha(5) = 6$ and $\alpha(6) = 5$.

Let $\alpha = (a_1, a_2)$ be a 2-cycle and β be an arbitrary permutation of E. Then the effect of applying α to β can be described as follows:

- If a_1 and a_2 are in the same cycle in β, then this cycle is broken into two smaller cycles in $\alpha\beta$ (or $\beta\alpha$), that is, α functions as a *split* operation of β. For example, if $\alpha = (1, 2)$ and $\beta = (1, 4, 5, 2, 3)$, then $\alpha\beta = (1, 4, 5)(2, 3)$ and $\beta\alpha = (3, 1)(4, 5, 2)$.
- If a_1 and a_2 are in two different cycles in β, then these two cycles are joined into a bigger cycle in $\alpha\beta$ (or $\beta\alpha$), that is, α functions as a *join* operation of β. For example, if $\alpha = (1, 3)$ and $\beta = (1, 4, 5)(2, 3)$, then $\alpha\beta = (1, 4, 5, 3, 2)$ and $\beta\alpha = (4, 5, 1, 2, 3)$.

Basically, every permutation α of E can be expressed as a product of 2-cycles (notably, in which 1-cycles are not written explicitly). There are, however, many ways of expressing α as a product of 2-cycles. For example, $(a_1, a_2, \ldots, a_k) = (a_1, a_2)(a_2, a_3)\ldots(a_{k-1}, a_k) = (a_1, a_k)(a_1, a_{k-1})\ldots(a_1, a_2)$, where $k \geq 3$. The *norm* of α, denoted by $\|\alpha\|$, is defined to be the minimum number k such that α can be expressed by a product of k 2-cycles. Let $n_c(\alpha)$ denote the number of disjoint cycles in the cycle decomposition of α. It should be noticed that $n_c(\alpha)$ counts also the non-expressed 1-cycles. For example, $\alpha = (1, 3, 2)(5, 6)$ is a permutation of $\{1, 2, \ldots, 6\}$ and then $n_c(\alpha) = 3$, instead of $n_c(\alpha) = 2$, since $\alpha = (1, 3, 2)(4)(5, 6)$. For two permutations α and β, α is said to *divide* β, simply denoted by $\alpha|\beta$, if and only if $\|\beta\alpha^{-1}\| = \|\beta\| - \|\alpha\|$. For instance, let $\alpha = (1, 2)$ and $\beta = (1, 4, 5, 2, 3)$ be two permutations of $E = \{1, 2, \ldots, 5\}$. Then $\beta\alpha^{-1} = (1, 3)(2, 4, 5)$. According to Lemma 2.1 below, we have $\|\beta\alpha^{-1}\| = 3$, $\|\beta\| = 4$ and $\|\alpha\| = 1$. As a result, $\|\beta\alpha^{-1}\| = \|\beta\| - \|\alpha\|$ and hence $\alpha|\beta$.

The following six lemmas are basic and useful, and the reader can refer to a textbook of algebra or Appendix A here for the details of their proofs.

Lemma 2.1. *For any permutation α of E, $\|\alpha\| = |E| - n_c(\alpha)$.*

Corollary 2.1. *Let α and β be any two permutations of E. Then $\alpha|\beta$ if and only if $n_c(\beta\alpha^{-1}) = n_c(\beta) + \|\alpha\|$.*

Lemma 2.2. *Let α and β be any two permutations of E. Then $\|\alpha \cdot \beta\| = \|\beta\|$.*

Lemma 2.3. *Let α and β be any two permutations of E. Then $\|\alpha\beta\| = \|\beta\alpha\|$.*

Lemma 2.4. *Let α and β be any two permutations of E. Then $\|\alpha\beta\| \leq \|\alpha\| + \|\beta\|$.*

Lemma 2.5. *Let α, β and γ be any three permutations of E and $\alpha = \beta\gamma$. If $\beta|\alpha$ or $\gamma|\alpha$, then $\|\alpha\| = |\beta\| + \|\gamma\|$.*

Lemma 2.6. *Let α, β and γ be any three permutations of E. If $\alpha|\beta$ and $\beta|\gamma$, then $\alpha|\gamma$.*

Lemma 2.7. *Let $a_1, a_2, \ldots, a_k \in E$ and β be any permutation of E. Then $a_1, a_2, \ldots, a_k$ are in the same cycle of β and they appear in this cycle in the order of $a_1, a_2, \ldots, a_k$ if and only if $(a_1, a_2, \ldots, a_k)|\beta$.*

Proof. We prove this lemma by induction on k. First, assume that $k = 2$. Let $\alpha = (a_1, a_2)$, and let $\beta' = (b_1 = a_1, \ldots, b_i = a_2, \ldots, b_j)$ be a cycle in β, where $2 \leq i < j$, and $\beta = \beta''\beta'$. Then $\beta\alpha^{-1} = \beta''\beta'(b_1, b_i)^{-1} = \beta''(b_2, b_3, \ldots, b_i)(b_{i+1}, \ldots, b_j, b_1)$. Clearly, $n_c(\beta\alpha^{-1}) = n_c(\beta) + 1$ and $\|\alpha\| = 1$. By Lemma 2.1, $\|\beta\alpha^{-1}\| = |E| - n_c(\beta\alpha^{-1}) = |E| - n_c(\beta) - 1 = \|\beta\| - \|\alpha\|$, resulting in that $\alpha|\beta$. Conversely, suppose that $\alpha|\beta$. Then by Corollary 2.1, $n_c(\beta\alpha^{-1}) = n_c(\beta) + \|\alpha\| = n_c(\beta) + 1$. If a_1 and a_2 are not in the same cycle of β, then $\beta\alpha^{-1} = \beta(a_1, a_2)$ that causes two disjoint cycles in β, one containing a_1 and the other containing a_2, to be joined together. This implies that $n_c(\beta\alpha^{-1}) = n_c(\beta) - 1$, a contradiction.

Next, assume that the lemma holds for some k. Let $\alpha = (a_1, a_2, \ldots, a_{k+1})$ and $\alpha' = (a_1, a_2, \ldots, a_k)$. Clearly, $\alpha = \alpha'(a_k, a_{k+1})$ and $\|\alpha\| = \|\alpha'\| + 1 = k + 1$. Now, suppose that $a_1, a_2, \ldots, a_{k+1}$ appear in this order in a cycle β' of β, where we let $\beta = \beta''\beta'$. Then $\beta\alpha^{-1} = \beta''\beta'(a_k, a_{k+1})\alpha'^{-1}$, in which β' is partitioned into two disjoint cycles, say β'_1 containing $a_1, a_2, \ldots, a_k$ in this order and β'_2 containing a_{k+1}. That is, $\beta\alpha^{-1} = \beta''\beta'_1\beta'_2\alpha'^{-1}$ (therefore, $n_c(\beta\alpha^{-1}) = n_c(\beta''\beta'_1\beta'_2\alpha'^{-1})$) and $n_c(\beta) = n_c(\beta''\beta'_1\beta'_2) - 1$. Moreover,

$\alpha'|\beta''\beta_1'\beta_2'$ by the induction hypothesis and hence $n_c(\beta''\beta_1'\beta_2'\alpha'^{-1}) = n_c(\beta''\beta_1'\beta_2') + \|\alpha'\|$. As a result, $n_c(\beta\alpha^{-1}) = n_c(\beta) + \|\alpha\|$ and, therefore, $\alpha|\beta$ by Corollary 2.1. Conversely, suppose that $\alpha|\beta$. Then $(a_1, a_k, a_{k+1})|\alpha$ and $\alpha'|\alpha$ based on the induction hypothesis and, consequently, $(a_1, a_k, a_{k+1})|\beta$ and $\alpha'|\beta$ by Lemma 2.6. By the induction hypothesis again, we understand that a_1, a_k, a_{k+1} appear in this order in a cycle of β, and $a_1, a_2, \ldots, a_k$ also appear in this order in a cycle of β. As a result, $a_1, a_2, \ldots, a_{k+1}$ appear in this order in the same cycle of β. ■

3. PERMUTATION GROUPS VERSUS GENOME REARRANGEMENTS

As mentioned before, a gene is usually represented by a signed integer in the genome rearrangement studies. Here, we follow this convention, although it can be any label (e.g., gene name) commonly used by biologists. To properly model a DNA, which is well known as a double stranded molecule, we let $E = \{-1, 1, -2, 2, \ldots, -n, n\}$, in which n is the number of genes and each gene i has counterpart gene $-i$ in the same location in the opposite strand. Let $\Gamma = (1, -1)(2, -2)\ldots(n, -n)$. Clearly, $\Gamma^2 = \mathbf{1}$, that is, $\Gamma^{-1} = \Gamma$. A cycle is said to be *admissible* if it does not contain i and $-i$ simultaneously for some gene $i \in E$. Then an admissible n-cycle can be used to represent a DNA strand that is constituted by n genes in some order. Given a DNA strand, say π_1, $\pi_2 = \Gamma \cdot \pi_1^{-1}$ is its *reverse complement*, since π_1^{-1} is the reverse of π_1 and $\Gamma \cdot \pi_1^{-1}$ is the complement of π_1^{-1}. For our purpose, we here represent the DNA molecule, named π, by the product of the two strands π_1 and π_2, that is, $\pi = \pi_1\pi_2 = \pi_2\pi_1$ (since π_1 and π_2 are disjoint). In such a representation, flipping π into $\Gamma \cdot \pi^{-1}$ does not affect the DNA molecule, since $\Gamma \cdot \pi^{-1} = \Gamma \cdot (\pi_1\pi_2)^{-1} = \Gamma \cdot (\pi_2^{-1}\pi_1^{-1}) = \Gamma \cdot (\Gamma\pi_1\Gamma\pi_1^{-1}) = \Gamma^2\pi_1\Gamma\pi_1^{-1}\Gamma = \pi_1\pi_2 = \pi$. This representation of a DNA molecule (or chromosome), which can certainly be applied to a genome with multiple chromosomes, was first introduced in the pioneering work by Meidanis and Dias (2000) in the study of genome rearrangement using permutation groups in algebra. In the following, we shall explain how to model elementary rearrangement operations in a genome, such as reversals and block-interchanges acting on single chromosome and fusions, fissions and translocations acting on multiple chromosomes, in a simple way, particularly from the permutation group point of view.

Lemma 3.1. *Let $\pi = \pi_1\pi_2$ be a (single-/multi-chromosomal) genome with n genes. Then $\pi\Gamma = \Gamma\pi^{-1}$ and $\Gamma\pi = \pi^{-1}\Gamma$.*

Proof. First, we have $\pi = \pi_1\pi_2 = \pi_1\Gamma\pi_1^{-1}\Gamma$, since $\pi_2 = \Gamma \cdot \pi_1^{-1} = \Gamma\pi_1^{-1}\Gamma^{-1}$ and $\Gamma^{-1} = \Gamma$. Then $\pi^{-1} = (\pi_1\Gamma\pi_1^{-1}\Gamma)^{-1} = \Gamma^{-1}\pi_1\Gamma^{-1}\pi_1^{-1} = \Gamma\pi_1\Gamma\pi_1^{-1}$ and hence, $\Gamma\pi^{-1}\Gamma = \pi$ and $\Gamma\pi\Gamma = \pi^{-1}$ (since $\Gamma^2 = \mathbf{1}$). Consequently, $\pi\Gamma = \Gamma\pi^{-1}$ and $\Gamma\pi = \pi^{-1}\Gamma$. ■

Suppose that $\pi = \pi_1\pi_2$ is a single chromosomal genome. Essentially, a reversal operation acting on π can be simply considered as a kind of block-interchange between π_1 and π_2. For the purpose of illustration, let us take Figure 2 for an example. Note that the previous gene of a gene x in π_1 is $\pi^{-1}(x)$ and its counterpart in π_2 is $\Gamma(x)$. Hence, the counterpart of $\pi^{-1}(x)$ in π_2 is $\Gamma\pi^{-1}(x)$, which equals to $\pi\Gamma(x)$ by Lemma 3.1. Suppose that the genome π in Figure 2 is rearranged by a reversal γ that in effect replaces the path A_1 of genes from u to v (including u but excluding v) in strand π_1 with its reverse complement, and simultaneously the path A_2 of genes from $\Gamma(v)$ to $\Gamma(u)$ (including $\Gamma(u)$ but excluding $\Gamma(v)$) in strand π_2 with its reverse complement. Notably, the reverse complement of A_1 is A_2, and vice versa. As a result, the rearrangement of $\gamma\pi$ can be done simply by an interchange between A_1 in π_1 and A_2 in π_2. Most particularly, this genome rearrangement can be modeled by the composition of two 2-cycles and π, as represented in the following lemma.

Lemma 3.2. *If u and v are in the same strand of π, then $\gamma = (v, \pi\Gamma(u))(u, \pi\Gamma(v))$ is a reversal operation acting on π.*

The following details the rearrangement result that is exemplified in Figure 2 by applying the reversal $\gamma = (v, \pi\Gamma(u))(u, \pi\Gamma(v))$ to genome $\pi = \pi_1\pi_2$.

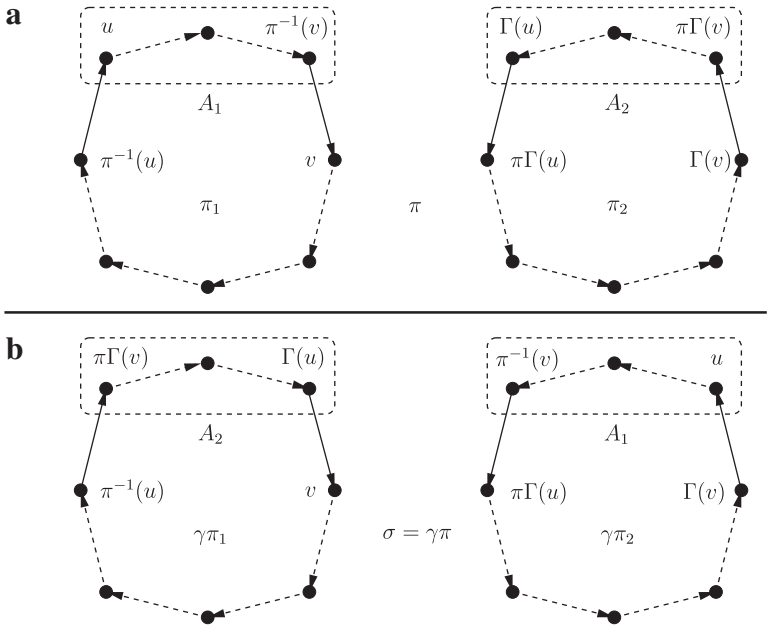

FIG. 2. (**a**) A chromosome $\pi = \pi_1\pi_2$, where solid arrows indicate consecutive genes and dashed arrows indicate paths of solid arrows. (**b**) The resulting chromosome σ by applying a reversal γ to π.

$$\gamma\pi = (v, \pi\Gamma(u))(u, \pi\Gamma(v))\pi_1\pi_2$$

$$= (v, \pi\Gamma(u))(u, \pi\Gamma(v))\overbrace{(u, \ldots, \pi^{-1}(v)}^{A_1}, v, \ldots, \pi^{-1}(u))\overbrace{(\pi\Gamma(v), \ldots, \Gamma(u)}^{A_2}, \pi\Gamma(u), \ldots, \Gamma(v))$$

$$= (v, \pi\Gamma(u))(u, \ldots, \pi^{-1}(v), v, \ldots, \pi^{-1}(u), \pi\Gamma(v), \ldots, \Gamma(u), \pi\Gamma(u), \ldots, \Gamma(v))$$

(i.e., $(u, \pi\Gamma(v))$ functions as a join operation.)

$$= \overbrace{(\pi\Gamma(v), \ldots, \Gamma(u)}^{A_2}, v, \ldots, \pi^{-1}(u))\overbrace{(u, \ldots, \pi^{-1}(v)}^{A_1}, \pi\Gamma(u), \ldots, \Gamma(v))$$

(i.e., $(v, \pi\Gamma(u))$ functions as a split operation.)

As demonstrated above, $\gamma = (v, \pi\Gamma(u))(u, \pi\Gamma(v))$ indeed acts as a reversal rearrangement when applied to chromosome $\pi = \pi_1\pi_2$, by reversing the segment of π from gene u to gene v (but excluding v). In the composition of $\gamma\pi$, intriguingly, $(u, \pi\Gamma(v))$ operates as a join of $\pi_1\pi_2$ and $(v, \pi\Gamma(u))$ as a split of $(u, \pi\Gamma(v))\pi_1\pi_2$.

In fact, a block-interchange rearrangement on a chromosome $\pi = \pi_1\pi_2$ can also be implemented by the composition of four 2-cycles and $\pi_1\pi_2$, just based on our previous work on sorting by block-interchanges (Lin et al., 2005). Here, we simply take Figure 3 for an illustration. Suppose that β is a block-interchange that affects π by exchanging the path A_1 of genes from v to $\pi^{-1}(w)$ and the path B_1 of genes from x to $\pi^{-1}(u)$, and also exchanging the path A_2 of genes from $\pi\Gamma(w)$ to $\Gamma(v)$ and the path B_2 of genes from $\pi\Gamma(u)$ to $\Gamma(x)$. Then we have the following lemma immediately, where $(v, x)(u, w)$ functions as the block-interchange on π_1 and $(\pi\Gamma(w), \pi\Gamma(u))(\pi\Gamma(x), \pi\Gamma(v))$ as the block-interchange on π_2.

Lemma 3.3. *If u,v,w and x are in the same strand of π in this order, then $\beta = (v, x)(u, w)$ $(\pi\Gamma(w), \pi\Gamma(u))(\pi\Gamma(x), \pi\Gamma(v))$ is a block-interchange operation acting on π.*

The genome rearrangement result obtained by the composition of β and π, as exemplified in Figure 3, is detailed as follows.

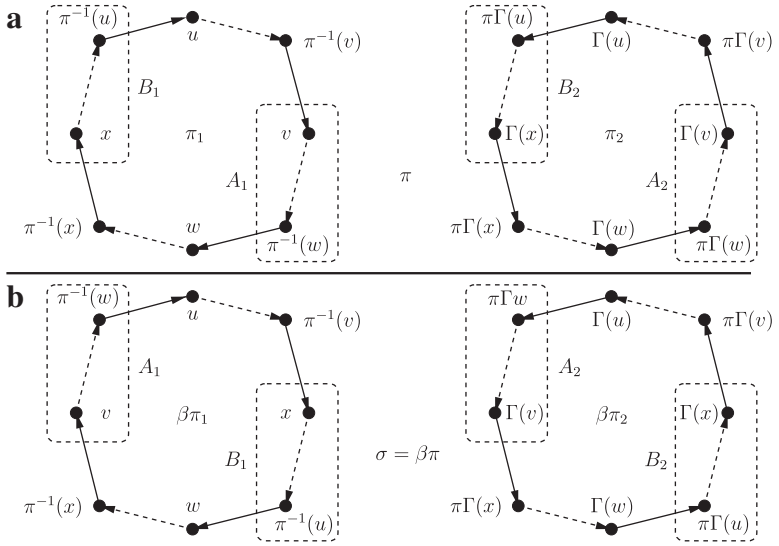

FIG. 3. (**a**) A chromosome $\pi = \pi_1\pi_2$. (**b**) The resulting chromosome σ by applying a block-interchange β to π.

$$
\begin{aligned}
\beta\pi &= (v,x)(u,w)(\pi\Gamma(w),\pi\Gamma(u))(\pi\Gamma(x),\pi\Gamma(v))\pi_1\pi_2 \\
&= (v,x)(u,w)(u,\ \ldots,\pi^{-1}(v),\overbrace{v,\ \ldots,\pi^{-1}(w)}^{A_1},w,\ \ldots,\pi^{-1}(x),\overbrace{x,\ \ldots,\pi^{-1}(u)}^{B_1} \\
&\quad (\pi\Gamma(w),\pi\Gamma(u))(\pi\Gamma(x),\pi\Gamma(v))(\pi\Gamma(x),\ \ldots,\Gamma(w),\overbrace{\pi\Gamma(w),\ \ldots,\Gamma(v)}^{A_2},\pi\Gamma(v),\ \ldots, \\
&\quad \Gamma(u),\overbrace{\pi\Gamma(u),\ \ldots,\Gamma(x)}^{B_2}) \\
&= (v,x)(u,\ \ldots,\pi^{-1}(v),v,\ \ldots,\pi^{-1}(w))(w,\ \ldots,\pi^{-1}(x),x,\ \ldots,\pi^{-1}(u))(\pi\Gamma(w),\pi\Gamma(u)) \\
&\quad (\pi\Gamma(x),\ \ldots,\Gamma(w),\pi\Gamma(w),\ \ldots,\Gamma(v))(\pi\Gamma(v),\ \ldots,\Gamma(u),\pi\Gamma(u),\ \ldots,\Gamma(x)) \\
&\quad \text{(i.e., } (u,w) \text{ and } (\pi\Gamma(x),\pi\Gamma(v)) \text{ are split operations of } \pi_1 \text{ and } \pi_2, \textit{respectively.)} \\
&= (u,\ \ldots,\pi^{-1}(v),\overbrace{x,\ \ldots,\pi^{-1}(u)}^{B_1},w,\ \ldots,\pi^{-1}(x),\overbrace{v,\ \ldots,\pi^{-1}(w)}^{A_1}) \\
&\quad (\pi\Gamma(x),\ \ldots,\Gamma(w),\overbrace{\pi\Gamma(u),\ \ldots,\Gamma(x)}^{B_2},\pi\Gamma(v),\ \ldots,\Gamma(u),\overbrace{\pi\Gamma(w),\ \ldots,\Gamma(v)}^{A_2}) \\
&\quad \text{(i.e., } (v,x) \text{ and } (\pi\Gamma(w),\pi\Gamma(u)) \text{ are join operations of } \pi_1 \text{ and } \pi_2, \text{ respectively.)}
\end{aligned}
$$

Now, suppose that $\pi = \pi_1\pi_2$ is a genome of multiple chromosomes and τ is a fission/fusion acting on π. Then, as demonstrated in the previous section, τ can easily be modeled by two 2-cycles, both functioning as split/join operations of π. For instance, as exemplified in Figure 4, if τ is a fission that affects a chromosome π by splitting it into two smaller chromosomes, then τ can be modeled as $(u,w)(\pi\Gamma(w),\pi\Gamma(u))$, where (u,w) and $(\pi\Gamma(w),\pi\Gamma(u))$ function as a split operation of π_1 and π_2, respectively. Clearly, on the other hand, the reverse process of the fission (by applying $(u,w)(\pi\Gamma(w),\pi\Gamma(u))$ to $\tau\pi$) becomes a fusion, in which both (u,w) and $(\pi\Gamma(w),\pi\Gamma(u))$ function as join operations of $\tau\pi$. In other words, the product of two 2-cycles, $(u,w)(\pi\Gamma(w),\pi\Gamma(u))$, acts as a fission for π and as a fusion for $\sigma = \tau\pi$.

Lemma 3.4. *If u and w are in the same strand of π, then $\tau = (u,w)(\pi\Gamma(w),\pi\Gamma(u))$ is a fission/fusion acting on π.*

Particularly, it should be noted here that translocations are equivalent to fusions for circular chromosomes (Alekseyev and Pevzner, 2008; Alekseyev, 2008). However, for linear chromosomes, fusions and fissions are just special cases of translocation. A translocation acts on two linear chromosomes by exchanging an end

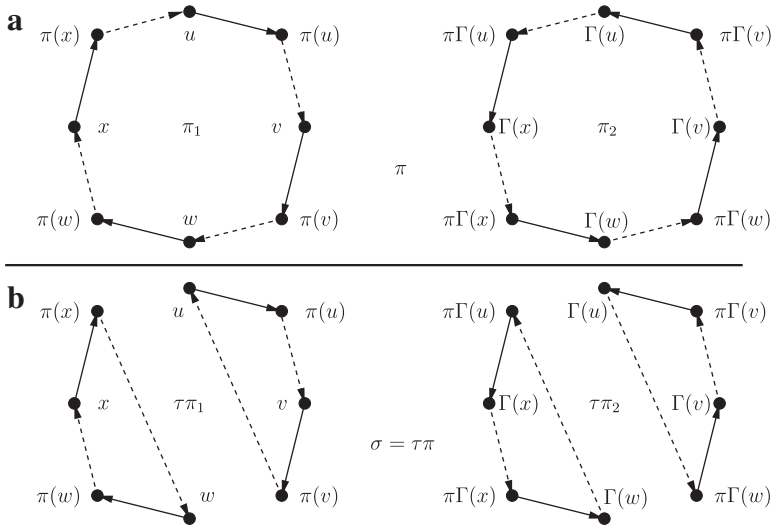

FIG. 4. (**a**) A genome $\pi = \pi_1\pi_2$. (**b**) The resulting genome σ by applying a fission τ to π. Note that its reverse process corresponds to a fusion.

segment of one chromosome with an end segment of the other chromosome. There are two types of translocations (i.e., prefix-prefix and prefix-suffix translocations) usually mentioned in related literature. However, we can mimic one type of translocation by a flip of one of the chromosomes, followed by a translocation of the other type, because, as mentioned before, flipping a chromosome entirely does not change the chromosome and is thus free. As exemplified in Figure 5, the translocation τ affects π with two linear chromosomes A and B by exchanging the end segment A_2 of A with the end segment B_2 of B. In fact, by representing the linear chromosomes using (circular) permutations, the above translocation can be modeled by four 2-cycles, as demonstrated as follows. Note that in this case where π is represented as a (circular) permutation, we have $\pi\Gamma(u) = \Gamma(y)$ and $\pi\Gamma(w) = \Gamma(z)$.

$$
\begin{aligned}
\tau\pi &= (u,w)(v,x)(\pi\Gamma(w),\pi\Gamma(u))(\pi\Gamma(x),\pi\Gamma(v))\pi \\
&= (u,w)(v,x)(\Gamma(z),\Gamma(y))(\pi\Gamma(x),\pi\Gamma(v))\pi \\
&= (u,w)(v,\ldots,y,u,\ldots,\pi^{-1}(v),x,\ldots,z,w,\ldots,\pi^{-1}(x)) \\
&\quad (\Gamma(z),\Gamma(y))(\pi\Gamma(x),\ldots,\Gamma(w),\Gamma(z),\ldots,\Gamma(x),\pi\Gamma(v),\ldots,\Gamma(u),\Gamma(y),\ldots,\Gamma(v)) \\
&= \overbrace{(u,\ldots,\pi^{-1}(v)}^{A_1},\overbrace{x,\ldots,z)}^{B_2}\overbrace{(w,\ldots,\pi^{-1}(x)}^{B_1},\overbrace{v,\ldots,y)}^{A_2} \\
&\quad \overbrace{(\Gamma(z),\ldots,\Gamma(x)}^{B_2},\overbrace{\pi\Gamma(v),\ldots,\Gamma(u))}^{A_1}\overbrace{(\Gamma(y),\ldots,\Gamma(v)}^{A_2},\overbrace{\pi\Gamma(x),\ldots,\Gamma(w))}^{B_1}
\end{aligned}
$$

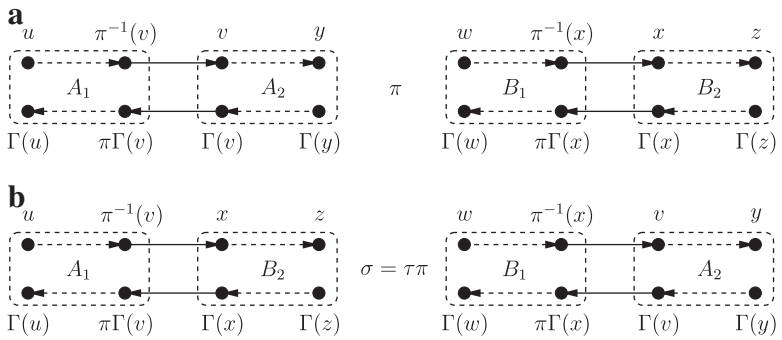

FIG. 5. (**a**) A genome π with two linear chromosomes A and B. (**b**) The resulting genome σ by applying a translocation τ to π.

Lemma 3.5. *If u and v are in the same strand of π, w and x are also in the same strand of π, and u and w are at the ends of π, then $\tau = (u,w)(v,x)(\pi\Gamma(w), \pi\Gamma(u))(\pi\Gamma(x), \pi\Gamma(v))$ is a translocation acting on π.*

4. ALGORITHMS FOR SORTING BY WEIGHTED REVERSALS AND BLOCK-INTERCHANGES

In this section, we shall first utilize the properties of permutation groups to design an efficient exact algorithm for solving the problem of transforming a circular chromosome $\pi = \pi_1\pi_2$ into another $\sigma = \sigma_1\sigma_2$ by reversals and block-interchanges with a weight ratio 1:2. For brevity, we denote this weighted genome rearrangement problem as SoRT(1,2) in the following. Next, we shall show that the problem of sorting by reversals and block-interchanges is equivalent for circular and linear chromosomes. It should be mentioned here that Mira and Meidanis (2007) have independently used the permutations groups to propose an $\mathcal{O}(n^2)$-time algorithm for solving the SoRT(1,2) problem with circular chromosomes, which is actually the same as ours in spirit, where n is the number of genes in the studied chromosome. Recall that any cycle can be expressed as a product of 2-cycles and moreover, every reversal (respectively, block-interchange) affecting a genome π can be implemented by a product of two (respectively, four) 2-cycles and π. Furthermore, the composition of $\sigma\pi^{-1}$ and π is σ, intuitively suggesting that $\sigma\pi^{-1}$ can be expressed as a product of 2-cycles that operates as a sequence of reversals and block-interchanges to optimally transform π into σ. We shall demonstrate in the sequel how to fulfill such an idea.

Lemma 4.1. *Let π and σ represent two different chromosomes. If α is a cycle in $\sigma\pi^{-1}$, then $(\pi\Gamma)\cdot\alpha^{-1}$ is also a cycle in $\sigma\pi^{-1}$.*

Proof. Let u and v be two consecutive elements in α. That is, $\alpha(u) = v$, or, equivalently, $\sigma\pi^{-1}(u) = v$, since α is known to be a cycle in (the cycle decomposition of) $\sigma\pi^{-1}$. In the following, we show that $\pi\Gamma(v)$ and $\pi\Gamma(u)$ are two consecutive elements in a cycle of $\sigma\pi^{-1}$, meaning that $\sigma\pi^{-1}(\pi\Gamma(v)) = \pi\Gamma(u)$.

$$\begin{aligned}
\sigma\pi^{-1}(\pi\Gamma(v)) &= \sigma\pi^{-1}\pi\Gamma(v) && \text{(by the definition of composition)}\\
&= \sigma\Gamma(v) && (\text{since } \pi^{-1}\pi = \mathbf{1})\\
&= \sigma\Gamma(\sigma\pi^{-1}(u)) && (\text{since } \sigma\pi^{-1}(u) = v)\\
&= \sigma\sigma^{-1}\Gamma\pi^{-1}(u) && (\text{since } \Gamma\sigma = \sigma^{-1}\Gamma, \text{by Lemma 3.1})\\
&= \Gamma\pi^{-1}(u) && (\text{since } \sigma\sigma^{-1} = \mathbf{1})\\
&= \pi\Gamma(u) && (\text{since } \Gamma\pi^{-1} = \pi\Gamma, \text{by Lemma 3.1})
\end{aligned}$$

It should be noticed that if α is a cycle of length 1, e.g., $\alpha(u) = u$, then with the similar discussion as described above, we can show that $\sigma\pi^{-1}(\pi\Gamma(u)) = \pi\Gamma(u)$, meaning that $\pi\Gamma(u)$ is fixed in $\sigma\pi^{-1}$. Clearly, $(\pi\Gamma)\cdot\alpha^{-1} = (\pi\Gamma(u))$ is another cycle in $\sigma\pi^{-1}$ and hence the lemma holds. Now, let $\alpha = (a_1, a_2, \ldots, a_k)$, where $2 \le k \le n$. According to the above discussion, $\sigma\pi^{-1}$ must contain the cycle $\alpha' = (\pi\Gamma(a_k), \pi\Gamma(a_{k-1}), \ldots, \pi\Gamma(a_1))$ and clearly, $\alpha' = (\pi\Gamma)\cdot\alpha^{-1}$. We claim that α and α' are two different cycles in $\sigma\pi^{-1}$. Suppose that α and α' are the same cycle in $\sigma\pi^{-1}$. Then $a_i \neq \pi\Gamma(a_i)$ for each $1 \le i \le k$; otherwise, for some i, we have $a_i = \pi\Gamma(a_i) = \pi(-a_i)$, causing both genes $-a_i$ and a_i to be in the same chromosome strand of π, which is not allowed because π is represented as an admissible chromosome. Without loss of generality, we let $a_1 = \pi\Gamma(a_j)$, where $2 \le j \le k$, and immediately, $a_i = \pi\Gamma(a_{j-i+1})$ for $2 \le i \le j$, due to the assumption of $\alpha = \alpha'$. In this case, if j is odd, then $a_{\lfloor \frac{j}{2} \rfloor + 1} = \pi\Gamma(a_{\lfloor \frac{j}{2} \rfloor + 1})$, a contradiction. In other words, j is even and consequently, we have $a_{\frac{j}{2}} = \pi\Gamma(a_{\frac{j}{2}+1})$ and $a_{\frac{j}{2}+1} = \pi\Gamma(a_{\frac{j}{2}})$. The fact, moreover, that $a_{\frac{j}{2}}$ and $a_{\frac{j}{2}+1}$ are consecutive in α of $\sigma\pi^{-1}$ leads to $\sigma\pi^{-1}(a_{\frac{j}{2}}) = a_{\frac{j}{2}+1}$, resulting in $\sigma\pi^{-1}\pi\Gamma(a_{\frac{j}{2}+1}) = a_{\frac{j}{2}+1}$ and consequently $\sigma\Gamma(a_{\frac{j}{2}+1}) = a_{\frac{j}{2}+1}$. This suggests, however, that $-a_{\frac{j}{2}+1}$ and $a_{\frac{j}{2}+1}$ should be two consecutive genes in the same chromosome strand of σ, which is again not allowed because σ is denoted as an admissible chromosome. Thus, α and α' are different cycles in $\sigma\pi^{-1}$ and the lemma holds completely. ■

According to Lemma 4.1, every cycle α in $\sigma\pi^{-1}$ has a *mate* cycle $(\pi\Gamma)\cdot\alpha^{-1}$, which also has its own mate cycle $(\pi\Gamma)^2\cdot\alpha$ that equals to α, since $(\pi\Gamma)^2 = \pi\Gamma\pi\Gamma = \pi\Gamma\Gamma\pi^{-1} = \pi\pi^{-1} = \mathbf{1}$ (by Lemma 3.1 and $\Gamma^2 = \mathbf{1}$).

That is, α and $(\pi\Gamma)\cdot\alpha^{-1}$ are each other's mate cycles in $\sigma\pi^{-1}$, also implying that $n_c(\sigma\pi^{-1})$ is even. Given a cycle α in $\sigma\pi^{-1}$, we say x and y to be *adjacent* in α if $\alpha(x)=y$ or $\alpha(y)=x$.

Lemma 4.2. *Let π and σ represent two different chromosomes. Suppose that $(a,b)|\pi$ for any two elements a and b in a cycle of $\sigma\pi^{-1}$ (i.e., $(a,b)|\sigma\pi^{-1}$). Then we have $\sigma\pi^{-1}=(\sigma_1\pi_1^{-1})(\sigma_2\pi_2^{-1})$ and $n_c(\sigma\pi^{-1})=n_c(\sigma_1\pi_1^{-1})+n_c(\sigma_2\pi_2^{-1})$ with $n_c(\sigma_1\pi_1^{-1})=n_c(\sigma_2\pi_2^{-1})$.*

Proof. Let $\pi_1=(a_1,a_2,\ldots,a_n)$ and $\pi_2=(b_1,b_2,\ldots,b_n)$. According to Lemma 2.7, the given assumption implies that all the elements in each cycle α of $\sigma\pi^{-1}$ belong to either π_1 or π_2. The fact that $\sigma=(\sigma\pi^{-1})\pi$ indicates clearly that σ_1 (respectively, σ_2) is a permutation of $\{a_1,a_2,\ldots,a_n\}$ (respectively, $\{b_1,b_2,\ldots,b_n\}$). Recall that $\sigma\pi^{-1}=\sigma_1\sigma_2\pi_2^{-1}\pi_1^{-1}$, in which, as indicated by the above property, both σ_1 and π_1^{-1} are disjoint to σ_2, as well as π_2^{-1}, and therefore, $\sigma\pi^{-1}=\sigma_1\pi_1^{-1}\sigma_2\pi_2^{-1}$and $n_c(\sigma\pi^{-1})=n_c(\sigma_1\pi_1^{-1})+n_c(\sigma_2\pi_2^{-1})$. Next, for simplicity of our discussion, we assume that all the numbers in π_1 have the same sign, say "+", and let $\alpha=(c_1,c_2,\ldots,c_k)$ belong to $\sigma_1\pi_1^{-1}$, where $1\le k$ (i.e., α can be a 1-cycle). By Lemma 4.1, α has a mate cycle $\alpha'=(\pi\Gamma)\cdot\alpha^{-1}$ in $\sigma\pi^{-1}$. By definition, $\alpha'=(\pi\Gamma(c_k),\pi\Gamma(c_{k-1}),\ldots,\pi\Gamma(c_1))=(\pi(-c_k),\pi(-c_{k-1}),\ldots,\pi(-c_1))=(\pi_2(-c_k),\pi_2(-c_{k-1}),\ldots,\pi_2(-c_1))$. Clearly, all the numbers in α' have the same sign of "−", suggesting that α' is a cycle in $\sigma_2\pi_2^{-1}$, instead of $\sigma_1\pi_1^{-1}$. In other words, for each cycle α in $\sigma_1\pi_1^{-1}$, we can find another cycle α' that is in $\sigma_2\pi_2^{-1}$. As a result, $n_c(\sigma_1\pi_1^{-1})=n_c(\sigma_2\pi_2^{-1})$. ■

Now, we suppose that the condition of Lemma 4.2 holds, that is, $(a,b)|\pi$ for any two elements a and b in a cycle of $\sigma\pi^{-1}$. By Lemma 4.2, we know that π_1 (respectively, π_2) is a permutation of σ_1 (respectively, σ_2) and $n_c(\sigma\pi^{-1})=n_c(\sigma_1\pi_1^{-1})+n_c(\sigma_2\pi_2^{-1})$, where $n_c(\sigma_1\pi_1^{-1})=n_c(\sigma_2\pi_2^{-1})$. Recall that in our previous study (Lin et al., 2005, see Theorem 1), we can use a minimum sequence of $k=\frac{|E|/2-n_c(\sigma_1\pi_1^{-1})}{2}=\frac{|E|-n_c(\sigma\pi^{-1})}{4}$ block-interchanges, say $\beta_1,\beta_2,\ldots,\beta_k$, to transform π_1 into σ_1 and moreover, each β_i, $1\le i\le k$, can be expressed by a product of two 2-cycles, say $\beta_i=(v,x)(u,w)$. We have also shown in (Lin et al., 2005) that u and w are adjacent in a cycle α_1 of $\sigma_1\pi_1^{-1}$ and v and x are adjacent in a cycle α_2 of $\sigma_1\pi_1^{-1}$ (u,w) such that (u,w) acts on π_1 as a split and (v,x) acts on $(u,w)\pi_1$ as a join. By Lemma 4.1, we can first find two adjacent elements $\pi\Gamma(w)$ and $\pi\Gamma(u)$ in the cycle $(\pi\Gamma)\cdot\alpha_1^{-1}$ of $\sigma_2\pi_2^{-1}$ and then two adjacent elements $\pi\Gamma(x)$ and $\pi\Gamma(v)$ in the cycle $(\pi\Gamma)\cdot\alpha_1^{-1}$ of $\sigma_2\pi_2^{-1}(\pi\Gamma(w),\pi\Gamma(u))$. Let $\beta_i'=(\pi\Gamma(w),\pi\Gamma(u))(\pi\Gamma(x),\pi\Gamma(v))$. By Lemma 3.3, $\beta_i\beta_i'$ is clearly a block-interchange of π. In other words, $\beta_1\beta_1',\beta_2\beta_2',\ldots,\beta_k\beta_k'$ are k block-interchanges that can transform π into σ. Therefore, we have the following lemma immediately. ■

Lemma 4.3. *Let π and σ represent two different chromosomes. Suppose that $(a,b)|\pi$ for any two elements a and b in a cycle of $\sigma\pi^{-1}$. Then π can be transformed into σ through a sequence of $\frac{|E|-n_c(\sigma\pi^{-1})}{4}$ block-interchanges.*

Lemma 4.4. *Let π and σ represent two different chromosomes. Suppose that there are at least two elements a and b in a cycle α of $\sigma\pi^{-1}$ such that $(a,b)\not|\pi$. Then we can find two 2-cycles $(u,\pi\Gamma(v))$ and $(v,\pi\Gamma(u))$ with $\alpha(u)=\pi\Gamma(v)$ and $(\pi\Gamma)\cdot\alpha^{-1}(v)=\pi\Gamma(u)$ such that $\gamma=(v,\pi\Gamma(u))(u,\pi\Gamma(v))$ acts on π as a reversal operation. Moreover, $n_c(\sigma(\gamma\pi)^{-1})=n_c(\sigma\pi^{-1})+2$.*

Proof. Let $\alpha=(a_1=u,a_2=\pi\Gamma(v),\ldots,a_k)$, where $k>1$. According to the given assumption, as well as Lemma 2.7, there are at least two adjacent elements, say u and $\pi\Gamma(v)$, in α such that for example, u is in π_1 and $\pi\Gamma(v)$ is in π_2. By Lemma 4.1, α has a mate cycle $(\pi\Gamma)\cdot\alpha^{-1}$, that by definition is $(\pi\Gamma(a_k),\ldots,\pi\Gamma(a_2)=v,\pi\Gamma(a_1)=\pi\Gamma(u))$ in $\sigma\pi^{-1}$. The above statement indicates that $\alpha(u)=\pi\Gamma(v)$ and $(\pi\Gamma)\cdot\alpha^{-1}(v)=\pi\Gamma(u)$. By Lemma 3.2, γ that is $(v,\pi\Gamma(u))(u,\pi\Gamma(v))$ acts on π as a reversal. $\sigma(\gamma\pi)^{-1}=\sigma\pi^{-1}\gamma^{-1}=\sigma\pi^{-1}(u,\pi\Gamma(v))(v,\pi\Gamma(u))$, resulting in both $\pi\Gamma(v)$ and $\pi\Gamma(u)$ being fixed. The reason is that $\alpha(u,\pi\Gamma(v))=(a_1,a_3,\ldots,a_k)(u,\pi\Gamma(v))^2=(a_1,a_3,\ldots,a_k)$ and similarly $(\pi\Gamma)\cdot\alpha^{-1}(v,\pi\Gamma(u))=(\pi\Gamma(a_k),\pi\Gamma(a_{k-1}),\ldots,\pi\Gamma(a_2))$. Therefore, $n_c(\sigma(\gamma\pi)^{-1})=n_c(\sigma\pi^{-1})+2$. ■

Lemma 4.5. *Let π and σ be two different chromosomes. For the SoRT(1,2) problem, let Φ be a minimum weighted sequence of reversals and block-interchanges needed to transform π into σ. Then the weight of Φ is greater than or equal to $\frac{|E|-n_c(\sigma\pi^{-1})}{2}$.*

Proof. Let Φ contain x reversals and y block-interchanges. Clearly, the weight of Φ is $x+2y$. As discussed previously, a reversal can by expressed by a product of two 2-cycles and a block-interchange by a product of four 2-cycles. Therefore, Φ can be written as a product of $2x+4y$ 2-cycles such that $\Phi\pi=\sigma$, equivalently meaning that $\sigma\pi^{-1}$ can be expressed by a product of $2x+4y$ 2-cycles and hence $\|\sigma\pi^{-1}\|\leq 2x+4y$. By Lemma 2.1, we have $2x+4y\geq|E|-n_c(\sigma\pi^{-1})$ and, consequently, the weight of Φ is greater than or equal to $\frac{|E|-n_c(\sigma\pi^{-1})}{2}$. ■

Suppose that there are at least two elements a and b in a cycle α of $\sigma\pi^{-1}$ such that $(a,b)\not\|\pi$. Then, based on Lemma 4.4, we can find from $\sigma\pi^{-1}$ a reversal $\gamma_1=(v,\pi\Gamma(u))(u,\pi\Gamma(v))$ to transform π into $\gamma_1\pi$. We continue with $\gamma_1\pi$ in the same way and finally find a sequence of ϕ reversals $\gamma_1,\gamma_2,\ldots,\gamma_\phi$ until $\sigma(\gamma_\phi\gamma_{\phi-1}\ldots\gamma_1\pi)^{-1}$ has no cycle in which there are two elements a and b such that $(a,b)\not\|\pi$. In this situation, π has become $\tau=\gamma_\phi\gamma_{\phi-1}\ldots\gamma_1\pi$, which by Lemma 4.3, can be further transformed into σ just by a sequence of $\psi=\frac{|E|-n_c(\sigma\tau^{-1})}{4}$ block-interchanges. In other words, we can find directly from $\sigma\pi^{-1}$ a sequence of ϕ reversals and ψ block-interchanges to transform π into σ, with total weight $\phi+2\psi$ that equals to $\frac{|E|-n_c(\sigma\pi^{-1})}{2}$, because $n_c(\sigma\tau^{-1})=n_c(\sigma\pi^{-1})+2\phi$ by Lemma 4.4 and hence $\phi+2\psi=\phi+\frac{|E|-n_c(\sigma\tau^{-1})}{2}=\phi+\frac{|E|-n_c(\sigma\pi^{-1})-2\phi}{2}=\frac{|E|-n_c(\sigma\pi^{-1})}{2}$. Furthermore, by Lemma 4.5, this sequence of ϕ reversals and ψ block-interchanges is optimal.

Take $\pi=(-5,3,2,4,-1)(1,-4,-2,-3,5)$ and $\sigma=(1,2,3,4,5)(-5,-4,-3,-2,-1)$ for an example. $\sigma\pi^{-1}=(1)(-5)(2,\ 4,\ 3,\ -4)(-1,\ 5,\ -2,\ -3)$ and clearly $n_c(\sigma\pi^{-1})=4$. By the discussion above, we immediately understand that there is a minimum weighted sequence of reversals and block-interchanges, whose total weight is $\frac{|E|-n_c(\sigma\pi^{-1})}{2}=\frac{10-4}{2}=3$, to transform π into σ. The optimal scenario is as follows. First, by Lemma 4.4, we can find a reversal $\gamma_1=(3,\ -4)(-1,\ 5)$ of π from $\sigma\pi^{-1}=(2,\ 4,\ 3)(3,\ -4)(-1,\ -2,\ -3)(-1,\ 5)$. After applying γ_1 to π, we obtain a new $\pi=(1,\ 3,\ 2,\ 4,\ 5)(-5,\ -4,\ -2,\ -3,\ -1)$ with $\sigma\pi^{-1}=(2,\ 4,\ 3)(-1,\ -2,\ -3)$. Then by Lemma 4.3, we can find a block-interchange $\beta_1\beta_1'$ from $\sigma\pi^{-1}$ that now is $(2,\ 4,\ 3)(-1,\ -2,\ -3)=(2,\ 3)(2,\ 4)(-2,\ -3)(-1,\ -3)$ by letting $\beta_1=(2,\ 3)(2,\ 4)$ and $\beta_1'=(-2,\ -3)(-1,\ -3)$ such that $\beta_1\beta_1'\pi=(1,2,3,4,5)(-5,\ -4,\ -3,\ -2,\ -1)=\sigma$. In summary, we use a reversal and a block-interchange to complete the above transformation. Based on the idea above, we have designed Algorithm 1 (SoRT; short for Sorting by Reversals and generalized Transpositions), which can be easily implemented using any program languages, to compute the *weighted genome rearrangement distance*, denoted by $\omega(\pi,\sigma)$, between two given chromosomes π and σ, and also generate an optimal scenario of the required rearrangement events.

Algorithm 1. SoRT

Input: Two chromosomes $\pi=\pi_1\pi_2$ and $\sigma=\sigma_1\sigma_2$.
Output: Weighted rearrangement distance $\omega(\pi,\ \sigma)$ and an optimal scenario Φ of operations.

1: Compute $\sigma\pi^{-1}$ and $\pi\Gamma$;
2: Let $\omega(\pi,\sigma)=\frac{|E|-n_c(\sigma\pi^{-1})}{2}$ and $\phi=0$;
3: **while** there are two elements a and b in a cycle α of $\sigma\pi^{-1}$ such that $(a,b)\not\|\pi$ **do**
3.1: Let $\phi=\phi+1$;
3.2: Find two adjacent elements a and b in α such that $(a,b)\not\|\pi$;
3.3: Let $\gamma_\phi=(\pi\Gamma(b),\pi\Gamma(a))(a,b)$;
3.4: Compute new $\pi=\gamma_\phi\pi$ and new $\pi\Gamma=\gamma_\phi\pi\Gamma$;
3.5: Obtain new $\sigma\pi^{-1}$ by removing b from α and $\pi\Gamma(a)$ from the mate cycle of α;
end while
4: Let $\psi=\frac{\omega(\pi,\sigma)-\phi}{2}$;
5: **for** $i=1$ to ψ **do**
5.1: Arbitrarily choose two adjacent elements a and b in a cycle of $\sigma\pi^{-1}$;
5.2: Find two adjacent elements c and d in $\sigma\pi^{-1}(a,b)$ such that $(c,d)\not\|(a,b)\pi$;
5.3: Let $\beta_i=(c,\ d)(a,\ b)$ and $\beta_i'=(\pi\Gamma(b),\pi\Gamma(a))(\pi\Gamma(d),\pi\Gamma(c))$;
5.4: Compute new $\pi=\beta_i\beta_i'\pi$ and new $\pi\Gamma=\beta_i\beta_i'\pi\Gamma$;
5.5: Obtain new $\sigma\pi^{-1}$ by removing $b,\ d,\ \pi\Gamma(c),\ \pi\Gamma(a)$ from the cycles in original $\sigma\pi^{-1}$;
end for
6: Output $\Phi=\{\gamma_1,\gamma_2,\ldots,\gamma_\phi,\beta_1\beta_1',\beta_2\beta_2',\ldots,\beta_\psi\beta_\psi'\}$;

Theorem 4.1. *Given two genomes π and σ, the SoRT(1,2) problem can be solved in $\mathcal{O}(\delta n)$ time, where δ is the number of reversals and block-interchanges needed to transform π into σ, and its weighted rearrangement distance is $\frac{|E|-n_c(\sigma\pi^{-1})}{2}$ that can be calculated in $\mathcal{O}(n)$ time.*

Proof. As we discussed previously, Algorithm 1 (SoRT) transforms π into σ using a minimum weighted sequence of ϕ reversals and ψ block-interchanges, whose total weight is $\frac{|E|-n_c(\sigma\pi^{-1})}{2}$ that clearly can be calculated in $\mathcal{O}(n)$ time. Hence, $\delta = \phi + \psi$. We analyze the time-complexity of Algorithm 1 (SoRT) as follows. Clearly, steps 1–2 can be done in $\mathcal{O}(n)$ time. For each iteration of step 3, its condition can be checked in worst-case $\mathcal{O}(n)$ time, because we only need to check every adjacent numbers a and b in each cycle of $\sigma\pi^{-1}$ and see whether $(a, b) \not\parallel \pi$, which can be verified in constant time according to Corollary 2.1. Then the execution time of step 3 is dominated by step 3.4 since the others need only constant time. Actually, each command of step 3.4 executes a join operation and a split operation, each of which can be done in constant time. Since step 3 is executed ϕ times and hence its total cost is $\mathcal{O}(\phi n)$. Step 4 is executed in constant time. As to step 5, it runs with ψ iterations and in each iteration, all substeps require only constant time, except for steps 5.2. Actually, step 5.2 can still be done in $\mathcal{O}(n)$ time in worst case. That is, the cost of step 5 is $\mathcal{O}(\psi n)$. The output of step 6 takes $\mathcal{O}(\delta)$ time. Consequently, the time-complexity of Algorithm 1 (SoRT) is $\mathcal{O}(\delta n)$. ■

Although the algorithm we presented above takes the circular chromosomes as the instances, it still works for the linear chromosomes, because, as described in the following theorem, it can be shown that the problem of sorting by reversals and block-interchanges is equivalent for circular and linear chromosomes, using a proof similar to that in Hartman and Sharan (2005). Basically, this proof is based on a property, that is, a reversal or block-interchange operating on a gene, say u, on a circular chromosome has an equivalent one that does not operate on u. This property for reversals can be verified by considering an example as shown in Figure 2. The effect of this reversal γ is the interchange between blocks A_1 in π_1 and A_2 in π_2, which equals to interchange $\pi_1 \backslash A_1$ and $\pi_2 \backslash A_2$ with each other. Similarly, as exemplified in Figure 3, the block-interchange β operating on a gene x on a circular chromosome π has also an equivalent block-interchange without operating on x.

Theorem 4.2. *The problem of sorting linear chromosomes by reversals and block-interchanges is equivalent to that of sorting circular chromosomes by reversals and block-interchanges.*

5. ALGORITHMS FOR SORTING BY WEIGHTED REVERSALS, BLOCK-INTERCHANGES, TRANSLOCATIONS

As mentioned early, translocations and fusions are not distinguishable for circular chromosomes (Alekseyev and Pevzner, 2008; Alekseyev, 2008), and hence, the problem of sorting a circular, multi-chromosomal genome Π into another circular, multi-chromosomal genome Σ by reversals, block-interchanges, and translocations (including fusions and fissions) is equivalent to that of sorting by reversals, block-interchanges, fusions, and fissions. Suppose that block-interchanges are weighted 2 and the others are all weighted 1. Then according to our previous work in Lu et al. (2006), it is not hard to show that in this case there is an optimal scenario of events to transform Π into Σ in the so-called *canonical order*, in which all fusions come before all reversals/block-interchanges, which then come before all fissions. In addition, it can be shown that we can derive a minimum series of 2-cycles from $\Sigma\Pi^{-1}$, acting on Π as fusions, to transform Π into Π', and then derive a minimum series of 2-cycle from $\Pi'\Sigma^{-1}$, acting on Σ as fusions, to transform Σ into Σ' (conversely, these fusions become fissions for transforming Σ' into Σ), and finally derive a minimum weighted series of reversals and block-interchanges from $\Sigma'\Pi'^{-1}$ to transform Π' into Σ' using Algorithm 1 (SoRT) we presented in the previous section. All of the above procedures can actually be done in $\mathcal{O}(\delta n)$ time, where δ is the number of fusions, reversals, block-interchanges and fissions needed to transform Π into Σ. Therefore, we have the following theorem immediately.

Theorem 5.1. *Given two circular, multi-chromosomal genomes Π and Σ, the problem of sorting Π into Σ by using a minimum weighted sequence of reversals, block-interchanges and fusions (or, equivalently, translocations) and fissions can be solved in $\mathcal{O}(\delta n)$ time, where block-interchanges are weighted 2 and the others are weighted 1.*

For linear chromosomes, fusions and fissions can be considered as special cases of translocation. In the following, therefore, we shall consider the SoRT^2(1,2,1) problem, which aims to sort a linear, multi-chromosomal genome Π into another Σ by reversals, block-interchanges and translocations (including fusions and fissions) with a weight proportion 1:2:1, and present an efficient and easily implemented algorithm to solve this problem.

Let $\Pi = \{\pi^1, \pi^2, \ldots, \pi^M\}$ and $\Sigma = \{\sigma^1, \sigma^2, \ldots, \sigma^N\}$ be two linear, multi-chromosomal genomes defined on the same set E of genes, where M and N denote the numbers of chromosomes in Π and Σ, respectively, and, without loss of generality, we assume that $M \geq N$. Let m_i and n_j be the number of genes in π^i and σ^j, respectively, where $1 \leq i \leq M$ and $1 \leq j \leq N$. Recall that in permutation group formalism, we represent a chromosome, say π^i, by the product $\pi^i_1\pi^i_2$ of its two strands, where $\pi^i_1 = (\pi^i_1(1), \pi^i_1(2), \ldots, \pi^i_1(m_i))$ and $\pi^i_2 = (\pi^i_2(1), \pi^i_2(2), \ldots, \pi^i_2(m_i))$. By following the convention, we call $\pi^i_1(1)$ and $\pi^i_2(1)$ as *tails* of π^i. Let $C = \{c_k = n + k + 1 : 0 \leq k \leq 2M - 1\}$ be a set of $2M$ distinct positive integers, called *caps*, which are different from genes in E. Let $\widehat{E} = E \cup \{\pm c_k : 0 \leq k \leq 2M - 1\}$ and $\widehat{\Gamma} = (1, -1)(2, -2)\ldots(n + 2M, -n - 2M)$. In fact, these caps are introduced to serve as chromosome delimiters when we use permutation group to model translocations of multiple linear chromosomes later. For this purpose, we first extend genome Σ by adding $M - N$ null chromosomes, resulting in $\Sigma = \{\sigma^1, \sigma^2, \ldots, \sigma^M\}$ that contains the same number of chromosomes as Π does. Then we obtain a capping genome $\widehat{\Pi} = \{\hat{\pi}^1, \hat{\pi}^2, \ldots, \hat{\pi}^M\}$ from Π by adding caps to the ends of each chromosome π^i, where $1 \leq i \leq M$, such that $\hat{\pi}^i_1 = (\hat{\pi}^i_1(1), \hat{\pi}^i_1(2), \ldots, \hat{\pi}^i_1(m_i + 2)) = (c_{2(i-1)}, \pi^i_1(1), \ldots, \pi^i_1(m_i), c_{2(i-1)+1})$ and $\hat{\pi}^i_2 = (\hat{\pi}^i_2(1), \hat{\pi}^i_2(2), \ldots, \hat{\pi}^i_2(m_i + 2)) = (\widehat{\Gamma}(c_{2(i-1)+1}), \pi^i_2(1), \ldots, \pi^i_2(m_i), \widehat{\Gamma}(c_{2(i-1)}))$. Using the same way to cap Σ, we have $\widehat{\Sigma} = \{\hat{\sigma}^1, \hat{\sigma}^2, \ldots, \hat{\sigma}^M\}$ with $\hat{\sigma}^i_1 = (\hat{\sigma}^i_1(1), \hat{\sigma}^i_1(2), \ldots, \hat{\sigma}^i_1(n_i + 2)) = (c_{2(i-1)}, \sigma^i_1(1), \ldots, \sigma^i_1(n_i), c_{2(i-1)+1})$ and $\hat{\sigma}^i_2 = (\hat{\sigma}^i_2(1), \hat{\sigma}^i_2(2), \ldots, \hat{\sigma}^i_2(n_i + 2)) = (\widehat{\Gamma}(c_{2(i-1)+1}), \sigma^i_2(1), \ldots, \sigma^i_2(n_i), \widehat{\Gamma}(c_{2(i-1)}))$. Since a single stranded DNA sequence is always written in the $5' \rightarrow 3'$ direction in biology, we call above $\hat{\pi}^i_1(1), \hat{\pi}^i_2(1), \hat{\sigma}^i_1(1)$ and $\hat{\sigma}^i_2(1)$ as $5'$ *caps* and the others as $3'$ *caps*. Since the $5'$ caps are considered as tails in the capping genomes, $\widehat{\Pi}$ and $\widehat{\Sigma}$ above are clearly *co-tailed*, which means to have the same set of tails, even though their original Π and Σ may not be.

Given a signed number $x \in \widehat{E}$, we use $\mathsf{char}(x, \widehat{\Pi})$ to define its character in a capping chromosome, say $\hat{\pi}^i$, in $\widehat{\Pi}$ as follows:

$$\mathsf{char}(x, \widehat{\Pi}) = \begin{cases} \text{C5}, & \text{if } x = \hat{\pi}^i_1(1) \text{ or } x = \hat{\pi}^i_2(1) \\ & \text{(i.e., } x \text{ serves as a } 5' \text{ cap of } \pi^i\text{).} \\ \text{C3}, & \text{if } (x = \hat{\pi}^i_1(m_i + 2) \text{ or } x = \hat{\pi}^i_2(m_i + 2)) \text{ and } \hat{\pi}^i \text{ is not null} \\ & \text{(i.e., } x \text{ serves as a } 3' \text{ cap of non-null } \pi^i\text{).} \\ \text{N3}, & \text{if } (x = \hat{\pi}^i_1(m_i + 2) \text{ or } x = \hat{\pi}^i_2(m_i + 2)) \text{ and } \hat{\pi}^i \text{ is null} \\ & \text{(i.e., } x \text{ serves as a } 3' \text{ cap of null } \pi^i\text{).} \\ \text{T}, & \text{if } x = \hat{\pi}^i_1(2) \text{ or } x = \hat{\pi}^i_2(2) \\ & \text{(i.e., } x \text{ serves as a tail of } \pi^i\text{).} \\ \text{O}, & \text{otherwise.} \end{cases}$$

According to Lemma 3.5, a translocation $\hat{\tau}$ acting on the capping $\widehat{\Pi}$ of the linear genome Π can be mimicked by using four 2-cycles in permutation group formalism, and the 2-cycles of $\hat{\tau}$ functioning as join operators must not be a (C5, C5) character pair and the 2-cycles functioning as split operators must be. In addition, if both of the character pairs for the 2-cycles of join operators are in {(C3, C3), (C3, N3), (T, T), (T, N3), (N3, N3)}, then the effect of the translocation $\hat{\tau}$ on $\widehat{\Pi}$ equals to the exchange of its caps, consequently leaving Π unaffected. Particularly, if they are both (T, C3) (respectively, (O, N3)), then the translocation on $\widehat{\Pi}$ corresponds to a fusion (respectively, fission) on Π. Intriguingly, we can use an internal translocation acting on $\widehat{\Pi}$ to mimic a translocation/fusion/fission on Π, where a translocation is called *internal* if it is neither a fusion nor a fission (Hannenhalli and Pevzner, 1995). In addition, we can use an internal reversal/block-interchange on $\widehat{\Pi}$ to mimic a reversal/block-interchange on Π, where a reversal/block-interchange is called *internal* if it does not involve the ends of the capped chromosome (Hannenhalli and Pevzner, 1995). For simplicity, therefore, the word "genome" mentioned in the rest of this section refers to a linear genome and "translocation/reversal/block-interchange" acting on a capping genome refers to an internal translocation/reversal/block-interchange.

Based on the properties described above, we can design an efficient algorithm as detailed below in Algorithm 2 (SoRT^2) to solve the problem of sorting capping, linear, multi-chromosomal genomes by reversals, block-interchanges, and translocations whose weights are 1, 2, and 1, respectively. For simplicity, we let CEpair = {(C3, C3), (C3, N3), (T, T), (T, N3), (N3, N3)}, TLpair = {(O, O), (O, C3), (O, T), (O, N3), (T, C3)} and, for a signed number x, define $\mathsf{5cap}(x)$ to be the signed number in the 5′ cap of the chromosome strand containing x.

Algorithm 2. SoRT^2

Input: Two linear genomes $\Pi = \{\pi^1, \pi^2, \ldots, \pi^M\}$ and $\Sigma = \{\sigma^1, \sigma^2, \ldots, \sigma^N\}$, where $M \geq N$.
Output: Weighted rearrangement distance $\omega(\Pi, \Sigma)$ and an optimal scenario Φ of operations.

1: Extend Σ by adding $M - N$ null chromosomes;
Obtain $\widehat{\Pi} = \{\hat{\pi}^1, \hat{\pi}^2, \ldots, \hat{\pi}^M\}$ and $\widehat{\Sigma} = \{\hat{\sigma}^1, \hat{\sigma}^2, \ldots, \hat{\sigma}^M\}$ by capping Π and Σ;
2: Compute $\widehat{\Sigma}\widehat{\Pi}^{-1}$ and $\widehat{\Pi}\widehat{\Gamma}$;
3: Let $n_\chi = n_\tau = n_\gamma = n_\beta = 0$;
4: **/* Preprocessing step for cap exchange */**
while there are a and b in a cycle of $\widehat{\Sigma}\widehat{\Pi}^{-1}$ (i.e., $(a, b)|\widehat{\Sigma}\widehat{\Pi}^{-1}$) such that $(\mathsf{char}(a, \widehat{\Pi}), \mathsf{char}(b, \widehat{\Pi})) \in$ CEpair **do**
4.1: Let $n_\chi = n_\chi + 1$;
4.2: Find a and b in a cycle of $\widehat{\Sigma}\widehat{\Pi}^{-1}$ with $(\mathsf{char}(a, \widehat{\Pi}), \mathsf{char}(b, \widehat{\Pi})) \in$ CEpair;
4.3: Let $\chi_{n_\chi} = (\mathsf{5cap}(a), \mathsf{5cap}(b))(a, b)(\widehat{\Pi}\widehat{\Gamma}(\mathsf{5cap}(b)), \widehat{\Pi}\widehat{\Gamma}(\mathsf{5cap}(a)))(\widehat{\Pi}\widehat{\Gamma}(b), \widehat{\Pi}\widehat{\Gamma}(a))$;
4.4: Compute new $\widehat{\Pi} = \chi_{n_\chi}\widehat{\Pi}$, new $\widehat{\Pi}\widehat{\Gamma} = \chi_{n_\chi}\widehat{\Pi}\widehat{\Gamma}$ and new $\widehat{\Sigma}\widehat{\Pi}^{-1} = \widehat{\Sigma}\widehat{\Pi}^{-1}\chi_{n_\chi}^{-1}$;
end while
5: **/* To derive translocations (including fusions and fissions) */**
while there are a and b in a cycle of $\widehat{\Sigma}\widehat{\Pi}^{-1}$ such that $(a, b)\not|\widehat{\Pi}$, $(a, \widehat{\Gamma}(b))\not|\widehat{\Pi}$ and $(\mathsf{char}(a, \widehat{\Pi}), \mathsf{char}(b, \widehat{\Pi})) \in$ TLpair **do**
/* Note that $(a, b)\not|\widehat{\Pi}$ and $(a, \widehat{\Gamma}(b))\not|\widehat{\Pi}$ mean that a and b are in the different chromosomes in $\widehat{\Pi}$. */
5.1: Let $n_\tau = n_\tau + 1$;
5.2: Find two adjacent elements a and b in a cycle of $\widehat{\Sigma}\widehat{\Pi}^{-1}$ such that $(a, b)\not|\widehat{\Pi}$, $(a, \widehat{\Gamma}(b))\not|\widehat{\Pi}$ and $(\mathsf{char}(a, \widehat{\Pi}), \mathsf{char}(b, \widehat{\Pi})) \in$ TLpair;
5.3: Let $\tau_{n_\tau} = (\mathsf{5cap}(a), \mathsf{5cap}(b))(a, b)(\widehat{\Pi}\widehat{\Gamma}(\mathsf{5cap}(b)), \widehat{\Pi}\widehat{\Gamma}(\mathsf{5cap}(a)))(\widehat{\Pi}\widehat{\Gamma}(b), \widehat{\Pi}\widehat{\Gamma}(a))$;
5.4: Compute new $\widehat{\Pi} = \tau_{n_\tau}\widehat{\Pi}$, new $\widehat{\Pi}\widehat{\Gamma} = \tau_{n_\tau}\widehat{\Pi}\widehat{\Gamma}$ and new $\widehat{\Sigma}\widehat{\Pi}^{-1} = \widehat{\Sigma}\widehat{\Pi}^{-1}\tau_{n_\tau}^{-1}$;
end while
6: **/* To derive reversals */**
while there are a and b in a cycle of $\widehat{\Sigma}\widehat{\Pi}^{-1}$ such that $(a, \widehat{\Gamma}(b))|\widehat{\Pi}$ **do**
/* Note that $(a, \widehat{\Gamma}(b))|\widehat{\Pi}$ means that a and b are in the same chromosome but different chromosome strands in $\widehat{\Pi}$. */
6.1: Let $n_\gamma = n_\gamma + 1$;
6.2: Find two adjacent elements a and b in a cycle of $\widehat{\Sigma}\widehat{\Pi}^{-1}$ such that $(a, \widehat{\Gamma}(b))|\widehat{\Pi}$;
6.3: Let $\gamma_{n_\gamma} = (\widehat{\Pi}\widehat{\Gamma}(b), \widehat{\Pi}\widehat{\Gamma}(a))(a, b)$;
6.4: Compute new $\widehat{\Pi} = \gamma_{n_\gamma}\widehat{\Pi}$, new $\widehat{\Pi}\widehat{\Gamma} = \gamma_{n_\gamma}\widehat{\Pi}\widehat{\Gamma}$ and new $\widehat{\Sigma}\widehat{\Pi}^{-1} = \widehat{\Sigma}\widehat{\Pi}^{-1}\gamma_{n_\gamma}^{-1}$;
end while
7: **/* To derive block-interchanges */**
while $\widehat{\Sigma}\widehat{\Pi}^{-1} \neq \mathbf{1}$ **do**
7.1: Let $n_\beta = n_\beta + 1$;
7.2: Arbitrarily choose two adjacent elements a and b in a cycle of $\widehat{\Sigma}\widehat{\Pi}^{-1}$;
7.3: Find two adjacent elements c and d in a cycle of $\widehat{\Sigma}\widehat{\Pi}^{-1}(a, b)$ such that $(c, d)\not|(a, b)\widehat{\Pi}$;
7.4: Let $\beta_{n_\beta} = (c, d)(a, b)(\widehat{\Pi}\widehat{\Gamma}(b), \widehat{\Pi}\widehat{\Gamma}(a))(\widehat{\Pi}\widehat{\Gamma}(d), \widehat{\Pi}\widehat{\Gamma}(c))$;
7.5: Compute new $\widehat{\Pi} = \beta_{n_\beta}\widehat{\Pi}$, new $\widehat{\Pi}\widehat{\Gamma} = \beta_{n_\beta}\widehat{\Pi}\widehat{\Gamma}$ and new $\widehat{\Sigma}\widehat{\Pi}^{-1} = \widehat{\Sigma}\widehat{\Pi}^{-1}\beta_{n_\beta}^{-1}$;
end while
8: Output $\omega(\Pi, \Sigma) = n_\tau + n_\gamma + 2 \times n_\beta$ and $\Phi = \{\tau_1, \ldots, \tau_{n_\tau}, \gamma_1, \ldots, \gamma_{n_\gamma}, \beta_1, \ldots, \beta_{n_\beta}\}$;

Now, we demonstrate Algorithm 2 (SoRT^2) by letting $\Pi = (-5, 3)(-3, 5)(2, 4, -1)(1, -4, -2)$ and $\Sigma = (1, 2, 3, 4, 5)(-5, -4, -3, -2, -1)$. Initially, we first add a null chromosome into Σ and then derive

from Π and Σ the capping genomes $\widehat{\Pi} = (6, -5, 3, 7)(-7, -3, 5, -6)(8, 2, 4, -1, 9)$ $(-9, 1, -4, -2, -8)$ and $\widehat{\Sigma} = (6, 1, 2, 3, 4, 5, 7)(-7, -5, -4, -3, -2, -1, -6)(8, 9)(-9, -8)$, respectively. Then we calculate $\widehat{\Sigma}\widehat{\Pi}^{-1} = (1, -8, -1, 5, -2, -3, -5)(-6, 7, 4, 3, -4, 2, 9)$, in which we can find -6 and 7 in a cycle with $(\mathsf{char}(-6, \widehat{\Pi}), \mathsf{char}(7, \widehat{\Pi})) = (\mathrm{C3}, \mathrm{C3})$. Based on the step 4 in Algorithm 2 (SoRT^2), there is an operation $\chi_1 = (-7, 6)(-6, 7)(-7, 6)(-3, -5)$ performing on $\widehat{\Pi}$ as a cap exchange. After applying χ_1 to $\widehat{\Pi}$, we can obtain a new capping genome $\widehat{\Pi} = (-7, -5, 3, -6)(6, -3, 5, 7)(8, 2, 4, -1, 9)(-9, 1, -4, -2, -8)$. In new $\widehat{\Sigma}\widehat{\Pi}^{-1}$ that is now $(1, -8, -1, 5, -2, -3)(-6, 4, 3, -4, 2, 9)$, we can still find another cap exchange operation $\chi_2 = (8, -7)(9, -6)(6, -9)(-3, 1)$ that transforms $\widehat{\Pi}$ into $(-7, -5, 3, 9)(-9, -3, 5, 7)(8, 2, 4, -1, -6)(6, 1, -4, -2, -8)$. Then we have new $\widehat{\Sigma}\widehat{\Pi}^{-1} = (-8, -1, 5, -2, -3)(4, 3, -4, 2, 9)(6, -9)(8, -7)$, in which there are -3 and -8 in a cycle with $(\mathsf{char}(-3, \widehat{\Pi}), \mathsf{char}(-8, \widehat{\Pi})) = (\mathrm{T}, \mathrm{C3})$ that clearly are in the different chromosomes in $\widehat{\Pi}$. Then based on the step 5, we can find a translocation $\tau_1 = (-9, 6)(-3, -8)(8, -7)(2, 9)$ that is a fusion actually and transforms $\widehat{\Pi}$ into $(-7, -5, 3, 2, 4, -1, -6)(6, 1, -4, -2, -3, 5, 7)(8, 9)(-9, -8)$, indicating that there is only a chromosome $(-5, 3, 2, 4, -1)(1, -4, -2, -3, 5)$ in the current Π whose gene content clearly equals to that of Σ. As was demonstrated in the previous section, this single chromosomal genome Π can be further transformed into Σ by using a reversal, followed by a block-interchange. In fact, these two intra-chromosomal operations can be derived as follows. Now, $\widehat{\Sigma}\widehat{\Pi}^{-1} = (2, 4, 3, -4)(-1, 5, -2, -3)$, in which -1 and 5 are in the same cycle but in the different chromosome strand (i.e, $(-1, \widehat{\Gamma}(5))|\widehat{\Pi}$), where $\widehat{\Gamma}(5) = -5$. Therefore, based on the step 6, we have $\gamma_1 = (3, -4)(-1, 5)$ that transforms $\widehat{\Pi}$ into $(-7, -5, -4, -2, -3, -1, -6)(6, 1, 3, 2, 4, 5, 7)$ $(8, 9)(-9, -8)$, which leads to new $\widehat{\Sigma}\widehat{\Pi}^{-1} = (2, 4, 3)(-1, -2, -3)$. According to the step 7, we can find a block-interchange $\beta_1 = (2, 3)(2, 4)(-2, -3)(-1, -3)$ that transforms $\widehat{\Pi}$ into $(-7, -5, -4, -3, -2, -1, -6)(6, 1, 2, 3, 4, 5, 7)$ that is equal to $\widehat{\Sigma}$.

Basically, Algorithm 2 (SoRT^2) is a greedy method, in which step 4 can be considered just as a preprocessing procedure that aims to exchange caps between chromosomes, and steps 5, 6 and 7 are to derive inter-chromosomal translocations, intra-chromosomal reversals and intra-chromosomal block-interchanges, respectively. As was demonstrated in the above example, we can express $\widehat{\Sigma}\widehat{\Pi}^{-1}$ as a product of 2-cycles that functions as a sequence of translocations, reversals and block-interchanges for optimally transforming Π into Σ. In total, Algorithm 2 (SoRT^2) derives n_χ cap exchange operations $\chi_1, \chi_2, \ldots, \chi_{n_\chi}$ and $n_\tau + n_\gamma + n_\beta$ rearrangement operations consisting of n_τ translocations $\tau_1, \tau_2 \ldots, \tau_{n_\tau}, n_\gamma$ reversals $\gamma_1, \gamma_2, \ldots, \gamma_{n_\gamma}$, and n_β block-interchanges $\beta_1, \beta_2, \ldots, \beta_{n_\beta}$. To prove the correctness of Algorithm 2 (SoRT^2), we first show that the output Φ of Algorithm 2 (SoRT^2) is a feasible solution to the problem and then continue to show that this feasible solution is optimal. For simplicity, we let $\widehat{\Phi} = \{\phi_1, \phi_2, \ldots, \phi_\Delta\} = \{\chi_1, \ldots, \chi_{n_\chi}, \tau_1, \ldots, \tau_{n_\tau}, \gamma_1, \ldots, \gamma_{n_\gamma}, \beta_1, \ldots, \beta_{n_\beta}\}$ and let $\widehat{\Pi}_0 = \widehat{\Pi}$ and $\widehat{\Pi}_i = \phi_i\widehat{\Pi}_{i-1}$, where $\Delta = n_\chi + n_\tau + n_\gamma + n_\beta$ and $1 \le i \le \Delta$. Since an internal reversal/block-interchange/translocation does not change the set of tails of a capping genome, we have the following observation immediately.

Observation 5.1. *Genomes $\widehat{\Pi}_0, \widehat{\Pi}_1, \ldots, \widehat{\Pi}_\Delta$ are all co-tailed.*

The above observation indicates the fact that for a signed number $x \in \widehat{E}$, if x is characterized as a $5'$ (respectively, $3'$) cap in $\widehat{\Pi}_0$, then it is also characterized as a $5'$ (respectively, $3'$) cap in $\widehat{\Pi}_i$, suggesting that once a $5'/3'$ cap, always a $5'/3'$ cap.

Observation 5.2. *Let x be a $5'$ cap of a chromosome in $\widehat{\Pi}_0$ (i.e., $\mathsf{char}(x, \widehat{\Pi}_0) = C5$). Then x is fixed in $\widehat{\Sigma}\widehat{\Pi}_0^{-1}$.*

The following five lemmas are essential for designing our algorithm, and their proofs can be found in Appendix B.

Lemma 5.1. *For $0 \le i \le n_\chi + n_\tau$, there are no two elements a and b with $\mathsf{char}(a, \widehat{\Pi}_i) = C5$ and $\mathsf{char}(b, \widehat{\Pi}_i) \neq C5$ such that a and b are both in the same cycle of $\widehat{\Sigma}\widehat{\Pi}_i^{-1}$ (i.e., $(a, b)|\widehat{\Sigma}\widehat{\Pi}_i^{-1}$).*

Lemma 5.2. *Given a capping genome $\widehat{\Pi}$ and a signed number $x \in \widehat{E}$, if $\mathsf{char}(x, \widehat{\Pi})$ is C3 (respectively, T), then $\mathsf{char}(\widehat{\Pi}\widehat{\Gamma}(x), \widehat{\Pi})$ is T (respectively, C3) and if $\mathsf{char}(x, \widehat{\Pi})$ is O (respectively, N3 and C5), then $\mathsf{char}(\widehat{\Pi}\widehat{\Gamma}(x), \widehat{\Pi})$ is O (respectively, N3 and C5).*

According to Lemma 5.2, it can be observed that if we can derive from $\widehat{\Sigma}\widehat{\Pi}_i^{-1}$ four 2-cycles, where $n_\chi \le i \le n_\chi + n_\tau$ and each of these four 2-cycles divides $\widehat{\Sigma}\widehat{\Pi}_i^{-1}$, that act on $\widehat{\Pi}_i$ as a cap exchange, then there must be a 2-cycle (a, b) among these four 2-cycles such that $\mathsf{char}(a, \widehat{\Pi}_i) \in \{\mathrm{C3}, \mathrm{N3}\}$ and $\mathsf{char}(b, \widehat{\Pi}_i) \in \{\mathrm{C3}, \mathrm{N3}\}$. This indicates that there is at least a cycle in $\widehat{\Sigma}\widehat{\Pi}_i^{-1}$ that contains the two elements a and b with $\mathsf{char}(a, \widehat{\Pi}_i) \in \{\mathrm{C3}, \mathrm{N3}\}$ and $\mathsf{char}(b, \widehat{\Pi}_i) \in \{\mathrm{C3}, \mathrm{N3}\}$. Conversely, if there is no cycle in $\widehat{\Sigma}\widehat{\Pi}_i^{-1}$ that contains two elements a and b with $\mathsf{char}(a, \widehat{\Pi}_i) \in \{\mathrm{C3}, \mathrm{N3}\}$ and $\mathsf{char}(b, \widehat{\Pi}_i) \in \{\mathrm{C3}, \mathrm{N3}\}$, then we cannot derive a cap exchange from $\widehat{\Sigma}\widehat{\Pi}_i^{-1}$.

Lemma 5.3. *For $n_\chi \le i \le n_\chi + n_\tau$, there are no two elements a and b in the same cycle of $\widehat{\Sigma}\widehat{\Pi}_i^{-1}$ such that* $(\mathsf{char}(a, \widehat{\Pi}_i), \mathsf{char}(b, \widehat{\Pi}_i)) \in$ CEpair.

Lemma 5.4. *$m = n_\chi + n_\tau$ and let a and b be any two non-C5 elements of $\widehat{\Pi}_m$ that are both in a cycle of $\widehat{\Sigma}\widehat{\Pi}_m^{-1}$. Then a and b are on the same chromosome in $\widehat{\Pi}_m$.*

Lemma 5.5. *Let $m = n_\chi + n_\tau$. For each chromosome $\hat{\pi}^i$ in $\widehat{\Pi}_m$, there is a corresponding chromosome $\hat{\sigma}^j$ in $\widehat{\Sigma}$ such that the gene content of $\hat{\pi}^i$ equals to that of $\hat{\sigma}^j$, where $1 \le i, j \le M$.*

Corollary 5.1. *Let $m = n_\chi + n_\tau$. For each uncapping chromosome π^i in $\widehat{\Pi}_m$, there is a corresponding, uncapping chromosome σ^j in $\widehat{\Sigma}$ such that the gene content of π^i equals to that of σ^j, where $1 \le i, j \le M$. In addition, all the C5 elements of $\widehat{\Pi}_m$ are fixed in $\widehat{\Sigma}\widehat{\Pi}_m^{-1}$.*

Proof. As was shown in Lemma 5.5, for each chromosome $\hat{\pi}^i$ in $\widehat{\Pi}_m$, there is a corresponding chromosome $\hat{\sigma}^j$ in $\widehat{\Sigma}$ such that they have the same gene content, where $1 \le i, j \le M$. Actually, their uncapping chromosomes π^i and σ^j still contain the same set of genes, because $\hat{\pi}^i$ and $\hat{\sigma}^j$ are co-tailed genomes. In addition, it can be verified that both of the C5 elements in $\hat{\pi}^i$ are fixed in $\widehat{\Sigma}\widehat{\Pi}_m^{-1}$. As a result, the C5 elements of $\widehat{\Pi}_m$ are all fixed in $\widehat{\Sigma}\widehat{\Pi}_m^{-1}$. ■

Based on Corollary 5.1, each uncapping chromosome π^i in $\widehat{\Pi}_m$ has a corresponding, uncapping chromosome σ^j in $\widehat{\Sigma}$ such that π^i and of σ^j have the same set of genes, where $1 \le i, j \le M$. As was already demonstrated in Algorithm 1 (SoRT), π^i can be transformed into σ^j using a series of reversals and block-interchanges, which serve as internal reversals and block-interchanges accordingly for transforming $\hat{\pi}^i$ into $\hat{\sigma}^j$, as implemented in the steps 6 and 7 of Algorithm 2 (SoRT^2), respectively. Therefore, the output Φ of Algorithm 2 (SoRT^2) is a feasible solution to the problem of sorting Π by reversals, block-interchanges and translocations (including fusions and fissions). In the following, we will show that Φ is also an optimal solution of the problem.

Lemma 5.6. *Given two linear genomes Π and Σ with multiple chromosomes, Algorithm 2 (SoRT^2) transforms Π into Σ by using a minimum weighted sequence of reversals, block-interchanges and translocations (including fusions and fissions).*

Proof. As demonstrated above, the output $\Phi = \{\phi_{m+1}, \phi_{m+2}, \ldots, \phi_\Delta\}$ of Algorithm 2 (SoRT^2) is a feasible solution of the problem, whose number of rearrangement operations is $n_f = n_\tau + n_\gamma + n_\beta$ and whose weight is $\omega_f = n_\tau + n_\gamma + 2n_\beta$. Let $\Phi_{opt} = \{\rho_1, \rho_2, \ldots, \rho_{n_o}\}$ be an optimal solution required to transform Π into Σ and its weight be denoted by ω_o. Clearly, we have $\omega_o \le \omega_f$. As mentioned before, each ρ_i, where $1 \le i \le n_o$, can be modeled by applying a corresponding permutation ρ_i' of two or four 2-cycles to the affected genome. Let $\widehat{\Theta}$ be the capping genome obtained from the capping $\widehat{\Sigma}$ of Σ by performing the sequence of operations $\rho_{n_o}', \rho_{n_o-1}', \ldots, \rho_1'$ (i.e., $\widehat{\Theta} = \rho_1'\rho_2' \ldots \rho_{n_o}'\widehat{\Sigma}$). Let $\widehat{\Theta} = \{\widehat{\theta}^1, \widehat{\theta}^2, \ldots, \widehat{\theta}^M\}$ and Θ be the uncapped genome of $\widehat{\Theta}$. It is clear that $\Theta = \Pi$ and $\widehat{\Theta}$ and $\widehat{\Pi}_{n_\chi}$ (obtained from $\widehat{\Pi}_0$ by n_χ cap exchanges) are co-tailed, indicating that we can transform $\widehat{\Pi}_{n_\chi}$ into $\widehat{\Theta}$ using a series of cap exchanges derived from $\widehat{\Theta}\widehat{\Pi}_{n_\chi}^{-1}$, in which all the numbers characterized as O elements in $\widehat{\Pi}_{n_\chi}$ are fixed, and there is no cycle containing both C5 and non-C5 elements simultaneously. Theoretically, we have $\widehat{\Sigma}\widehat{\Pi}_{n_\chi}^{-1} = \widehat{\Sigma}\widehat{\Theta}^{-1}\widehat{\Theta}\widehat{\Pi}_{n_\chi}^{-1}$. By Lemma 5.1, $\widehat{\Sigma}\widehat{\Pi}_{n_\chi}^{-1}$ has no cycle that contains both C5 and non-C5 elements of $\widehat{\Pi}_{n_\chi}$ at the same time. Since the above property holds for both $\widehat{\Sigma}\widehat{\Pi}_{n_\chi}^{-1}$ and $\widehat{\Theta}\widehat{\Pi}_{n_\chi}^{-1}$, it still holds for $\widehat{\Sigma}\widehat{\Theta}^{-1}$. To simplify the

following discussion, we denote by X, Y and Z, respectively, the products of all cycles of comprising non-C5 elements in $\widehat{\Sigma}\widehat{\Pi}_{n_\chi}^{-1}$, $\widehat{\Sigma}\widehat{\Theta}^{-1}$ and $\widehat{\Theta}\widehat{\Pi}_{n_\chi}^{-1}$ (i.e., ignoring the cycles consisting of C5 elements). Then it can be verified that $\omega_f = \|X\|/2$ according to the design of Algorithm 2 (SoRT^2). As mentioned above, $\widehat{\Theta} = \rho'_1\rho'_2 \ldots \rho'_{n_o}\widehat{\Sigma}$. For each $1 \le i \le n_o$, if ρ'_i is a translocation (whose weight is 1), then it can be expressed by two 2-cycles of (non-C5, non-C5) character pair and two 2-cycles of (C5, C5), while if ρ'_i is a reversal/block-interchange (whose weight is $1/2$), then it can be expressed by two/four 2-cycles of (non-C5, non-C5) character pair. Then the total number of the above 2-cycles of (non-C5, non-C5) is $2\omega_o$, and it should be greater than or equal to $\|Y\|$. In other words, $\omega_o \ge \|Y\|/2$ and hence $\|X\| \ge \|Y\|$ since $\omega_f \ge \omega_o$. Suppose that $\|X\| > \|Y\|$. Then we claim that there are two elements a and b in a cycle of $\widehat{\Sigma}\widehat{\Pi}_{n_\chi}^{-1}$ such that $(\mathsf{char}(a, \widehat{\Pi}_{n_\chi}), \mathsf{char}(b, \widehat{\Pi}_{n_\chi})) \in$ CEpair. (The correctness of this claim will be proved later.) However, this result clearly contradicts to Lemma 5.3. In other words, $\|X\| = \|Y\|$ and, therefore, Φ is an optimal solution.

Below, we prove the above claim by contradiction method. It should be noted that the characters of all elements we mention in the rest of this proof are with respect to $\widehat{\Pi}_{n_\chi}$. Suppose that there are no two elements a and b in a cycle of $\widehat{\Sigma}\widehat{\Pi}_{n_\chi}^{-1}$ such that $(\mathsf{char}(a, \widehat{\Pi}_{n_\chi}), \mathsf{char}(b, \widehat{\Pi}_{n_\chi})) \in$ CEpair. This indicates that for each cycle, say α, in X with containing at least a C3, N3 or T element, either this cycle has exactly a C3, N3 or T element, or it has exactly a C3 element and a T element. For the latter case, we let $\alpha = (a_1, a_2, \ldots, a_i, \ldots, a_j)$, where $(\mathsf{char}(a_i, \widehat{\Pi}_{n_\chi}), \mathsf{char}(a_j, \widehat{\Pi}_{n_\chi}))$ is equal to (C3, T) or (T, C3) and the others are all O elements. Then we can express α as a product of $(a_1, a_2, \ldots, a_i)(a_{i+1}, \ldots, a_j)(a_i, a_j)$ and $(a_i, a_j)|X$ by Lemma 2.7. Using this approach, we can express $\mathrm{X} = \mathrm{X}'\mathrm{X}''$, where each cycle in X' contains at most a C3, N3 or T element, and each cycle in X'' is a 2-cycle whose character pair is either (C3, T) or (T, C3). In addition, we have $X''|X$, since it can be verified that $n_c(XX''^{-1}) - n_c(X) = \|X''\|$ (and hence $X''|X$ by Corollary 2.1). Then $\|X\| = \|X'\| + \|X''\|$ by Lemma 2.5. As mentioned above, each O element x is fixed in Z (i.e., $Z(x) = x$) and hence $X(x) = YZ(x) = Y(x)$. This further suggests that for each cycle in X', say $\beta = (b_1, b_2, \ldots, b_j)$ where $\mathsf{char}(b_i, \widehat{\Pi}_{n_\chi}) = \mathrm{O}$ for $1 \le i \le j-1$, all of its elements appear consecutively in a cycle of Y in the order of $b_1, b_2, \ldots, b_j$. In other words, $\beta|Y$ by Lemma 2.7 and, moreover, $X'|Y$ since it can be verified that $n_c(YX'^{-1}) - n_c(Y) = \|X'\|$. Let $Y = X'Y''$ (i.e., $Y'' = X'^{-1}Y$). Then $\|Y\| = \|X'\| + \|Y''\|$ by Lemma 2.5. Based on the above assumption that $\|X\| > \|Y\|$, we have $\|X''\| > \|Y''\|$. For convenience, we let $\|X''\| = k$ and $X'' = (a_1, b_1)(a_2, b_2) \ldots (a_k, b_k)$, where $(\mathsf{char}(a_i, \widehat{\Pi}_{n_\chi}), \mathsf{char}(b_i, \widehat{\Pi}_{n_\chi})) = (\mathrm{C3}, \mathrm{T})$ for each $1 \le i \le k$. Let $Z(a_i) = c_i$. Then $\mathsf{char}(c_i, \widehat{\Pi}_{n_\chi})$ is either C3 or N3 (i.e., $\mathsf{char}(c_i, \widehat{\Pi}_{n_\chi}) \ne \mathrm{T}$) and hence $c_i \ne b_i$, which is due to the fact that $\Theta = \Pi$ and $\widehat{\Theta}$ and $\widehat{\Pi}_{n_\chi}$ are co-tailed. Since $X'' = Y''Z$, we have $b_i = X''(a_i) = Y''Z(a_i) = Y''(c_i)$, indicating that c_i and b_i are consecutive in a cycle in Y''. This further implies that $\|Y''\| \ge k$. As a result, $\|X''\| \le \|Y''\|$, which contradicts to that $\|X''\| > \|Y''\|$. In other words, there are two elements a and b in a cycle of $\widehat{\Sigma}\widehat{\Pi}_{n_\chi}^{-1}$ such that $(\mathsf{char}(a, \widehat{\Pi}_{n_\chi}), \mathsf{char}(b, \widehat{\Pi}_{n_\chi})) \in$ CEpair. ■

Theorem 5.2. *Given two linear, multi-chromosomal genomes* Π *and* Σ*, the SoRT*^{2}*(1, 2, 1) problem can be solved in* $\mathcal{O}(\delta n)$ *time, where* δ *is the number of reversals, block-interchanges and translocations* (*including fusions and fissions*) *needed to transform* Π *into* Σ.

Proof. According to Lemma 5.6, Algorithm 2 (SoRT^2) can transform Π into Σ using a minimum weighted sequence of reversals, block-interchanges and translocations (including fusions and fissions). Below, we analyze its time-complexity. It is clear that steps 1–2 can be done in $\mathcal{O}(n)$ time and step 3 in constant time. Suppose that $M \ge N$. Then there are exactly $2M$ C3 and N3 elements. As mentioned in the front of Lemma 5.3, there must be a 2-cycle in a cap exchange operation whose character pair is either (C3, C3), (C3, N3) or (N3, N3). This means that the composition of this cycle, its mate cycle and two additional (C5, C5) cycles comprises an operation of cap exchange, suggesting that the number of iterations in step 4 is at most M. Recall that once an element is a $3'$ cap, it is always a $3'$ cap in the whole process. Hence, we can first identify all the C3/N3 elements in the initial $\widehat{\Sigma}\widehat{\Pi}^{-1}$ by costing $\mathcal{O}(2n + 2M)$ time, where $M \le n$. Then by constant time, we can determine if we need to enter the iteration of step 4 to perform the cap exchange that actually requires only constant time. As a result, the total cost of step 4 is $\mathcal{O}(n)$. For each iteration of steps 5 and 6, the most time-consuming operation is to check and find if there is any two adjacent elements in each cycle of $\widehat{\Sigma}\widehat{\Pi}^{-1}$ that satisfy the required conditions, which totally can be done in $\mathcal{O}(n)$ time. As to step 7, its execution time is dominated by step 7.3 that can be finished in $\mathcal{O}(n)$ time in worst case. Step 8 needs only constant time. Consequently, based on the above discussion, if the number of needed reversals, block-interchanges and translocations in Algorithm 2 (SoRT^2) is δ, then the total time complexity of Algorithm 2 (SoRT^2) is $\mathcal{O}(\delta n)$. ■

Corollary 5.2. *Given two linear, multi-chromosomal genomes* Π *and* Σ*, their weighted rearrangement distance* $\omega(\Pi, \Sigma)$ *can be calculated in* $\mathcal{O}(n)$ *time.*

Proof. According to Algorithm 2 (SoRT^2), it can be realized that $\omega(\Pi, \Sigma) = \frac{\|\widehat{\Sigma}\widehat{\Pi}^{-1}\| - 2n_\chi}{2}$, where n_χ is the number of cap exchanges needed in Algorithm 2 (SoRT^2). Clearly, $\|\widehat{\Sigma}\widehat{\Pi}^{-1}\|$ can be computed in $\mathcal{O}(n)$ time by Lemma 2.1 and n_χ can also be calculated in $\mathcal{O}(n)$ time as shown in Theorem 5.2. Therefore, $\omega(\Pi, \Sigma)$ can be calculated in $\mathcal{O}(n)$ time. ■

6. CONCLUSION

In this article, we demonstrated that the permutation group formalism can be utilized to model genome rearrangements, such as reversals, (generalized) transpositions, and translocations (including fusions and fissions), and design novel algorithms, which can be easily implemented using simple data structure of 1-dimensional arrays, for efficiently sorting linear/circular, multi-chromosomal genomes. For future work, it would be interesting to discuss relative advantages and disadvantages of all the different formalisms (e.g., the permutation group and breakpoint/adjacent graphs) for solving the genome rearrangement problems with regard to both the theoretical and practical advances in the design of efficient algorithms.

7. APPENDIX A

Lemma 2.1. *For any permutation* α *of* E, $\|\alpha\| = |E| - n_c(\alpha)$.

Proof. We refer the reader to (Lin et al., 2005) for the detailed proof of this lemma. ■

Lemma 2.2. *Let* α *and* β *be any two permutations of* E*. Then* $\|\alpha \cdot \beta\| = \|\beta\|$.

Proof. As mentioned before, $\alpha \cdot \beta$ has the same cycle structure as β, meaning that $n_c(\alpha \cdot \beta) = n_c(\beta)$. By Lemma 2.1, $\|\alpha \cdot \beta\| = \|\beta\|$. ■

Lemma 2.3. *Let* α *and* β *be any two permutations of* E*. Then* $\|\alpha\beta\| = \|\beta\alpha\|$.

Proof. By Lemma 2.2, $\|\alpha\beta\| = \|\beta \cdot (\alpha\beta)\| = \|\beta\alpha\beta\beta^{-1}\| = \|\beta\alpha\|$. ■

Lemma 2.4. *Let* α *and* β *be any two permutations of* E*. Then* $\|\alpha\beta\| \leq \|\alpha\| + \|\beta\|$.

Proof. Let $\|\alpha\| = i$ and $\|\beta\| = j$. Then we can express α as a product of i 2-cycles, say $\alpha = a_1 a_2 \ldots a_i$ and β as a product of j 2-cycles, say $\beta = b_1 b_2 \ldots b_j$. Therefore, $\alpha\beta = a_1 a_2 \ldots a_i b_1 b_2 \ldots b_j$ and, consequently, $\|\alpha\beta\| \leq i + j = \|\alpha\| + \|\beta\|$. ■

Lemma 2.5. *Let* α*,* β *and* γ *be any three permutations of* E *and* $\alpha = \beta\gamma$*. If* $\beta|\alpha$ *or* $\gamma|\alpha$*, then* $\|\alpha\| = \|\beta\| + \|\gamma\|$.

Proof. Suppose that $\beta|\alpha$. Then $\|\alpha\beta^{-1}\| = \|\alpha\| - \|\beta\|$ by definition. Moreover, $\|\alpha\beta^{-1}\| = \|\beta^{-1}\alpha\|$ by Lemma 2.3 and $\|\beta^{-1}\alpha\| = \|\gamma\|$ since $\alpha = \beta\gamma$. Therefore, $\|\alpha\| = \|\beta\| + \|\gamma\|$. Suppose that $\gamma|\alpha$. Then $\|\alpha\gamma^{-1}\| = \|\alpha\| - \|\gamma\|$ by definition and $\|\alpha\gamma^{-1}\| = \|\beta\|$ since $\alpha = \beta\gamma$. Therefore, $\|\alpha\| = \|\beta\| + \|\gamma\|$. ■

Lemma 2.6. *Let* α*,*β *and* γ *be any three permutations of* E*. If* $\alpha|\beta$ *and* $\beta|\gamma$*, then* $\alpha|\gamma$.

Proof. Suppose that $\alpha|\beta$ and $\beta|\gamma$. Then $\|\beta\alpha^{-1}\| = \|\beta\| - \|\alpha\|$ and $\|\gamma\beta^{-1}\| = \|\gamma\| - \|\beta\|$, which means that $\|\beta\alpha^{-1}\| + \|\gamma\beta^{-1}\| = \|\gamma\| - \|\alpha\|$. Note that $\gamma\alpha^{-1} = \gamma\beta^{-1}\beta\alpha^{-1}$. By Lemma 2.4, we have $|\gamma\alpha^{-1}\| \leq \|\gamma\beta^{-1}\| + \|\beta\alpha^{-1}\|$, implying that $\|\gamma\alpha^{-1}\| \leq \|\gamma\| - \|\alpha\|$. Since $\gamma = \gamma\alpha^{-1}\alpha$, we have $\|\gamma\| = \|\gamma\alpha^{-1}\alpha\|$. By Lemma 2.4 again, $\|\gamma\| \leq \|\gamma\alpha^{-1}\| + \|\alpha\|$, meaning that $\|\gamma\alpha^{-1}\| \geq \|\gamma\| - \|\alpha\|$. Consequently, $\|\gamma\alpha^{-1}\| = \|\gamma\| - \|\alpha\|$ and, therefore, $\alpha|\gamma$. ■

8. APPENDIX B

Lemma 5.1. *For $0 \le i \le n_\chi + n_\tau$, there are no two elements a and b with* $\mathsf{char}(a, \widehat{\Pi}_i) = C5$ *and* $\mathsf{char}(b, \widehat{\Pi}_i) \neq C5$ *such that a and b are both in the same cycle of* $\widehat{\Sigma}\widehat{\Pi}_i^{-1}$ *(i.e.,* $(a, b) | \widehat{\Sigma}\widehat{\Pi}_i^{-1}$*).*

Proof. We prove this lemma by induction on i. Recall that once a signed number is a C5 element in $\widehat{\Pi}_0$, it is always a C5 element in $\widehat{\Pi}_j$, where $1 \le j \le n_\chi + n_\tau$. Basically, the lemma is true when $i = 0$, since each C5 element is fixed in $\widehat{\Sigma}\widehat{\Pi}_0^{-1}$ according to Observation 5.2. Suppose that this lemma holds for $i = k$, where $0 \le k < n_\chi + n_\tau$. Then each cycle in $\widehat{\Sigma}\widehat{\Pi}_i^{-1}$ either consists only of C5 or non-C5 elements. Note that $\widehat{\Pi}_{i+1} = \phi_{i+1}\widehat{\Pi}_i$ and, therefore, $\widehat{\Sigma}\widehat{\Pi}_{i+1}^{-1} = \widehat{\Sigma}\widehat{\Pi}_i^{-1}\phi_{i+1}^{-1}$. Since ϕ_{i+1} acts on $\widehat{\Pi}_i$ as a cap exchange or translocation, the (C5, C5)-cycle in ϕ_{i+1}^{-1} acting on $\widehat{\Sigma}\widehat{\Pi}_i^{-1}$ either divides a cycle with only C5 elements into two smaller cycles, or joins two cycles, each with C5 elements only, into a bigger one. As a result, there is no cycle in $\widehat{\Sigma}\widehat{\Pi}_{i+1}^{-1}$ that contains a C5 element and a non-C5 element simultaneously and, therefore, the lemma still holds when $i = k + 1$. ■

Lemma 5.2. *Given a capping genome* $\widehat{\Pi}$ *and a signed number* $x \in \widehat{E}$*, if* $\mathsf{char}(x, \widehat{\Pi})$ *is C3 (respectively, T), then* $\mathsf{char}(\widehat{\Pi}\widehat{\Gamma}(x), \widehat{\Pi})$ *is T (respectively, C3) and if* $\mathsf{char}(x, \widehat{\Pi})$ *is O (respectively, N3 and C5), then* $\mathsf{char}(\widehat{\Pi}\widehat{\Gamma}(x), \widehat{\Pi})$ *is O (respectively, N3 and C5).*

Proof. Suppose that $\mathsf{char}(x, \widehat{\Pi}) = \mathrm{C3}$. Then $\mathsf{char}(\widehat{\Gamma}(x), \widehat{\Pi}) = \mathrm{C5}$, since $\widehat{\Gamma}(x)$ is the complement of x in $\widehat{\Pi}$, and clearly $\mathsf{char}(\widehat{\Pi}\widehat{\Gamma}(x), \widehat{\Pi}) = \mathrm{T}$. Conversely, suppose that $\mathsf{char}(x, \widehat{\Pi}) = \mathrm{T}$. Then $\mathsf{char}(\widehat{\Gamma}(x), \widehat{\Pi}) = \mathrm{O}$ and $\mathsf{char}(\widehat{\Pi}\widehat{\Gamma}(x), \widehat{\Pi}) = \mathrm{C3}$. Suppose that $\mathsf{char}(x, \widehat{\Pi}) = \mathrm{O}$. Then either $\mathsf{char}(\widehat{\Gamma}(x), \widehat{\Pi}) = \mathrm{O}$ or $\mathsf{char}(\widehat{\Gamma}(x), \widehat{\Pi}) = \mathrm{T}$. Whatever the case, however, we have $\mathsf{char}(\widehat{\Pi}\widehat{\Gamma}(x), \widehat{\Pi}) = \mathrm{O}$. Suppose that $\mathsf{char}(x, \widehat{\Pi}) = \mathrm{N3}$. Then $\mathsf{char}(\widehat{\Gamma}(x), \widehat{\Pi}) = \mathrm{C5}$ and $\mathsf{char}(\widehat{\Pi}\widehat{\Gamma}(x), \widehat{\Pi}) = \mathrm{N3}$. Suppose that $\mathsf{char}(x, \widehat{\Pi}) = \mathrm{C5}$. Then either $\mathsf{char}(\widehat{\Gamma}(x), \widehat{\Pi}) = \mathrm{C3}$ or $\mathsf{char}(\widehat{\Gamma}(x), \widehat{\Pi}) = \mathrm{N3}$. Whatever the case, we have $\mathsf{char}(\widehat{\Pi}\widehat{\Gamma}(x), \widehat{\Pi}) = \mathrm{C5}$. ■

Lemma 5.3. *For $n_\chi \le i \le n_\chi + n_\tau$, there are no two elements a and b in the same cycle of* $\widehat{\Sigma}\widehat{\Pi}_i^{-1}$ *such that* $(\mathsf{char}(a, \widehat{\Pi}_i), \mathsf{char}(b, \widehat{\Pi}_i)) \in \mathrm{CEpair}$.

Proof. We prove this lemma by induction on i. Suppose that $i = n_\chi$. Then Algorithm 2 (SoRT^2) has just finished its step 4 and hence there are no a and b in a cycle of $\widehat{\Sigma}\widehat{\Pi}_i^{-1}$ such that $(\mathsf{char}(a, \widehat{\Pi}_i), \mathsf{char}(b, \widehat{\Pi}_i)) \in \mathrm{CEpair}$. Thus, the lemma is true when $i = n_\chi$. Suppose that the lemma holds when $i = k$, where $n_\chi \le k < n_\chi + n_\tau$, and that we can find a translocation $\phi_{i+1} = (\mathsf{5cap}(x), \mathsf{5cap}(y))(x, y)(\widehat{\Pi}_i\widehat{\Gamma}(\mathsf{5cap}(y)), \widehat{\Pi}_i\widehat{\Gamma}(\mathsf{5cap}(x)))(\widehat{\Pi}_i\widehat{\Gamma}(y), \widehat{\Pi}_i\widehat{\Gamma}(x))$ by choosing x and y from a non-fixed cycle α of $\widehat{\Sigma}\widehat{\Pi}_i^{-1}$ with satisfying $(\mathsf{char}(x, \widehat{\Pi}_i), \mathsf{char}(y, \widehat{\Pi}_i)) \in \mathrm{TLpair}$. Then α and its mate cycle contain no C5 element according to Lemmas 5.1 and 5.2 and also contain no two elements a and b with $\mathsf{char}(a, \widehat{\Pi}_i) \in \{\mathrm{C3}, \mathrm{N3}\}$ and $\mathsf{char}(b, \widehat{\Pi}_i) \in \{\mathrm{C3}, \mathrm{N3}\}$ by the induction hypothesis. Moreover, $(\mathsf{5cap}(x), \mathsf{5cap}(y))$ is a (C5, C5) cycle and so is $(\widehat{\Pi}_i\widehat{\Gamma}(\mathsf{5cap}(y)), \widehat{\Pi}_i\widehat{\Gamma}(\mathsf{5cap}(x)))$ by Lemma 5.2. Since $\widehat{\Pi}_{i+1} = \phi_{i+1}\widehat{\Pi}_i$, we have $\widehat{\Sigma}\widehat{\Pi}_{i+1}^{-1} = \widehat{\Sigma}\widehat{\Pi}_i^{-1}\phi_{i+1}^{-1}$, where the 2-cycles $(x, y)^{-1}$ and $(\widehat{\Pi}_i\widehat{\Gamma}(y), \widehat{\Pi}_i\widehat{\Gamma}(x))^{-1}$ in ϕ_{i+1}^{-1} act on α and its mate cycle in $\widehat{\Sigma}\widehat{\Pi}_i^{-1}$ respectively by splitting them into four smaller cycles, and the other two 2-cycles in ϕ_{i+1}^{-1} act on some cycles of $\widehat{\Sigma}\widehat{\Pi}_i^{-1}$ that consist only of C5 elements. Note that an element z is a 3′ cap in $\widehat{\Pi}_i$ if and only if z is a 3′ cap in $\widehat{\Pi}_{i+1}$ (since $\widehat{\Pi}_{i+1}$ and $\widehat{\Pi}_i$ are co-tailed). Based on this property, as well as the discussion above, we cannot find a cycle in $\widehat{\Sigma}\widehat{\Pi}_{i+1}$ that contains two elements a and b with $\mathsf{char}(a, \widehat{\Pi}_{i+1}) \in \{\mathrm{C3}, \mathrm{N3}\}$ and $\mathsf{char}(b, \widehat{\Pi}_{i+1}) \in \{\mathrm{C3}, \mathrm{N3}\}$. Therefore, the lemma still holds when $i = k + 1$. ■

Lemma 5.4. *Let $m = n_\chi + n_\tau$ and let a and b be any two non-C5 elements of* $\widehat{\Pi}_m$ *that are both in a cycle of* $\widehat{\Sigma}\widehat{\Pi}_m^{-1}$*. Then a and b are on the same chromosome in* $\widehat{\Pi}_m$.

Proof. By Lemma 5.3, we have $(\mathsf{char}(a, \widehat{\Pi}_m), \mathsf{char}(b, \widehat{\Pi}_m)) \notin \mathrm{CEpair}$ and hence $(\mathsf{char}(a, \widehat{\Pi}_m), \mathsf{char}(b, \widehat{\Pi}_m)) \in \mathrm{TLpair}$. Recall that Algorithm 2 (SoRT^2) produces $\widehat{\Pi}_m$ when it has finished its step 5. It indicates that there are no two non-C5 elements x and y in $\widehat{\Pi}_m$ that are both in a cycle of $\widehat{\Sigma}\widehat{\Pi}_m^{-1}$ such that $(x, y) \not| \widehat{\Pi}_m$,

$(x, \widehat{\Gamma}(y))|\widehat{\Pi}_m$ and $(\mathsf{char}(x, \widehat{\Pi}_m), \mathsf{char}(y, \widehat{\Pi}_m)) \in$ TLpair. This suggests that $(a, b)|\widehat{\Pi}_m$ or $(a, \widehat{\Gamma}(b))|\widehat{\Pi}_m$. The former implies that a and b are on the same chromosome strand in $\widehat{\Pi}_m$, and the latter implies that a and b are on the same chromosome but different strands. Whatever the case, they are on the same chromosome in $\widehat{\Pi}_m$. ■

Lemma 5.5. *Let $m = n_\chi + n_\tau$. For each chromosome $\hat{\pi}^i$ in $\widehat{\Pi}_m$, there is a corresponding chromosome $\hat{\sigma}^j$ in $\widehat{\Sigma}$ such that the gene content of $\hat{\pi}^i$ equals to that of $\hat{\sigma}^j$, where $1 \le i, j \le M$.*

Proof. For convenience, we use $\mathrm{GC}(\hat{\pi}^i)$ and $\mathrm{GC}(\hat{\sigma}^j)$ to denote the gene contents of chromosomes $\hat{\pi}^i$ and $\hat{\sigma}^j$, respectively. First, we claim that if $\mathrm{GC}(\hat{\pi}^i) \cap \mathrm{GC}(\hat{\sigma}^j) \neq \emptyset$, then $\mathrm{GC}(\hat{\sigma}^j) \subseteq \mathrm{GC}(\hat{\pi}^i)$. Suppose that $\mathrm{GC}(\hat{\pi}^i) \cap \mathrm{GC}(\hat{\sigma}^j) \neq \emptyset$ and $\mathrm{GC}(\hat{\sigma}^j) \setminus \mathrm{GC}(\hat{\pi}^i) \neq \emptyset$. Then there are at least two elements a and b in $\mathrm{GC}(\hat{\sigma}^j)$ such that $a \in \mathrm{GC}(\hat{\pi}^i) \cap \mathrm{GC}(\hat{\sigma}^j), b \in \mathrm{GC}(\hat{\sigma}^j) \setminus \mathrm{GC}(\hat{\pi}^i)$ and $\hat{\sigma}^j(a) = b$.

Case 1. Suppose that $\mathsf{char}(b, \widehat{\Pi}_m) \neq$ C5. Then let $\hat{\pi}^i(a) = c$. Note that if $\hat{\pi}^i$ contains a single gene, then $c = a$; otherwise, $c \neq a$. Whatever the case may be, $c \in \mathrm{GC}(\hat{\pi}^i)$. Since $\widehat{\Pi}_m^{-1}(c) = a$ and $\widehat{\Sigma}(a) = b$, we have $\widehat{\Sigma}\widehat{\Pi}_m^{-1}(c) = b$, implying that c and b are adjacent in a cycle of $\widehat{\Sigma}\widehat{\Pi}_m^{-1}$. Since $\mathsf{char}(b, \widehat{\Pi}_m) \neq$ C5, $\mathsf{char}(c, \widehat{\Pi}_m) \neq$ C5 according to Lemma 5.1. Further by Lemma 5.4, b and c are on the same chromosome in $\widehat{\Pi}_m$. That is, $b \in \mathrm{GC}(\hat{\pi}^i)$, a contradiction to the assumption that $b \in \mathrm{GC}(\hat{\sigma}^j) \setminus \mathrm{GC}(\hat{\pi}^i)$.

Case 2. Suppose that $\mathsf{char}(b, \widehat{\Pi}_m) =$ C5. Then $\mathsf{char}(\widehat{\Gamma}(b), \widehat{\Pi}_m) =$ C3. Let $\hat{\pi}^i(a) = c$. As discussed in case 1, c is adjacent to b in a cycle of $\widehat{\Sigma}\widehat{\Pi}_m^{-1}$. Since $\mathsf{char}(b, \widehat{\Pi}_m) =$ C5, $\mathsf{char}(c, \widehat{\Pi}_m) =$ C5 by Lemma 5.1. This implies that $\mathsf{char}(a, \widehat{\Pi}_m) =$ C3 and hence $\mathsf{char}(\widehat{\Gamma}(a), \widehat{\Pi}_m) =$ C5. Since $\widehat{\Sigma}$ and $\widehat{\Pi}_m$ are co-tailed, each of the two strands $\hat{\sigma}_1^j$ and $\hat{\sigma}_2^j$ in $\hat{\sigma}^j$ contains exactly one number that is characterized as a C5 element with respect to $\widehat{\Pi}_m$. Clearly, this C5 element in the strand of $\hat{\sigma}^j$ with containing $\widehat{\Gamma}(b)$, say $\hat{\sigma}_2^j$, is $\widehat{\Gamma}(a)$. Based on the result derived in case 1, all the elements following $\widehat{\Gamma}(a)$ in the $\hat{\sigma}_2^j$ strand in the $5' \to 3'$ direction are all non-C5 elements with respect to $\widehat{\Pi}_m$ and hence they must belong to $\mathrm{GC}(\hat{\pi}^i)$, suggesting that $\widehat{\Gamma}(b)$, as well as b, belongs to $\mathrm{GC}(\hat{\pi}^i)$, a contradiction.

Next, recall that $\widehat{\Pi}_m$ and $\widehat{\Sigma}$ are co-tailed genomes over the same set $\widehat{E}$ of genes. The co-tailed property further implies that $\widehat{\Pi}_m$ and $\widehat{\Sigma}$ have the same number of chromosomes. As was shown in the above claim, for each chromosome $\hat{\sigma}^j$ in $\widehat{\Sigma}$, there is a corresponding chromosome $\hat{\pi}^i$ in $\widehat{\Pi}_m$ such that $\mathrm{GC}(\hat{\sigma}^j) \subseteq \mathrm{GC}(\hat{\pi}^i)$. If $\mathrm{GC}(\hat{\sigma}^j) \subset \mathrm{GC}(\hat{\pi}^i)$, then it is clear that the number of chromosomes in $\widehat{\Pi}_m$ is not equal to that in $\widehat{\Sigma}$, a contradiction. Therefore, $\mathrm{GC}(\hat{\pi}^i) = \mathrm{GC}(\hat{\sigma}^j)$ and the lemma holds. ■

DISCLOSURE STATEMENT

No competing financial interests exist.

REFERENCES

Alekseyev, M.A. 2008. Multi-break rearrangements and breakpoint re-uses: from circular to linear genomes. *J. Comput. Biol.* 15, 1117–1131.

Alekseyev, M.A., and Pevzner, P.A. 2008. Multi-break rearrangements and chromosomal evolution. *Theoret. Comput. Sci.* 395, 193–202.

Bader, D.A., Moret, B.M., and Yan, M. 2001. A linear-time algorithm for computing inversion distance between signed permutations with an experimental study. *J. Comput. Biol.* 8, 483–491.

Bafna, V., and Pevzner, P.A. 1998. Sorting by transpositions. *SIAM J. Discrete Math.* 11, 221–240.

Belda, E., Moya, A., and Silva, F.J. 2005. Genome rearrangement distances and gene order phylogeny in γ-Proteobacteria. *Mol. Biol. Evol.* 22, 1456–1467.

Bergeron, A., Mixtacki, J., and Stoye, J. 2006a. On sorting by translocations. *J. Comput. Biol.* 13, 567–578.

Bergeron, A., Mixtacki, J., and Stoye, J. 2006b. A unifying view of genome rearrangements. *Lect. Notes Comput. Sci.* 4175, 163–173.

Blanchette, M., Kunisawa, T., and Sankoff, D. 1996. Parametric genome rearrangement. *Gene* 172, GC11–GC17.

Christie, D.A. 1996. Sorting permutations by block-interchanges. *Inform. Process. Lett.* 60, 165–169.

Elias, I., and Hartman, T. 2005. A 1.375-approximation algorithm for sorting by transpositions. *Lect. Notes Comput. Sci.* 3692, 204–215.

Eriksen, N. 2002. $(1+\varepsilon)$-Approximation of sorting by reversals and transpositions. *Theoret. Comput. Sci.* 289, 517–529.

Fraleigh, J.B. 2003. *A First Course in Abstract Algebra*, 7th ed. Addison-Wesley, Boston.

Hannenhalli, S. 1996. Polynomial-time algorithm for computing translocation distance between genomes. *Discrete Appl. Math.* 71, 137–151.

Hannenhalli, S., and Pevzner, P.A. 1995. Transforming men into mice (polynomial algorithm for genomic distance problem). *Proc. 36th IEEE Symp. Found. Comput. Sci.* 581–592.

Hannenhalli, S., and Pevzner, P.A. 1999. Transforming cabbage into turnip: polynomial algorithm for sorting signed permutations by reversals. *J. ACM* 46, 1–27.

Hartman, T., and Sharan, R. 2005. A 1.5-approximation algorithm for sorting by transpositions and transreversals. *J. Comput. Syst. Sci.* 70, 300–320.

Kaplan, H., Shamir, R., and Tarjan, R.E. 1999. Faster and simpler algorithm for sorting signed permutations by reversals. *SIAM J. Comput.* 29, 880–892.

Lin, Y.C., Lu, C.L., Chang, H.-Y., et al. 2005. An efficient algorithm for sorting by block-interchanges and its application to the evolution of vibrio species. *J. Comput. Biol.* 12, 102–112.

Lu, C.L., Huang, Y.L., Wang, T.C., et al. 2006. Analysis of circular genome rearrangement by fusions, fissions and block-interchanges. *BMC Bioinform.* 7.

Meidanis, J., and Dias, Z. 2000. An alternative algebraic formalism for genome rearrangements, 213–223. *In* Sankoff, D., and Nadeau, J.H., eds. *Comparative Genomics: Empirical and Analytical Approaches to Gene Order Dynamics, Map Alignment and Evolution of Gene Families*. Kluwer Academic Press, Amsterdam.

Meidanis, J., and Dias, Z. 2001. Genome rearrangements distance by fusion, fission, and transposition is easy. *Proc. 8th Int. Symp. String Process. Inform. Retrieval* 250–253.

Mira, C., and Meidanis, J. 2007. Sorting by block-interchanges and signed reversals. *Proc. Int. Conf. Inform. Technol.* 670–676.

Ozery-Flato, M., and Shamir, R. 2006. An $O(n^{3/2}\sqrt{\log n})$ algorithm for sorting by reciprocal translocations. *Lect. Notes Comput. Sci.* 4009, 258–269.

Pevzner, P.A., and Tesler, G. 2003. Genome rearrangements in mammalian evolution: lessons from human and mouse genomes. *Genome Res.* 13, 37–45.

Sankoff, D., Leduc, G., Antoine, N., et al. 1992. Gene order comparisons for phylogenetic inference: evolution of the mitochondrial genome. *Proc. Natl. Acad. Sci. USA* 89, 6575–6579.

Tannier, E., Bergeron, A., and Sagot, M.-F. 2007. Advances on sorting by reversals. *Discrete Appl. Math.* 155, 881–888.

Yancopoulos, S., Attie, O., and Friedberg, R. 2005. Efficient sorting of genomic permutations by translocation, inversion and block interchange. *Bioinformatics* 21, 3340–3346.

Address correspondence to:
Dr. C.L. Lu
Institute of Bioinformatics and Systems Biology
Department of Biological Science and Technology
National Chiao Tung University
Hsinchu 300, Taiwan

E-mail: cllu@mail.nctu.edu.tw

This article has been cited by:

1. Chi-Long Li, Kun-Tze Chen, Chin Lung Lu. 2013. Assembling contigs in draft genomes using reversals and block-interchanges. *BMC Bioinformatics* **14**:Suppl 5, S9. [CrossRef]
2. Keng-Hsuan Huang, Kun-Tze Chen, Chin Lung Lu. 2011. Sorting permutations by cut-circularize-linearize-and-paste operations. *BMC Genomics* **12**:Suppl 3, S26. [CrossRef]
3. Y.-L. Huang, C.-C. Huang, C. Y. Tang, C. L. Lu. 2010. SoRT2: a tool for sorting genomes and reconstructing phylogenetic trees by reversals, generalized transpositions and translocations. *Nucleic Acids Research* **38**:Web Server, W221-W227. [CrossRef]