Sorting Signed Circular Permutations by Super Short Reversals

Gustavo Galvão, Christian Baudet, Zanoni Dias

To cite this version:

Gustavo Galvão, Christian Baudet, Zanoni Dias. Sorting Signed Circular Permutations by Super Short Reversals. 11th International Symposium on Bioinformatics Research and Applications, Jun 2015, Norfolk, Virginia, United States. pp.272-283, 10.1007/978-3-319-19048-8_23 . hal-01316993

HAL Id: hal-01316993
https://hal.inria.fr/hal-01316993

Submitted on 12 Jun 2017

Sorting signed circular permutations by super short reversals

Gustavo Rodrigues Galvão[1], Christian Baudet[2,3], and Zanoni Dias[1]

[1] Institute of Computing, University of Campinas, Campinas, Brazil
{ggalvao,zanoni}@ic.unicamp.br
[2] Laboratoire Biométrie et Biologie Evolutive, Université de Lyon, Université Lyon 1, CNRS, Villeurbanne, UMR5558, France
[3] Inria Grenoble - Rhône-Alpes, Erable Team, France
christian.baudet@inria.fr

Abstract. We consider the problem of sorting a circular permutation by reversals of length at most 2, a problem that finds application in comparative genomics. Polynomial-time solutions for the unsigned version of this problem are known, but the signed version remained open. In this paper, we present the first polynomial-time solution for the signed version of this problem. Moreover, we perform an experiment for inferring distances and phylogenies for published *Yersinia* genomes and compare the results with the phylogenies presented in previous works.

1 Introduction

Distance-based methods form one of the three large groups of methods to infer phylogenetic trees from sequence data [8, Chapter 5]. Such methods proceed in two steps. First, the evolutionary distance is computed for every sequence pair and this information is stored in a matrix of pairwise distances. Then, a phylogenetic tree is constructed from this matrix using a specific algorithm, such as *Neighbor-Joining* [9]. Note that, in order to complete the first step, we need some method to estimate the evolutionary distance between a sequence pair. Assuming the sequence data correspond to complete genomes, we can resort to the genome rearrangement approach [4] in order to estimate the evolutionary distance.

In genome rearrangements, one estimates the evolutionary distance between two genomes by finding the rearrangement distance between them, which is the length of the shortest sequence of rearrangement events that transforms one genome into the other. Assuming genomes consist of a single chromosome, share the same set of genes, and contain no duplicated genes, we can represent them as permutations of integers, where each integer corresponds to a gene. If, besides the order, the orientation of the genes is also regarded, then each integer has a sign, + or −, and the permutation is called a signed permutation (similarly, we also refer to a permutation as an unsigned permutation when its elements do not have signs). Moreover, if the genomes are circular, then the permutations are also circular; otherwise, they are linear.

A number of publications address the problem of finding the rearrangement distance between two permutations, which can be equivalently stated as a problem of sorting a permutation into the identity permutation (for a detailed survey, the reader is referred to the book of Fertin *et al.* [4]). This problem varies according to the rearrangement events allowed to sort a permutation. Reversals are the most common rearrangement event observed in genomes. They are responsible for reversing the order and orientation of a sequence of genes within a genome. Although the problem of sorting a permutation by reversals is a well-studied problem, most of the works concerning it do not take into account the length of the reversals (*i.e.* the number of genes affected by it). Since it has been observed that short reversals are prevalent in the evolution of some species [1, 2, 7, 10], recent efforts have been made to address this issue [3, 5].

In this paper, we add to those efforts and present a polynomial-time solution for the problem of sorting a signed circular permutation by super short reversals, that is, reversals which affect at most 2 elements (genes) of a permutation (genome). This solution closes a gap in the literature since polynomial-time solutions are known for the problem of sorting an unsigned circular permutation [3, 6], for the problem of sorting an unsigned linear permutation [6], and for the problem of sorting a signed linear permutation [5]. Moreover, we reproduce the experiment performed by Egri-Nagy *et. al.* [3] to infer distances and phylogenies for published *Yersinia* genomes, but this time we consider the orientation of the genes (they have ignored it in order to treat the permutations as unsigned).

The rest of this paper is organized as follows. Section 2 succinctly presents the solution developed by Jerrum [6] for the problem of sorting by cyclic super short reversals. Section 3 builds upon the previous section and presents the solution for the problem of sorting by signed cyclic super short reversals. Section 4 briefly explains how we can use the solutions described in Sect(s). 2 and 3 to solve the problem of sorting a (signed) circular permutation by super short reversals. Section 5 presents experimental results performed on *Yersinia pestis* data. Finally, Sect. 6 concludes the paper.

2 Sorting by Cyclic Super Short Reversals

A *permutation* π is a bijection of $\{1, 2, \ldots, n\}$ onto itself. A classical notation used in combinatorics for denoting a permutation π is the two-row notation

$$\pi = \begin{pmatrix} 1 & 2 & \ldots & n \\ \pi_1 & \pi_2 & \ldots & \pi_n \end{pmatrix},$$

$\pi_i \in \{1, 2, \ldots, n\}$ for $1 \leq i \leq n$. This notation indicates that $\pi(1) = \pi_1$, $\pi(2) = \pi_2$, $\ldots$, $\pi(n) = \pi_n$. The notation used in genome rearrangement literature, which is the one we will adopt, is the one-row notation $\pi = (\pi_1\ \pi_2\ \ldots\ \pi_n)$. We say that π has size n. The set of all permutations of size n is S_n.

A *cyclic reversal* $\rho(i, j)$ is an operation that transforms a permutation $\pi = (\pi_1\ \pi_2\ \ldots\ \pi_{i-1}\ \underline{\pi_i\ \pi_{i+1}\ \ldots\ \pi_{j-1}\ \pi_j}\ \pi_{j+1}\ \ldots\ \pi_n)$ into the permutation $\pi \cdot \rho(i, j) = (\pi_1\ \pi_2\ \ldots\ \pi_{i-1}\ \underline{\pi_j\ \pi_{j-1}\ \ldots\ \pi_{i+1}\ \pi_i}\ \pi_{j+1}\ \ldots\ \pi_n)$ if $1 \leq i < j \leq n$ and transforms a permutation $\pi = (\underline{\pi_1\ \pi_2\ \ldots\ \pi_i}\ \pi_{i+1}\ \ldots\ \pi_{j-1}\ \underline{\pi_j\ \pi_{j+1}\ \ldots\ \pi_n})$ into the permutation $\pi \cdot \rho(i,\, j) = (\underline{\pi_j\ \pi_{j+1}\ \ldots\ \pi_n}\ \pi_{i+1}\ \ldots\ \pi_{j-1}\ \underline{\pi_i\ \pi_{i-1}\ \ldots\ \pi_1})$ if $1 \leq j < i \leq n$. The cyclic reversal $\rho(i,\, j)$ is called a *cyclic k-reversal* if $k \equiv j - i + 1 \pmod n$. It is called *super short* if $k = 2$.

The problem of sorting by cyclic super short reversals consists in finding the minimum number of cyclic super short reversals that transform a permutation $\pi \in S_n$ into $\iota_n = (1\ 2\ \ldots\ n)$. This number is referred to as the *cyclic super short reversal distance* of permutation π and it is denoted by $d(\pi)$.

Let $S(\pi_i,\, \pi_j)$ denote the act of switching the positions of the elements π_i and π_j in a permutation π. Note that the cyclic 2-reversal $\rho(i,\, j)$ can be alternatively denoted by $S(\pi_i,\, \pi_j)$. Given a sequence S of cyclic super short reversals and a permutation $\pi \in S_n$, let $R_S(\pi_i)$ be the number of cyclic 2-reversals of the type $S(\pi_i,\, \pi_j)$ and let $L_S(\pi_i)$ be the number of cyclic 2-reversals of the type $S(\pi_k,\, \pi_i)$. In other words, $R_S(\pi_i)$ denotes the number of times a cyclic 2-reversal moves the element π_i to the right and $L_S(\pi_i)$ denotes the number of times a cyclic 2-reversal moves the element π_i to the left. We define the *net displacement* of an element π_i with respect to S as $d_S(\pi_i) = R_S(\pi_i) - L_S(\pi_i)$. The *displacement vector* of π with respect to S is defined as $d_S(\pi) = (d_S(\pi_1),\, d_S(\pi_2),\, \ldots,\, d_S(\pi_n))$.

Lemma 1. *Let $S = \rho_1,\, \rho_2,\, \ldots,\, \rho_t$ be a sequence of cyclic super short reversals that sorts a permutation $\pi \in S_n$. Then, we have that*

$$\sum_{i=1}^{n} d_S(\pi_i) = 0, \tag{1}$$

$$\pi_i - d_S(\pi_i) \equiv i \pmod n. \tag{2}$$

Proof. Let L_S be the number of times a cyclic super short reversal of S moves an element to the left and let R_S be the number of times a cyclic super short reversal of S moves an element to the right. Then, $L_S = R_S$ because a cyclic super reversal always moves two elements, one for each direction. Therefore, we have that $\sum_{i=1}^{n} d_S(\pi_i) = \sum_{i=1}^{n} (R_S(\pi_i) - L_S(\pi_i)) = R_S - L_S = 0$ and equation 1 follows. The equation 2 follows from the fact that, once the permutation is sorted, all of its elements must be in the correct position. □

Note that, in one hand, we can think of a sequence of cyclic super short reversals as specifying a displacement vector. On the other hand, we can also think of a displacement vector as specifying a sequence of cyclic super short reversals. Let $x = (x_1, x_2, \ldots, x_n) \in Z^n$ be a vector and $\pi \in S_n$ be a permutation. We say that x is a *valid vector* for π if $\sum_i x_i = 0$ and $\pi_i - x_i \equiv i \pmod n$. Given a vector $x = (x_1,\, x_2,\, \ldots,\, x_n) \in Z^n$ and two distinct integers $i,\, j \in \{1,\, 2,\, \ldots,\, n\}$, let $r = i - j$ and $s = (i + x_i) - (j + x_j)$. The *crossing number* of i and j with respect to x is defined by

$$c_{ij}(x) = \begin{cases} |\{k \in [r, s] : k \equiv 0 \pmod n\}| & \text{if } r \leq s, \\ -|\{k \in [s, r] : k \equiv 0 \pmod n\}| & \text{if } r > s. \end{cases}$$

The crossing number of x is defined by $C(x) = \frac{1}{2}\sum_{i,j} |c_{ij}(x)|$. Intuitively, if S is a sequence of cyclic super short reversals that sorts a permutation π and $d_S(\pi) = x$, then $c_{ij}(x)$ measures the number of times the elements π_i and π_j must "cross", that is, the number of cyclic 2-reversals of type $S(\pi_i, \pi_j)$ minus the number of cyclic 2-reversals of type $S(\pi_j, \pi_i)$. Using the notion of crossing number, Jerrum [6] was able to prove the following fundamental lemma.

Lemma 2 (Jerrum [6]). *Let S be a minimum-length sequence of cyclic super short reversals that sorts a permutation $\pi \in S_n$ and let $x \in Z^n$ be a valid vector for π. If $d_S(\pi) = x$, then $d(\pi) = C(x)$.*

The Lemma 2 allows the problem of sorting a permutation π by cyclic super short reversals to be recast as the optimisation problem of finding a valid vector $x \in Z^n$ for π with minimum crossing number. More specifically, as Jerrum [6] pointed out, this problem can formulated as the integer program:

$$\begin{gathered}\text{Minimize } C(x) \text{ over } Z^n \\ \text{subject to } \textstyle\sum_i x_i = 0,\ \pi_i - x_i \equiv i \pmod{n}.\end{gathered}$$

Although solving an integer program is NP-hard in the general case, Jerrum [6] presented a polynomial-time algorithm for solving this one.

Firstly, Jerrum [6] introduced a transformation $T_{ij} : Z^n \to Z^n$ defined as follows. For any vector $x \in Z^n$, the result, $x' = T_{ij}(x)$, of applying T_{ij} to x is given by $x'_k = x_k$ for $k \notin \{i,j\}$, $x'_i = x_i - n$, and $x'_j = x_j + n$. Lemma 3 shows what is the effect of this transformation on the crossing number of a vector.

Lemma 3. *Let x and x' be two vectors over Z^n such that $x' = T_{ij}(x)$. Then, $C(x') - C(x) = 2(n + x_j - x_i)$.*

Proof. The proof of this lemma is given by Jerrum [6, Theorem 3.9]. We note, however, that he mistakenly wrote that $C(x') - C(x) = 4(n + x_j - x_i)$. In other words, he forgot to divide the result by 2. This division is necessary because the crossing number of a vector is the half of the sum of the crossing numbers of its indices. □

Let $\max(x)$ and $\min(x)$ respectively denote the maximum and minimum component values of a vector $x \in Z^n$. The transformation T_{ij} is said to *contract* x iff $x_i = \max(x)$, $x_j = \min(x)$ and $x_i - x_j \geq n$. Moreover, T_{ij} is said to *strictly contract* x iff, in addition, the final inequality is strict. The algorithm proposed by Jerrum [6] starts with a feasible solution to the integer program and performs a sequence of strictly contracting transformations which decrease the value of the crossing number. When no further strictly contracting transformation can be performed, the solution is guaranteed to be optimal. This is because, as showed by Jerrum [6], any two local optimum solutions (*i.e* solutions which admit no strictly contracting transformation) can be brought into agreement with each other via a sequence of contracting transformations. The detailed algorithm is given below (Algorithm 1).

```
   Data: A permutation $\pi \in S_n$.
   Result: Number of cyclic super short reversals applied for sorting $\pi$.
 1 Let $x$ be a $n$ dimension vector
 2 for $k = 1$ to $n$ do
 3     $x_k \leftarrow \pi_k - k$
 4 end
 5 while $\max(x) - \min(x) > n$ do
 6     Let $i$,$j$ be two integers such that $x_i = \max(x)$ and $x_j = \min(x)$
 7     $x_i \leftarrow x_i - n$
 8     $x_j \leftarrow x_j + n$
 9 end
10 return $C(x)$
```

Algorithm 1: Algorithm for sorting by cyclic super short reversals.

Regarding the time complexity of Algorithm 1, we have that line 1 and the for loop of lines 2-4 take $O(n)$ time. Jerrum [6] observed that none of the variables x_i changes value more than once, therefore the while loop iterates only $O(n)$ times. As the lines 6-8 take $O(n)$ time, the while loop takes $O(n^2)$ time to execute. Since we can compute the value of $C(x)$ in $O(n^2)$ time, the overall complexity of the algorithm is $O(n^2)$.

Note that, in this section, we have focused on the problem of computing the cyclic super short reversal distance of a permutation rather than finding the minimum number of cyclic super short reversals that sorts it. As Jerrum [6] remarked, his proofs are constructive and directly imply algorithms for finding the sequence of cyclic super short reversals.

3 Sorting by Signed Cyclic Super Short Reversals

A *signed permutation* π is a bijection of $\{-n, \ldots, -2, -1, 1, 2, \ldots, n\}$ onto itself that satisfies $\pi(-i) = -\pi(i)$ for all $i \in \{1, 2, \ldots, n\}$. The two-row notation for a signed permutation is

$$\pi = \begin{pmatrix} -n & \ldots & -2 & -1 & 1 & 2 & \ldots & n \\ -\pi_n & \ldots & -\pi_2 & -\pi_1 & \pi_1 & \pi_2 & \ldots & \pi_n \end{pmatrix},$$

$\pi_i \in \{1, 2, \ldots, n\}$ for $1 \leq i \leq n$. The notation used in genome rearrangement literature, which is the one we will adopt, is the one-row notation $\pi = (\pi_1\ \pi_2\ \ldots\ \pi_n)$. Note that we drop the mapping of the negative elements since $\pi(-i) = -\pi(i)$ for all $i \in \{1, 2, \ldots, n\}$. By abuse of notation, we say that π has size n. The set of all signed permutations of size n is $S_n^{\pm}$.

A *signed cyclic reversal* $\rho(i, j)$ is an operation that transforms a signed permutation $\pi = (\pi_1\ \pi_2 \ldots \pi_{i-1}\ \underline{\pi_i\ \pi_{i+1}\ \ldots\ \pi_{j-1}\ \pi_j}\ \pi_{j+1} \ldots \pi_n)$ into the signed permutation $\pi \cdot \rho(i, j) = (\pi_1\ \pi_2\ \ldots\ \pi_{i-1}\ \underline{-\pi_j\ -\pi_{j-1}\ \ldots\ -\pi_{i+1}\ -\pi_i}\ \pi_{j+1} \ldots \pi_n)$ if $1 \leq i \leq j \leq n$ and transforms a signed permutation $\pi = (\underline{\pi_1\ \pi_2\ \ldots\ \pi_i}\ \pi_{i+1} \ldots \pi_{j-1}$ $\underline{\pi_j\ \pi_{j+1}\ \ldots\ \pi_n})$ into the signed permutation $\pi \cdot \rho(i, j) = (\underline{-\pi_j\ -\pi_{j+1}\ \ldots\ -\pi_n}$

$\pi_{i+1}\ \ldots\ \pi_{j-1}\ \underline{\text{-}\pi_i\ -\pi_{i-1}\ \ldots\ -\pi_1}$) if $1 \leq j < i \leq n$. The signed cyclic reversal $\rho(i,\, j)$ is called a *signed cyclic k-reversal* if $k \equiv j - i + 1 \pmod{n}$. It is called *super short* if $k \leq 2$.

The problem of sorting by signed cyclic super short reversals consists in finding the minimum number of signed cyclic super short reversals that transform a permutation $\pi \in S_n^{\pm}$ into ι_n. This number is referred to as the *signed cyclic super short reversal distance* of permutation π and it is denoted by $d^{\pm}(\pi)$.

Let $S(|\pi_i|,\, |\pi_j|)$ denote the act of switching the positions and flipping the signs of the elements π_i and π_j in a signed permutation π. Note that the signed cyclic 2-reversal $\rho(i,\, j)$ can be alternatively denoted by $S(|\pi_i|,\, |\pi_j|)$. Given a sequence S of cyclic signed super short reversals and a signed permutation $\pi \in S_n^{\pm}$, let $R_S(\pi_i)$ be the number of signed cyclic 2-reversals of the type $S(|\pi_i|,\, |\pi_j|)$ and let $L_S(\pi_i)$ be the number of signed cyclic 2-reversals of the type $S(|\pi_k|,\, |\pi_i|)$. We define the *net displacement* of an element π_i with respect to S as $d_S(\pi_i) = R_S(\pi_i) - L_S(\pi_i)$. The *displacement vector* of π with respect to S is defined as $d_S(\pi) = (d_S(\pi_1),\, d_S(\pi_2),\, \ldots,\, d_S(\pi_n))$. The following lemma is the signed analog of Lemma 1. We omit the proof because it is the same as of the proof of Lemma 1.

Lemma 4. *Let $S = \rho_1,\, \rho_2,\, \ldots,\, \rho_t$ be a sequence of signed cyclic super short reversals that sorts a signed permutation $\pi \in S_n^{\pm}$. Then, we have that*

$$\sum_{i=1}^{n} d_S(\pi_i) = 0, \tag{3}$$

$$|\pi_i| - d_S(\pi_i) \equiv i \pmod{n}. \tag{4}$$

Let $x \in Z^n$ be a vector and $\pi \in S_n^{\pm}$ be a signed permutation. We say that x is a *valid vector* for π if $\sum_i x_i = 0$ and $|\pi_i| - x_i \equiv i \pmod{n}$. Given a valid vector x for the signed permutation π, we define the set *podd*$(\pi,\, x)$ as *podd*$(\pi,\, x) = \{i : \pi_i > 0 \text{ and } |x_i| \text{ is odd}\}$ and we define the set *neven*$(\pi,\, x)$ as *neven*$(\pi,\, x) = \{i : \pi_i < 0 \text{ and } |x_i| \text{ is even}\}$. Moreover, let $U(\pi,\, x)$ denote the union of these sets, that is, $U(\pi,\, x) = \mathit{podd}(\pi, x) \cup \mathit{neven}(\pi, x)$. The following lemma is the signed analog of Lemma 2.

Lemma 5. *Let S be a minimum-length sequence of signed cyclic super short reversals that sorts a signed permutation $\pi \in S_n^{\pm}$ and let $x \in Z^n$ be a valid vector for π. If $d_S(\pi) = x$, then $d^{\pm}(\pi) = C(x) + |U(\pi, x)|$.*

Proof. Note that the sequence S can be decomposed into two distinct subsequences S_1 and S_2 such that S_1 is formed by the signed cyclic 1-reversals of S and S_2 is formed by the signed cyclic 2-reversals of S. Moreover, we can assume without loss of generality that the signed cyclic reversals of subsequence S_2 are applied first. We argue that $|S_1| = |U(\pi, x)|$ regardless the size of S_2. To see this, suppose that we apply a signed cyclic 2-reversal $\rho(i,\, j)$ of S_2 in π, obtaining a signed permutation π'. Moreover, let S' be the resulting sequence after we remove $\rho(i,\, j)$ from S. We have that $d_{S'}(\pi'_k) = d_S(\pi_k)$ for $k \notin \{i{,}j\}$, $d_{S'}(\pi'_i) = d_S(\pi_i) - 1$, and $d_{S'}(\pi'_j) = d_S(\pi_j) + 1$. Then, assuming the vector $x' \in Z^n$ is equal to $d_{S'}(\pi')$, we can conclude that $U(\pi', x') = U(\pi, x)$ because $\rho(i,\, j)$ has changed both the parities of $|x_i|$ and $|x_j|$ and the signs of π_i and π_j. Since $|S_1|$ $= |U(\pi, x)|$ regardless the size of S_2 and we know from Lemma 2 that $|S_2| \geq C(x)$, we can conclude that $|S_2| = C(x)$, therefore the lemma follows. □

The Lemma 5 allows the problem of sorting a signed permutation π by signed cyclic super short reversals to be recast as the optimisation problem of finding a valid vector $x \in Z^n$ for π which minimizes the sum $C(x) + |U(\pi, x)|$. The next theorem shows how to solve this problem in polynomial time.

Theorem 1. *Let $\pi \in S_n^{\pm}$ be a signed permutation. Then, we can find a valid vector $x \in Z^n$ which minimizes the sum $C(x) + |U(\pi, x)|$ in polynomial time.*

Proof. We divide our analysis into two cases:

i) n is even. In this case, we have that the value of $|U(\pi, x)|$ is the same for any feasible solution x. This is because, in order to be a feasible solution, a vector x has to satisfy the restriction $|\pi_i| - x_i \equiv i \pmod{n}$. This means that x_i is congruent modulo n with $a = |\pi_i| - i$ and belongs to the equivalent class $\{\ldots,\, a - 2n,\, a - n,\, a,\, a + n,\, a + 2n,\, \ldots\}$. Since n is even, the parities of the absolute values of the elements in this equivalence class are the same, therefore the value of $|U(\pi, x)|$ is the same for any feasible solution x. It follows that we can only minimize the value of $C(x)$ and this can be done by performing successive strictly contracting transformations.

ii) n is odd. In this case, it is possible to minimize the values of $|U(\pi, x)|$ and $C(x)$. Firstly, we argue that minimizing $C(x)$ leads to a feasible solution x'' such that $C(x'') + |U(\pi, x'')|$ is at least as low as $C(x') + |U(\pi, x')|$, where x' can be any feasible solution such that $C(x')$ is not minimum. To see this, let x' be a feasible solution such that $C(x')$ is not minimum. Then, we can perform a sequence of strictly contracting transformations which decrease the value of $C(x)$. When no further strictly contracting transformation can be performed, we obtain a solution x'' such that $C(x'')$ is minimum. On one hand, we know from Lemma 3 that each strictly contracting transformation T_{ij} decreases $C(x)$ by at least 2 units. On the other hand, since n is odd, its possible that the parities of $|x_i|$ and $|x_j|$ have been changed in such a way that the value of $|U(\pi, x)|$ increases by 2 units. Therefore, in the worst case, each strictly contracting transformation does not change the value of $C(x)$ $+ |U(\pi, x)|$, so $C(x') + |U(\pi, x')| \geq C(x'') + |U(\pi, x'')|$. Now, we argue that, if there exists more than one feasible solution x such that $C(x)$ is minimum, then it is still may be possible to minimize the value of $|U(\pi, x)|$.
Jerrum [6, Theorem 3.9] proved that if there is more than one feasible solution such that $C(x)$ is minimum, then each of these solutions can be brought into agreement with each other via a sequence of contracting transformations. Note that a contracting transformation T_{ij} does not change the value of $C(x)$, but it can change the value of $|U(\pi, x)|$ because n is odd and the

parities of $|x_i|$ and $|x_j|$ change when T_{ij} is performed. This means that, among all feasible solutions such that $C(x)$ is minimum, some of them have minimum $|U(\pi, x)|$ and these solutions are optimal. Therefore, we can obtain an optimal solution by first obtaining a feasible solution with minimum $C(x)$ (this can be done by performing successive strictly contracting transformations) and then we can apply on it every possible contracting transformation T_{ij} which decreases the value of $|U(\pi, x)|$. □

The proof of Theorem 1 directly implies an exact algorithm for sorting by signed cyclic super short reversals. Such an algorithm is described below (Algorithm 2). Regarding its time complexity, we know from previous section that lines 1-9 take $O(n^2)$ time. Since lines 13-23 take $O(1)$ time, we can conclude that the nested for loops take $O(n^2)$ times to execute. Finally, we can compute $C(x)$ + $|U(\pi, x)|$ in $O(n^2)$, therefore the overall complexity of Algorithm 2 is $O(n^2)$.

```
   Data: A permutation $\pi \in S_n^{\pm}$.
   Result: Number of signed cyclic super short reversals applied for sorting $\pi$.
1  Let $x$ be a $n$ dimension vector
2  for $k = 1$ to $n$ do
3      $x_k \leftarrow |\pi_k| - k$
4  end
5  while $\max(x) - \min(x) > n$ do
6      Let $i$,$j$ be two integers such that $x_i = \max(x)$ and $x_j = \min(x)$
7      $x_i \leftarrow x_i - n$
8      $x_j \leftarrow x_j + n$
9  end
10 if $n$ is odd then
11     for $i = 1$ to $n - 1$ do
12         for $j = i + 1$ to $n$ do
13             if $x_i > x_j$ then
14                 $min \leftarrow j$
15                 $max \leftarrow i$
16             else
17                 $min \leftarrow i$
18                 $max \leftarrow j$
19             end
20             if $x_{max} - x_{min} = n$ and $min \in U(\pi, x)$ and $max \in U(\pi, x)$ then
21                 $x_{max} \leftarrow x_i - n$
22                 $x_{min} \leftarrow x_j + n$
23             end
24         end
25     end
26 end
27 return $C(x)$ + $|U(\pi, x)|$
```

Algorithm 2: Algorithm for sorting by signed cyclic super short reversals.

Note that, in this section, we have focused on the problem of computing the signed cyclic super short reversal distance of a signed permutation rather than finding the minimum number of signed cyclic super short reversals that sorts it. We remark that the proofs are constructive and directly imply algorithms for finding the sequence of signed cyclic super short reversals.

4 Sorting Circular Permutations

In this section, we briefly explain how we can use the solution for the problem of sorting by (signed) cyclic super short reversals to solve the problem of sorting a (signed) circular permutation by super short reversals. This explanation is based on Sect. 2.3 of the work of Egri-Nagy *et al.* [3] and on Sect. 2.5 of the book of Fertin *et al.* [4], where one can find more details.

Note that a circular permutation can be "unrolled" to produce a linear permutation, such as defined in the two previous sections. This process can produce n different linear permutations, one for each possible rotation of the circular permutation. Moreover, since a circular permutation represents a circular chromosome, which lives in three dimension, it can also be "turned over" before being unrolled. This means that, for each possible rotation of the circular permutation, we can first turn it over and then unroll it, producing a linear permutation. Again, this process can produce n different linear permutations. The n linear permutations produced in the first process are different from the n linear permutations produced in the second process, thus both processes can produce a total of $2n$ different linear permutations. Each of these $2n$ linear permutations represents a different viewpoint from which to observe the circular permutation, therefore they are all equivalent.

The discussion of the previous paragraph leads us to conclude that, in order to sort a (signed) circular permutation by super short reversals, we can sort each of the $2n$ equivalent (signed) linear permutations by (signed) cyclic super short reversals, generating $2n$ different sorting sequences. Then, we can take the sequence of minimum length as the sorting sequence for the (signed) circular permutation and the *super short reversal distance* of the (signed) circular permutation is the length of this sequence. Note that this procedure takes $O(n^3)$ time because we have to execute Algorithm 1 or Algorithm 2 $O(n)$ times.

5 Experimental Results and Discussion

We implemented the procedure described in the previous section for computing the super short reversal distance of a signed circular permutation and we reproduced the experiment performed by Egri-Nagy *et. al.* [3] for inferring distances and phylogenies for published *Yersinia* genomes. In fact, we performed the same experiment, except that we considered the orientation of the genes rather than ignoring it and we considered that each permutation has 78 elements rather than 79[4]. More specifically, we obtained from Darling *et al.* [2] the signed circular permutations which represent eight *Yersinia* genomes. Then, we computed the super short reversal distance between every pair of signed circular permutation and this information was stored in a matrix of pairwise distances (Table 1). Finally, a phylogenetic tree was constructed from this matrix using *Neighbor-Joining* [9] method. The resulting phylogeny is shown in Fig. 1.

Table 1. Matrix of the super short reversal distances among the signed circular permutations which represent the *Yersinia* genomes. The names of the species were abbreviated so that YPK refers to *Y. pestis Kim*, YPA to *Y. pestis Antiqua*, YPM to *Y. pestis Microtus 91001*, YPC to *Y. pestis CO92*, YPN to *Y. pestis Nepal516*, YPP to *Y. pestis Pestoides F 15-70*, YT1 to *Y. pseudotuberculosis IP31758*, and YT2 to *Y. pseudotuberculosis IP32953*.

	YPK	**YPA**	**YPM**	**YPC**	**YPN**	**YPP**	**YT1**	**YT2**
YPK	0	243	752	205	338	533	764	760
YPA	243	0	772	352	279	510	724	773
YPM	752	772	0	728	747	643	361	385
YPC	205	352	728	0	381	656	776	760
YPN	338	279	747	381	0	547	617	624
YPP	533	510	643	656	547	0	434	457
YT1	764	724	361	776	617	434	0	189
YT2	760	773	385	760	624	457	189	0

Considering the pair of *Y. pseudotuberculosis* as outgroup, the obtained phylogeny shows that *Y. pestis Microtus 91001* was the first to diverge. It was followed then by the divergences of *Y. pestis Pestoides F 15-70*, *Y. pestis Nepal516*, *Y. pestis Antiqua* and the final divergence of *Y. pestis Kim* and *Y. pestis CO92*. This result is different of the one obtained by Egri-Nagy *et. al.* [3] which used super short reversal distance between unsigned permutations. On their results, the divergence of *Y. pestis Nepal516* happened before the divergence of *Y. pestis CO92* which occurred previous to the divergence of *Y. pestis Kim* and *Y. pestis Antiqua*.

In our work and in the work of Egri-Nagy *et. al.* [3], the use of super short reversals resulted on topologies which are different from the one of Darling *et al.* [2], which considered inversions of any size. The first difference observed on the result of Darling *et al.* [2] is that *Y. pestis Pestoides F 15-70* diverged before *Y. pestis Microtus 91001*. The second difference shows that *Y. pestis Nepal516*

[4] In their article, Darling *et al.* [2] state that they could identify 78 conserved segments (or blocks) using Mauve, but they provided permutations with elements ranging from 0 to 78. In a personal communication, Darling confirmed that there are actually 78 blocks, with 0 and 78 being part of the same block. Nevertheless, we performed another experiment, this time considering the permutations have 79 elements. Although the distances were greater, the topology of the tree was the same.

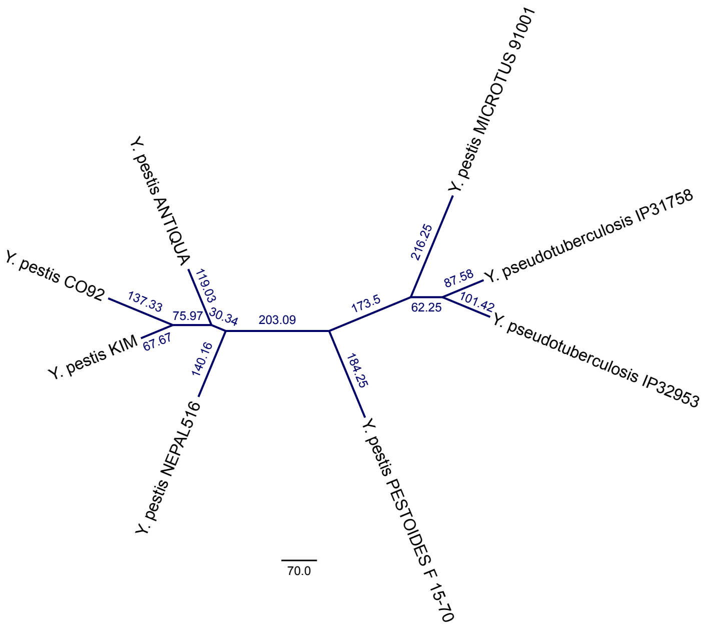

Fig. 1. Phylogeny of the *Yersinia* genomes based on the super short reversal distance of the signed circular permutations.

is sibling of *Y. pestis Kim*, that *Y. pestis CO92* is sibling of *Y. pestis Antiqua* and that these four bacteria have a common ancestor that is descendant of *Y. pestis Microtus 91001*.

If we look to the branch lengths of the two trees obtained with super short reversal distances and we compare with the branch lengths of the topology obtained by Darling *et al.* [2], we can see that our results are more consistent than the one obtained by Egri-Nagy *et al.* [3]. For instance, on our results the distance between the two *Y. pseudotuberculosis* is smaller than the one observed between the pair *Y. pestis Kim* and *Y. pestis Antiqua*, what agrees with the configuration obtained by Darling *et al.* [2].

6 Conclusions

In this paper, we presented a polynomial-time solution for the problem of sorting a signed circular permutation by super short reversals. From a theoretical perspective, this solution is important because it closes a gap in the literature. From a biological perspective, it is important because signed permutations constitute a more adequate model for genomes. Moreover, we performed an experiment to infer distances and phylogenies for published *Yersinia* genomes and compared the results with the phylogenies presented in previous works [2, 3]. Our obtained topology is similar to the one obtained by Egri-Nagy *et. al.* [3]. However, the distances calculated with our algorithm are more consistent with the topology obtained by Darling *et al.* [2]. Some theoretical questions remain open (for instance, the diameter of the super short reversal distance for signed permutations), and we intend to address them in our future research.

Acknowledgments. GRG acknowledges the support by the FAPESP and CAPES under grant #2014/04718-6, São Paulo Research Foundation (FAPESP). CB acknowledges the support by the French Project ANR MIRI BLAN08-1335497 and by the ERC Advanced Grant SISYPHE (FP7/2007-2013)/ERC grant agreement no. [247073]10. ZD acknowledges the support by the CNPq under grants 306730/2012-0, 477692/2012-5, and 483370/2013-4. Finally, the authors thank the Center for Computational Engineering and Sciences at Unicamp for financial support through the FAPESP/CEPID Grant 2013/08293-7.

References

1. D. A. Dalevi, N. Eriksen, K. Eriksson, and S. G. E. Andersson. Measuring genome divergence in bacteria: A case study using chlamydian data. *Journal of Molecular Evolution*, 55(1):24–36, 2002.
2. A. E. Darling, I. Miklós, and M. A. Ragan. Dynamics of genome rearrangement in bacterial populations. *PLoS Genetics*, 4(7):e1000128, 2008.
3. A. Egri-Nagy, V. Gebhardt, M. M. Tanaka, and A. R. Francis. Group-theoretic models of the inversion process in bacterial genomes. *Journal of Mathematical Biology*, 69(1):243–265, 2014.
4. G. Fertin, A. Labarre, I. Rusu, E. Tannier, and S. Vialette. *Combinatorics of Genome Rearrangements*. The MIT Press, Cambridge, 2009.
5. G. R. Galvão, O. Lee, and Z. Dias. Sorting signed permutations by short operations. *Algorithms for Molecular Biology*, 10(12), 2015.
6. M. R. Jerrum. The complexity of finding minimum-length generator sequences. *Theoretical Computer Science*, 36:265–289, 1985.
7. J. F. Lefebvre, N. El-Mabrouk, E. Tillier, and D. Sankoff. Detection and validation of single gene inversions. *Bioinformatics*, 19(suppl 1):i190–i196, 2003.
8. P. Lemey, M. Salemi, and A. Vandamme. *The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing*. Cambridge University Press, New York, 2009.
9. N. Saitou and M. Nei. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4(1):406–425, 1987.
10. C. Seoighe, N. Federspiel, T. Jones, N. Hansen, V. Bivolarovic, R. Surzycki, R. Tamse, C. Komp, L. Huizar, R. W. Davis, S. Scherer, E. Tait, D. J. Shaw, D. Harris, L. Murphy, K. Oliver, K. Taylor, M. A. Rajandream, B. G. Barrell, and K. H. Wolfe. Prevalence of small inversions in yeast gene order evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 97(26):14433–14437, 2000.