A LOWER BOUND ON THE TRANSPOSITION DIAMETER[*]

LINYUAN LU[†] AND YITING YANG[†]

Abstract. Sorting permutations by transpositions is an important and difficult problem in genome rearrangements. The transposition diameter $TD(n)$ is the maximum transposition distance among all pairs of permutations in S_n. It was previously conjectured [H. Eriksson et al., *Discrete Math.*, 241 (2001), pp. 289–300] that $TD(n) \leq \lceil \frac{n+1}{2} \rceil$. This conjecture was disproved by Elias and Hartman [*IEEE/ACM Trans. Comput. Biol. Bioinform.*, 3 (2006), pp. 369–379] by showing $TD(n) \geq \lfloor \frac{n+1}{2} \rfloor + 1$. In this paper we improved the lower bound to $TD(n) \geq \frac{17}{33}n + \frac{1}{33}$ via computation.

1. Introduction. It has been observed that the genomes of some close species have essentially the same set of genes, but in different orders. It is widely believed that this aspect is caused by a sequence of some evolutionary events, which transform one genome into another without changing the gene sequences. The genome rearrangement problem seeks to reconstruct the evolutionary history on these events. Significant progress has been made in last decade for the problem of genome rearrangements. Genomes are represented by permutations (or signed permutations). The basic problem is as follows: given two permutations σ and τ, find the shortest sequences of basic rearrangement operations that transform σ to τ. The basic rearrangement operations include reversals, transpositions, insertions, deletions, translocations, etc. A signed (or unsigned) reversal reverses a block of signed (or unsigned) permutation of genes. Hannenhalli and Pevzner [7] first proposed a polynomial-time algorithm for sorting signed permutations by reversals. Caprara [3] showed that sorting unsigned permutations by reversals was NP-hard. A transposition exchanges two adjacent blocks. The complexity problem of sorting permutations by transpositions is still open. Elias and Hartman [6] provided a polynomial-time 1.375-approximation algorithm for sorting by transpositions.

In 1996, Bafna and Pevzner [1] proved that the *diameter* of unsigned permutations of length n under reversal is $n - 1$. Three years later, Hannenhalli and Pevzner [7] found that the diameter of signed permutations of length n under reversal is $n + 1$. The transposition diameter $TD(n)$ has not been determined exactly. In 1998, Bafna and Pevzner [2] proved that

$$\left\lceil \frac{n}{2} \right\rceil \leq TD(n) \leq \left\lfloor \frac{3n}{4} \right\rfloor.$$

Eriksson et al. [5] improved the bounds to

$$\left\lceil \frac{n+1}{2} \right\rceil \leq TD(n) \leq \left\lfloor \frac{2n-2}{3} \right\rfloor.$$

Their lower bound is achieved by the reverse permutation $(n, n-1, \ldots, 2, 1)$.

Recently, Elias and Hartman [6] improved the lower bound to

$$TD(n) \geq \left\lfloor \frac{n+1}{2} \right\rfloor + 1.$$

Here we prove the following theorem.

THEOREM 1.

$$TD(n) \geq \frac{17n}{33} + \frac{1}{33}.$$

The main idea is to create permutations whose *breakpoint graphs* (defined later) are disjoint unions of *super-bad* cycles. Roughly speaking, the super-bad cycle (or permutation) is a cycle which cannot be broken easily to small good cycles and more importantly, cannot be merged into a good cycle. The detailed definition of a super-bad cycle is given in section 3. The super-bad cycles are quite rare. In fact, we performed an extensive computational search using a *Maple* program and found only one super-bad permutation of order 10. Each super-bad cycle will provide a nontrivial lower bound of $TD(n)$ as shown in the following theorem. Theorem 1 is a special case of Theorem 2 by setting $b = 10$.

THEOREM 2. *If there exists a super-bad permutation of order b, then*

$$TD(n) \geq \frac{(3b+4)}{6(b+1)}n + \frac{1}{3(b+1)}.$$

The remainder of the paper is organized as follows. Notations and lemmas are given in section 2. Theorem 2 will be proved in section 3. In section 4, we will show that a super-bad permutation of order 10 exists. Comments and remarks will be given in the last section.

2. Notations and lemmas.

2.1. Transpositions and transposition diameter. Let S_n be the group of all permutations over $[n] = \{1, 2, \ldots, n\}$. For a permutation $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n) \in S_n$, a *block* (or *segment*) of σ is a sequence of consecutive elements $\sigma_{i+1}, \sigma_{i+2}, \ldots, \sigma_j$ for some $j > i \geq 0$. Here i and j are called *cuts* of this segment, and this block is called an ij-*block*.

For any $0 \leq i < j < k \leq n$, a *transposition* $\rho_{i,j,k}$ is a mapping from S_n to S_n by swapping the ij-block with the jk-block. Namely, for any $\sigma \in S_n$, we have

$$\rho_{i,j,k}(\sigma) = \sigma_1, \ldots, \sigma_i, \sigma_{j+1}, \ldots, \sigma_k, \sigma_{i+1}, \ldots, \sigma_j, \sigma_{k+1}, \ldots, \sigma_n.$$

The *transposition graph* on S_n is a graph $G = (V, E)$ with $V = S_n$ and $E = \{\sigma\tau \mid \exists i < j < k \text{ so that } \tau = \rho_{i,j,k}(\sigma)\}$. The *transposition distance* $d_T(\sigma, \tau)$ is the graph distance of σ and τ in G. The *transposition diameter*, denoted by $TD(n)$, is the diameter of the transposition graph.

It is easy to see that S_n acts on G transitively. We have

$$d_T(\sigma, \tau) = d_T(id, \sigma\tau^{-1}) = d_T(id, \tau\sigma^{-1}),$$

where $id = (1, 2, \ldots, n)$ is the identity permutation.

The transposition diameter is the maximum transposition distance between a permutation and id over S_n.

2.2. Breakpoint graphs. The breakpoint graphs of permutations were first introduced by Bafna and Pevzner [1] and were widely used in the genome rearrangements literature [1, 2, 3, 4, 6, 7, 9].

Let $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n)$ be a permutation in S_n. The *breakpoint graph* $BP(\sigma)$ is an edge-colored graph on $2n + 2$ vertices $\{L_0, R_0, L_1, R_1, \ldots, L_n, R_n\}$. For $0 \leq i \leq n$, connect R_{σ_i} and $L_{\sigma_{i+1}}$ by a black edge and connect L_{i+1} and R_i by a gray edge. Here we used the convention that indices are modulo by $n+1$ (i.e., $L_{n+1} = L_0$, $R_{n+1} = R_0$, and $R_{\sigma_0} = R_0$).

We can orient black edges from R_* to L_* and gray edges from L_* to R_*. This is called the *natural orientation* of the breakpoint graph. Each vertex in the breakpoint graph has exactly one edge coming in and exactly one edge coming out. The in-edge and out-edge are in different colors. As a result, the breakpoint graph is decomposed into cycles of even lengths. The number of black edges and the number of gray edges in each cycle are the same. For $k \geq 1$, a cycle with exactly k black edges (also k gray edges) is called a k-cycle. Bafna and Pevzner [1] defined *odd* cycles to be k-cycles with odd k.

We draw $BP(\sigma)$ as follows. We put $(L_0, R_0, L_1, R_1, \ldots, L_n, R_n)$ on the vertices of regular $2(n + 1)$-polygon counterclockwise. All gray edges appear as sides of this polygon and are oriented counterclockwise. Black edges can be drawn as straight line segments. This drawing has an advantage of showing the structure of unions (to be defined later) in a permutation in comparison to the traditional drawing in [1, 2].

For $i < j < k$, the transposition $\rho_{i,j,k}$ acts on the breakpoint graph $BP(\sigma)$ by rewiring three black edges. It replaces $R_{\sigma_i} L_{\sigma_{i+1}}$, $R_{\sigma_j} L_{\sigma_{j+1}}$ and $R_{\sigma_k} L_{\sigma_{k+1}}$ by $R_{\sigma_i} L_{\sigma_{j+1}}$, $R_{\sigma_j} L_{\sigma_{k+1}}$, and $R_{\sigma_k} L_{\sigma_{i+1}}$.

The ordering inherited from the natural orientation is called the *natural ordering*, while the ordering inherited from the regular polygon is called the *drawing ordering*. The effects of the transposition $\rho_{i,j,k}$ acting on the breakpoint graph $BP(\sigma)$ can be described as follows. This is simply a rephrasing of Bafna and Pevzner's [1] result. Note that each three vertices in a cycle of a breakpoint graph receive two orderings.

Case 1. Three black edges belong to the same cycle C, and their natural ordering in C is different from the drawing ordering. The rewiring breaks this cycle into three smaller cycles.

Case 2. Three black edges belong to the same cycle C, and their natural ordering in C is the same as the drawing ordering. After rewiring, there is only one cycle.

Case 3. Two black edges belong to one cycle, and the third edge belongs to the other cycle. The rewiring cuts a part of the first cycle and pastes the part to the second cycle.

Case 4. This is a reverse action of the first case. All three black edges belong to different cycles. After rewiring, three smaller cycles are merged into a big cycle C, and their natural ordering in C is different from the drawing ordering.

Let $C_{odd}(\sigma)$ be the number of odd cycles in $BP(\sigma)$. The transposition $\rho_{i,j,k}$ affects the number of odd cycles by $\{-2, 0, 2\}$. For $i = -2, 0, 2$, a transposition is called an i-move if it affects the number of odd cycles by i. Since $C_{odd}(id) = n + 1$, we have the following lower bound (see [2]):

$$(1) \qquad d_T(\sigma) \geq \frac{n + 1 - C_{odd}(\sigma)}{2}.$$

2.3. Union. For two permutations $\sigma \in S_n$ and $\tau \in S_m$, the union of σ and τ, denoted by $\sigma \uplus \tau$, is a new permutation in S_{n+m+1}:

$$(\sigma \uplus \tau) = (\sigma_1, \ldots, \sigma_n, n + 1, n + 1 + \tau_1, \ldots, n + 1 + \tau_m).$$
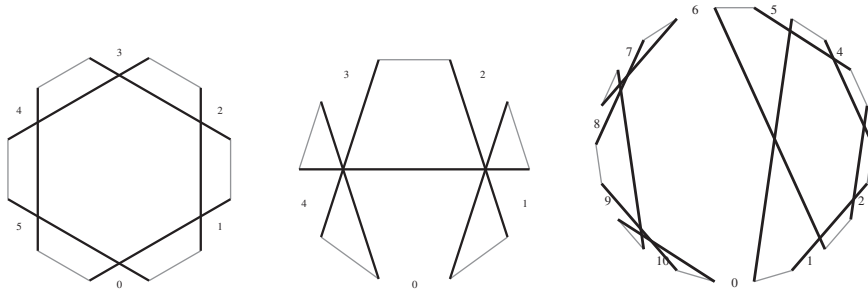
FIG. 1. *The breakpoint graphs of σ, τ, and $\sigma \uplus \tau$.*
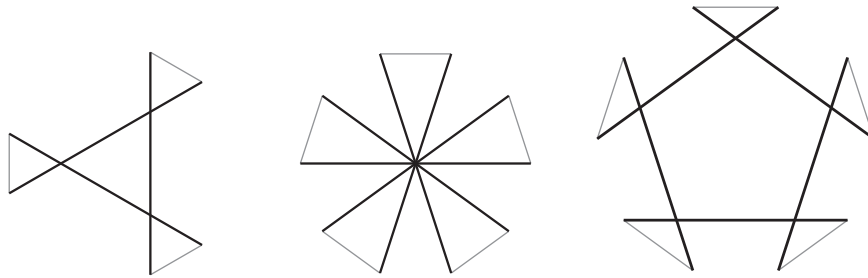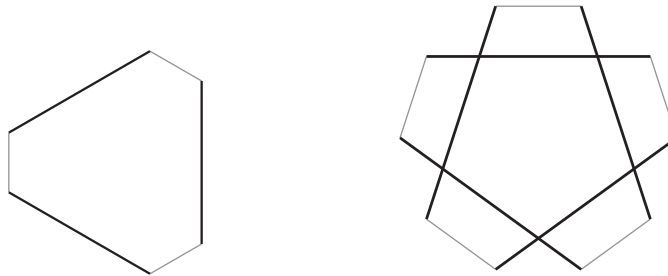

FIG. 2. *Examples of good cycles.*

FIG. 3. *Examples of bad cycles.*

For example, $\sigma = (5, 4, 3, 2, 1)$, $\tau = (2, 1, 4, 3)$, and we have $\sigma \uplus \tau = (5, 4, 3, 2, 1, 6, 8, 7, 10, 9)$ (see Figure 1).

In our drawing, the breakpoint graph of $\sigma \uplus \tau$ is the disjoint union of the breakpoint graphs of σ and τ under *homeomorphism* (defined later).

3. New lower bounds of transposition distance. In the breakpoint graph $BP(\sigma)$, odd cycles can be divided into two classes. A $2k + 1$-cycle is called *good* if it can be transformed into a set of 1-cycles only by k 2-moves while keeping other cycles invariant (see Figure 2). Otherwise it is called *bad* (see Figure 3). For example, Eriksson et al. [5] proved that the breakpoint graph of the reverse permutation $(r - 1, r - 2, \ldots, 2, 1)$ is a bad odd cycle when r is odd.

We denote the number of good and bad cycles in $BP(\sigma)$ by $C_{good}(\sigma)$ and $C_{bad}(\sigma)$.

LEMMA 3. *Suppose a permutation σ' can be obtained from a permutation σ by one transposition. Then*

$$(2) \qquad\qquad |C_{good}(\sigma') - C_{good}(\sigma)| \leq 2.$$

Proof. By symmetry, it suffices to show

$$(3) \qquad C_{good}(\sigma') - C_{good}(\sigma) \le 2.$$

Inequality (3) holds if the transposition creates at most two good odd cycles. It is possible that the transposition can create three good cycles. But in this case, the transposition breaks a big cycle into three good cycles. The big cycle must be good by definition. Thus, the number of good cycles increases by two in this case. □

Note that $C_{good}(id) = n + 1$. We have a useful corollary from the above lemma.

COROLLARY 4. *Let $\sigma \in S_n$ and $BP(\sigma)$ be its breakpoint graph. Then*

$$(4) \qquad d_T(\sigma) \ge \frac{n + 1 - c_{good}(\sigma)}{2}.$$

We say a permutation $\gamma = (\gamma_1, \gamma_2, \dots \gamma_{b-1}) \in S_{b-1}$ is *super-bad* if
1. $BP(\gamma)$ consists of exactly one odd cycle,
2. there is no 2-move that can be applied on $BP(\gamma)$,
3. $d_T(\gamma \uplus \gamma) \ge b + 1$.

We say an odd cycle C (in $BP(\sigma)$) is *homeomorphic* to $BP(\gamma)$ if there is a cyclic-order-preserving map π from the vertices of $BP(\gamma)$ to the vertices of $BP(\sigma)$ satisfying

$$\pi(BP(\gamma)) = C.$$

We say C is of γ-type if
1. C is homeomorphic to $BP(\gamma)$,
2. the vertices of C are consecutive in $BP(\sigma)$.

Let C_{bad}^{γ} be the number of bad cycles of γ-type. Let $C_{bad}^{\ne\gamma}(\sigma)$ be the number of bad cycles which are not γ-type.

THEOREM 5. *Suppose $\gamma \in S_{b-1}$ is a super-bad permutation. For any permutation $\sigma \in S_n$, we have*

$$(5) \qquad d_T(\sigma) \ge \frac{n + 1 - C_{good}(\sigma)}{2} - \frac{C_{bad}^{\ne\gamma}(\sigma)}{6} + \frac{C_{bad}^{\gamma}(\sigma)}{6}.$$

Proof. Let $f(\sigma) = \frac{C_{good}(\sigma)}{2} + \frac{C_{bad}^{\ne\gamma}(\sigma)}{6} - \frac{C_{bad}^{\gamma}(\sigma)}{6}$. The identity permutation id has $n + 1$ good cycles and no bad cycle. Thus $f(id) = \frac{n+1}{2}$. It suffices to show any transposition can change the value of f by at most 1. Namely, let $\sigma' = \rho(\sigma)$ be the permutation by applying the transposition ρ on a permutation σ. We claim

$$(6) \qquad f(\sigma') - f(\sigma) \le 1.$$

Note that any action of a transposition can affect at most four cycles. We have

$$\left(C_{bad}^{\ne\gamma}(\sigma') - C_{bad}^{\gamma}(\sigma') \right) - \left(C_{bad}^{\ne\gamma}(\sigma) - C_{bad}^{\gamma}(\sigma) \right) \le 4.$$

The equality only holds if ρ transforms a γ-type odd cycle into three bad not-γ-type bad cycles. This never happens because we assume that γ has no 2-move. Thus we have an improved inequality

$$C_{bad}^{\ne\gamma}(\sigma') - C_{bad}^{\gamma}(\sigma') - C_{bad}^{\ne\gamma}(\sigma) + C_{bad}^{\gamma}(\sigma) \le 3.$$

Note inequality (6) holds trivially if $C_{good}(\sigma') - C_{good}(\sigma) \le 1$. From now on, we assume

$$C_{good}(\sigma') - C_{good}(\sigma) = 2.$$

We will show

$$C_{bad}^{\neq\gamma}(\sigma') - C_{bad}^{\gamma}(\sigma') \leq C_{bad}^{\neq\gamma}(\sigma) - C_{bad}^{\gamma}(\sigma). \tag{7}$$

There are two cases of the action of ρ.

Case 1. ρ breaks one big cycle into three small cycles. At least two of three cycles are good (odd) cycles.

If the third small cycle is even (or good), then the big cycle is even (or good). In both cases, γ-type and non-γ-type cycles do not change. Inequality (7) holds.

If the third small cycle is bad, then the big cycle is also bad. Since γ has no 2-move, the big cycle is not of γ-type. If the third small cycle is not of γ-type, we have

$$C_{bad}^{\neq\gamma}(\sigma') = C_{bad}^{\neq\gamma}(\sigma),$$
$$C_{bad}^{\gamma}(\sigma') = C_{bad}^{\gamma}(\sigma).$$

If the third small cycle is of γ-type, we have

$$C_{bad}^{\neq\gamma}(\sigma') = C_{bad}^{\neq\gamma}(\sigma) - 1,$$
$$C_{bad}^{\gamma}(\sigma') = C_{bad}^{\gamma}(\sigma) + 1.$$

Inequality (7) still holds.

Case 2. ρ transforms two cycles C_1 and C_2 into two good cycles. The two cycles C_1 and C_2 are either both even or both bad.

If C_1 and C_2 are even cycles, the γ-type and non-γ-type cycles do not change. Inequality (7) holds.

Two γ-type cycles cannot be transformed into two good cycles because of $d(\gamma \uplus \gamma) \geq b + 1$. At least one of C_1 and C_2 is non-γ-type. Thus, inequality (7) still holds. The proof of the theorem is finished. $\square$

Proof of Theorem 2. Let γ be a super-bad permutation of order b. For any positive integer n, we write $n + 1 = (b + 1)k + r$ with $0 \leq r \leq b$. Let σ be the unions of k copies of γ and one copy of $\alpha = (r - 1, r - 2, \ldots, 1)$. We have

$$\sigma = \gamma \uplus (\gamma \uplus (\cdots \uplus (\gamma \uplus \alpha) \cdots)).$$

Eriksson et al. [5] showed the breakpoint graph of the reverse permutation $(r - 1, r - 2, \ldots, 2, 1)$ is a bad odd cycle when r is even, two bad odd cycles when $r = 4k + 1$, and two even cycles when $r = 4k + 3$. It implies $f_\gamma(\sigma) = k$, $C_{good}(\sigma) = 0$, and $C_{bad}^{\neq\gamma}(\sigma) = 1$ if r is even, and $C_{bad}^{\neq\gamma}(\sigma) = 0$ or 2 if r is odd.

Applying Theorem 5, we have

$$TD(n) \geq d(\sigma)$$

$$\geq \frac{n + 1 - C_{good}(\sigma)}{2} - \frac{C_\tau^{\neq\gamma}(\sigma)}{6} + \frac{k}{6}$$

$$\geq \frac{n + 1}{2} - \frac{2}{6} + \frac{k}{6}$$

$$= \frac{n}{2} + \frac{1}{6} + \frac{n + 1 - r}{6(b + 1)}$$

$$= \frac{3b + 4}{6(b + 1)} n + \frac{b + 2 - r}{6(b + 1)}$$

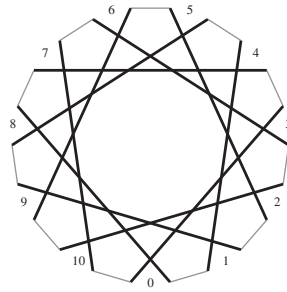$$\geq \frac{3b + 4}{6(b + 1)} n + \frac{1}{3(b + 1)}. \qquad \square$$

FIG. 4. *The break-point graph of* $\gamma = (8, 5, 2, 10, 7, 4, 1, 9, 6, 3)$.

4. Searching for super-bad permutations. We used *Maple* to search super-bad permutations. We discovered that $\gamma = (8, 5, 2, 10, 7, 4, 1, 9, 6, 3) \in S_{10}$ is super-bad.

From Figure 4, $BP(\gamma)$ contains one 11-cycle and is symmetric under rotations and reflections. To show that γ has no 2-move, we need only to check

$$\frac{1}{11}\binom{11}{3} = 15$$

cases. This was done by a *Maple* program exhaustively.

Since γ has no 2-move, $\gamma \uplus \gamma$ has no 2-move, too. For any transposition ρ, $\rho(\gamma \uplus \gamma)$ has exactly two cycles.

Claim A. These two cycles cannot both be good.

Applying Corollary 4 to $\rho(\gamma \uplus \gamma)$, we have

$$d(\rho(\gamma \uplus \gamma)) \geq \frac{21 + 1 - 1}{2} = 10.5.$$

Thus, $d(\rho(\gamma \uplus \gamma)) \geq 11$ for any ρ. We have

$$d(\gamma \uplus \gamma) \geq 12$$

as required.

To prove Claim A, we need check whether a given odd cycle is good. This is done by a bread-first search algorithm to try all possible 2-moves. We find that it always ends up with a permutation containing one of the two bad cycles in Figure 3.

The *Maple* code is posted at [8].

5. Concluding remarks. The main idea of this paper is the existence of a super-bad cycle. The super-bad cycle can not be broken into small cycles easily, and more important, the union of two such cycles is also difficult to break. Super-bad cycles are rare. We searched a special class of (symmetric) permutations and found only one super-bad permutation, namely, $\gamma = (8, 5, 2, 10, 7, 4, 1, 9, 6, 3)$. Note that the breakpoint graph of every permutation in S_{2k+1} consists of an even number of odd cycles. Thus, every super-bad permutation has even order. We also exhaust all permutations of orders $2, 4, 6, 8, 10$. No other super-bad permutation of order at most 10 exists. Thus, γ has the smallest order among all super-bad permutations. Apparently a new idea is needed in order to improve our lower bound.

A possible improvement is to find a super-bad permutation σ with an additional property "The union of three σ's cannot be merged into a good cycle in one step."

If such σ exists, then a similar argument shows $TD(n) \geq \frac{3|\sigma|+5}{6(|\sigma|+1)} + O(1)$. We do not know if such a permutation exists. In contrast, the union of three copies of $\gamma = (8, 5, 2, 10, 7, 4, 1, 9, 6, 3)$ can be merged into a good cycle.

Acknowledgments. We thank the anonymous referees for useful comments in preparing this version.

REFERENCES

[1] V. BAFNA AND P.A. PEVZNER, *Genome rearrangements and sorting by reversals*, SIAM J. Comput., 25 (1996), pp. 272–289.

[2] V. BAFNA AND P.A. PEVZNER, *Sorting by transpositions*, SIAM J. Discrete Math., 11 (1998), pp. 224–240.

[3] A. CAPRARA, *Sorting permutations by reversals and Eulerian cycle decompositions*, SIAM J. Discrete Math., 12 (1999), pp. 91–110.

[4] D.A. CHRISTIE, *Sorting permutations by block-interchanges*, Inform. Process. Lett., 60 (1996), pp. 165–169.

[5] H. ERIKSSON, K. ERIKSSON, J. KARLANDER, L. SVENSSON, AND J. WASSTLUND, *Sorting a bridge hand*, Discrete Math., 241 (2001), pp. 289–300.

[6] I. ELIAS AND T. HARTMAN, *A 1.375-approximation algorithm for sorting by transpositions*, IEEE/ACM Trans. Comput. Biol. Bioinform., 3 (2006), pp. 369–379.

[7] S. HANNENHALLI AND P.A. PEVZNER, *Transforming cabbage into turnip: Polynomial algorithm for sorting signed permutations by reversals*, J. ACM, 46 (1999), pp. 1–27.

[8] *Maple Code*, http://www.math.sc.edu/∼lu/tdiam.mw.

[9] J. MEIDANIS, M.E.M.T. WALTER, AND Z. DIAS, *A lower bound on the reversal and transposition diameter*, J. Comput. Biol., 9 (2002), pp. 743–745.