Journal of Mathematical Chemistry Vol. 42, No. 1, July 2007 (© 2006)
DOI: 10.1007/s10910-006-9095-8

Estimate the distance of genome rearrangements by reversals

Jianxiu Hao
Department of Mathematics, Zhejiang Normal University, Jinhua, Zhejiang 321004, P.R. China
E-mail: sx35@zjnu.cn

Received 31 January 2005; revised 15 February 2006

We study genome rearrangements by reversals and present a modified lower bound and an almost holding upper bound for this problem.

KEY WORDS: genome rearrangement, reversal, gene

AMS subject classification: 92D20

1. Introduction

In order to derive evolutional and fundamental relationships between genes, sequence comparison is a useful tool. However, classical alignment algorithms deal with only local mutations (i.e., insertions, deletions, and substitutions of nucleotides) and ignore the global rearrangements (e.g., reversals, transpositions and translocations of long fragments). In [1], Palmer and Herbon found that the rearrangements of mitochondrial genomes of Brassica (cabbage) and Brassica campestris (turnip) are with 99–99.9% identical genes. They discovered that these molecules, which are almost identical in gene sequence, differ in gene order. The classical methods of sequence comparison are not very useful to analyze highly rearranged genomes [2,3]. Genome rearrangement is a common mode of molecular evolution in mitochondrial, chloroplast, viral, bacterial DNA, and human red–green color blindness [4–9].

For example, two chromosomes with homolgous blocks [10]:

2 1 3 7 5 4 8 6 and 1 2 3 4 5 6 7 8.

A series of reversals that sort permutation 2 1 3 7 5 4 8 6 to 1–8 is given as follows:

$$\underline{21}375486,$$

$$123\underline{754}86,$$

$$12345\underline{786},$$

$$1234\underline{5768},$$

$$12345\underline{76}8,$$

$$12345678.$$

Let $\pi = (\pi_1, \ldots, \pi_n)$ be a permutation of $\{1, \ldots, n\}$, and denote I the identity permutation$(12\ldots n)$. A reversal of the interval (i, j) is an inversion of the subsequence $\pi_i \ldots \pi_j$ of π, represented by the permutation $\rho = (1 \ldots i-1 j \ldots i j+1 \ldots n)$. Composition of π with ρ yields $\pi\rho = (\pi_1 \ldots \pi_{i-1}\pi_j \ldots \pi_i \pi_{j+1} \ldots \pi_n)$, where elements $\pi_i \ldots \pi_j$ have been reversed. The problem of sorting a permutation by the minimum number of reversals (MIN-SBR) is defined as follows:

MIN–SBR: given a permutation π, find the shortest sequence of reversals $\rho_1 \ldots \rho_{d(\pi)}$ such that $\pi\rho_1 \ldots \rho_{d(\pi)} = I$.

The optimal solution value $d(\pi)$ is called the reversal distance of π [11].

Consider a permutation $\pi = (\pi_1 \ldots \pi_n)$ of $\{1 \ldots n\}$. A long strip of π is a subsequence $\pi_i \ldots \pi_j$ of π such that $j > i + 1$ and either $\pi_k = \pi_{k-1} + 1$ for $k = i+1, \ldots, j$ or $\pi_k = \pi_{k-1} - 1$ for $k = i+1, \ldots, j$. In other words, a long strip of π corresponds to three or more elements, which appear in the same order or reverse order in π and I. As far as MIN-SBR is concerned, Hannenhalli and Pevzner proved that one can assume without loss of generality that π does not contain any long strips; we therefore, make this assumption in the remainder of the paper. We also assume without loss of generality $\pi_1 \neq 1$ and $\pi_n \neq n$ [11].

The organization of the paper is as follows. In section 2, we present some definitions and lemmas, section 3 is devoted to the main result, we give some remarks in section 4.

2. Preliminaries

Definition 2.1 [11]. Define the breakpoint graph $G(\pi) = (V, R \cup B)$ of π as follows. Add to π the elements $\pi_0 := 0$ and $\pi_{n+1} := n + 1$, redefining $\pi := (0\pi_1 \ldots \pi_n n{+}1)$. Also, let the inverse permutation π^{-1} of π be defined by $\pi^{-1}_{\pi_i} := i$ for $i = 0, \ldots, n + 1$. Let $V := \{0, \ldots, n + 1\}$, where each vertex $v \in V$ represents an element of π. Graph $G(\pi)$ is bicolored, i.e., its edge set is partitioned into two subsets, each represented by a different color. R is the set of, say, red edges, each of the form (π_i, π_{i+1}), for all $i \in \{0, \ldots, n\}$ such that $|\pi_i - \pi_{i+1}| \neq 1$, i.e., elements, which in consecutive positions in π but not in the identity permutation I. Such a pair π_i, π_{i+1} is called a breakpoint of π. Let $b(\pi) := |R|$ be the number of breakpoints of π. B is the set of, say, blue edges, each of the form $(i, i + 1)$, for all $i \in \{0, \ldots, n\}$ such that $|\pi^{-1}_i - \pi^{-1}_{i+1}| \neq 1$, i.e., elements which are in consecutive positions in I but not in π. Note that each vertex $i \in V$ has either degree 2 or 4, and has the same number of incident blue and red edges. Therefore, $|R| = |B|(= b(\pi))$. The fact that $G(\pi)$ has no vertices of degree 0 follows from the assumption that π contains no long strips.

Definition 2.2 [11]. An alternating cycle of $G(\pi)$ is a sequence of edges $r_1, b_1, r_2, b_2, \ldots, r_m, b_m$, where $r_i \in R$, $b_i \in B$ for $i = 1, \ldots, m$; r_i and b_j are incident to a common vertex for $i = j$, $i = 1, 2, \ldots, m$ and for $i = j + 1$, $j = 1, \ldots, m$ (where $r_{m+1} := r_1$); and $r_i \neq r_j$, $b_i \neq b_j$ for $1 \leqslant i < j \leqslant m$. Note that an alternating cycle may be not a cycle in common sense. For example, two 3-cycles with one vertex in common may form an alternating cycle.

Definition 2.3 [11]. An alternating cycle decomposition of $G(\pi)$ is a collection of edge-disjoint alternating cycles such that every edge of $G(\pi)$ is contained in exactly one cycle of the collection. It is easy to see that $G(\pi)$ always admits an alternating cycle decomposition. For a given π let $c(\pi)$ be the maximum cardinality of an alternating cycle decomposition of $G(\pi)$.

Lemma 2.4 [6,12]. For every permutation π, $d(\pi) \geqslant b(\pi) - c(\pi)$, where $b(\pi)$ is defined in definition 2.1, $c(\pi)$ is defined in definition 2.3, and $d(\pi)$ is the reversal distance of π defined in section 1.

Definition 2.5. We call a graph G even graph if and only if it has no vertex with odd degree. For example, all eulerian graphs are even graphs.

Definition 2.6. For a given π let $c_1(\pi)$ be the maximum cardinality cycle decomposition of $G(\pi)$. Similarly, let $c_1(G)$ be the maximum cardinality cycle decomposition of even graph G.

Definition 2.7. A generalized tree GT is a plane even graph obtained from a tree T by using some cycles to replace some vertices of T (may not replace all vertices of T) and contracting each edge of T.

For example, trivially, an isolated vertex is a GT, a cycle is a GT too, the following graph G is a GT obtained from T:

$$V(T) = \{u_1, u_2, u_3, u_4, u_5\},$$

$$E(T) = \{u_1u_2, u_2u_3, u_3u_4, u_3u_5\}.$$

In order to obtain G, we use cycle $v_1v_2v_3v_4v_1$ to replace u_1, cycle $v_4v_5v_6v_4$ to replace u_2, cycle $v_5v_7v_{10}v_5$ to replace u_3, cycle $v_7v_8v_9v_7$ to replace u_4, cycle $v_{10}v_{11}v_{12}v_{10}$ to replace u_5, and contract every edge of T. That is,

$$\begin{aligned} V(G) &= \{v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8, v_9, v_{10}, v_{11}, v_{12}\}, \\ E(G) &= \{v_1v_2, v_2v_3, v_3v_4, v_4v_1, v_4v_5, v_5v_6, v_6v_4, \\ &\qquad v_5v_7, v_7v_{10}, v_{10}v_5, v_7v_8, v_8v_9, v_9v_7, v_{10}v_{11}, v_{11}v_{12}, v_{12}v_{10}\}. \end{aligned}$$

Note that in G we may put cycles $v_4v_5v_6$ and $v_1v_2v_3v_4$ into cycle $v_5v_7v_{10}$, and let cycles $v_7v_8v_9$ and $v_{10}v_{11}v_{12}$ remain in the exterior of cycle $v_5v_7v_{10}$.

Lemma 2.8 [13].

$$2|E(G)| = \sum_{v \in V(G)} d(v),$$

where $|E(G)|$ is the edge number of G, $d(v)$ is the degree of vertex v.

Lemma 2.9 [13]. If G is a plane graph, then

$$|F(G)| = |E(G)| - |V(G)| + \omega(G) + 1,$$

where $|F(G)|$ is the face number of G, $|E(G)|$ is the edge number of G, $|V(G)|$ is the vertex number of G, $\omega(G)$ is the component number of G.

Definition 2.10. Let G be a plane even graph. For the cycle decomposition of G, if each cycle constitues the boundary of a face by itself, we call the cycle decomposition further as non-intersecting cycle decomposition of G. Let $c_2(G)$ be the maximum cardinality of non-intersecting cycle decomposition of G. An example is given in definition 2.11.

Definition 2.11. For a plane even graph G, let its non-intersecting cycle decomposition be $C_1, C_2, \ldots, C_m$. When $m \geqslant 2$, a non-induced face is a face whose boundary is a cycle C_i, where $i = 1, 2, \ldots, m$. Otherwise, a face is called an induced face. When $m = 1$, we call the interior face a non-induced face and the exterior face an induced face. For example, define $G = (V, E)$ as follows:

$$V(G) = \{v_1, v_2, v_3, v_4, v_5, v_6, v_7\},$$

$$E(G) = \{v_1v_2, v_2v_3, v_3v_4, v_4v_6, v_6v_1, v_2v_5, v_5v_4, v_4v_7, v_7v_2\}.$$

Suppose the non-intersecting cycle decomposition of G is as follows:

$$C_1 = v_1v_2v_3v_4v_6v_1,$$

$$C_2 = v_2v_5v_4v_7v_2.$$

Since $\{C_1, C_2\}$ is a non-intersecting cycle decomposition of G, we assume that v_5 and v_7 are in the exterior of C_1 and v_2v_5, v_5v_4, v_4v_7, v_7v_2 are on the same side of C_1. The faces f_1, f_2, f_3, f_4 of G are defined as follows: the boundary of f_1 is C_1, the boundary of f_2 is C_2, the boundary of f_3 is $C_3 = v_2v_3v_4v_5v_2$ or $v_2v_3v_4v_7v_2$, f_4 is the unbounded face, that is, f_4 is the exterior face of G.

By the definition of induced face and non-induced face, we know that f_1 and f_2 are non-induced faces, f_3 and f_4 are induced faces.

Lemma 2.12 [10].

$$d(\pi) \geqslant \frac{1}{2}b(\pi),$$

where $d(\pi)$ is the reversal distance of π defined in section 1, $b(\pi)$ is defined in definition 2.1.

Lemma 2.12 is equivalent to the following Claim.
Claim: For permutation π, we have

$$d(\pi) \geqslant \lceil \frac{1}{4}(m+2n) \rceil,$$

where $d(\pi)$ is the reversal distance of π which is defined in Section 1, m the vertex number with degree 2 in $G(\pi)$, and n is the vertex number with degree 4 in $G(\pi)$.

In fact, by lemma 2.8 we have $|E(G(\pi))| = m + 2n$. By definition 2.1, we have $|E(G(\pi))| = 2b(\pi)$. Thus, $b(\pi) = \frac{1}{2}(m+2n)$. The equivalence follows.

3. Main result

Theorem 3.1. For permutation π, if $G(\pi)$ is a plane graph we have

$$d(\pi) \geqslant \frac{1}{2}m - \omega(G(\pi)).$$

Further, when $G(\pi)$ is not the union of some generalized trees, we have

$$d(\pi) \geqslant \frac{1}{2}m - \omega(G(\pi)) + 1,$$

where $G(\pi)$ is defined in definition 2.1, m denotes the number of vertices with degree 2 in $G(\pi)$, $\omega(G(\pi))$ is the number of components in $G(\pi)$, $d(\pi)$ is the reversal distance of π defined in Introduction.

Proof.
Claim 1: Let G be a plane even graph. When G is an empty graph, we have $c_1(G) = 0$. Otherwise, we have

$$\frac{1}{2}|F(G)| \leqslant c_1(G) \leqslant |F(G)| - 1;$$

$c_1(G) = |F(G)| - 1$ if and only if G is the union of some generalized trees. When $\frac{1}{2}|F(G)| = c_1(G)$, G contains exactly one non-trivial component without cut vertex, where $|F(G)| = |E(G)| - |V(G)| + \omega(G) + 1$, $\omega(G)$ denotes the number of components of G, $|F(G)|$ is the number of faces of G, $c_1(G)$ is defined in definition 2.6.

In fact, when G is an empty graph, Claim 1 holds obviously. In the following we suppose that G is not an empty graph.

Claim 2: If there exists a cycle decomposition of G, there exists a non-intersecting cycle decomposition of G, and $c_1(G) = c_2(G)$, where $c_1(G)$ and $c_2(G)$ are defined in definitions 2.6 and 2.10, respectively.

In fact, let the cycle decomposition of G be $C_1, C_2, \dots, C_i, \dots, C_j, \dots, C_k$, where C_i and C_j are two intersecting cycles in the cycle decomposition of G above and let

$$V(C_i) = \{u_1, u_2, \dots, u_a, u_{a+1}, \dots, u_b, u_{b+1}, \dots, u_s\},$$

$$E(C_i) = \{u_1u_2, u_2u_3, \dots, u_{a-1}u_a, \dots, u_bu_{b+1}, \dots, u_su_1\},$$

$$V(C_j) = \{v_1, v_2, \dots, v_t, v_{t+1}, \dots, v_h, v_{h+1}, \dots, v_g\},$$

$$E(C_j) = \{v_1v_2, v_2v_3, \dots, v_{t-1}v_t, v_tv_{t+1}, \dots, v_hv_{h+1}, \dots, v_gv_1\},$$

where $u_a = v_t$, $u_b = v_h$, $V(C_i) \cap V(C_j) \supseteq \{u_a, u_b\}$. Without loss of generality, let $\{v_{t+1}, v_{t+2}, \dots, v_{h-1}\}$ be in the interior of C_i.

We construct two new cycles as follows:

$$C_i^{''} = u_1u_2 \dots u_a v_{t+1} v_{t+2} \dots v_{h-1} u_b u_{b+1} \dots u_s u_1,$$

$$C_j^{''} = v_1v_2 \dots v_t u_{a+1} u_{a+2} \dots u_{b-1} v_h v_{h+1} \dots v_g v_1.$$

By definition 2.10, if $C_i^{''}$ and $C_j^{''}$ are not two non-intersecting cycles, we can do as above. At last, we obtain $C_i^{'}$ and $C_j^{'}$ are two non-intersecting cycles. It is easy to see that $C_1, C_2, \dots C_{i-1}, C_i^{'}, C_{i+1}, \dots, C_{j-1}, C_j^{'}, C_{j+1}, \dots, C_k$ is a cycle decomposition of G.

If there were two intersecting cycles in $C_1, \dots, C_{i-1}, C_i^{'}, C_{i+1}, \dots, C_{j-1}, C_j^{'}, C_{j+1}, \dots, C_k$ further, we could do as above. Hence, there exists a non-intersecting cycle decomposition of G.

By definitions 2.6 and 2.10, we have $c_2(G) \leqslant c_1(G)$. From the above construction we obtain $c_1(G) \leqslant c_2(G)$. Thus, $c_1(G) = c_2(G)$. The claim follows.

Let $c_1(G) = l$ and $C_1, C_2, \dots, C_l$ be the maximum cardinality non-intersecting cycle decomposition of G. For a plane embedding of G, we find the corresponding edges of $C_1, C_2, \dots, C_l$. Note that it is possible $|V(C_i) \cap V(C_j)| \geqslant 2$, where $i \neq j$. When the boundary of the exterior face is C_i, where $C_i \in \{C_1, C_2, \dots, C_l\}$, that is, the exterior face is a non-induced face, by definition 2.11, there exists one interior face which is an induced face. Hence, $|F(G)| > c_2(G)$. By Claim 2, we have

$$c_1(G) \leqslant |F(G)| - 1,$$

where $|F(G)|$ denotes the number of faces of G. When the boundary of the exterior face is constituted by the edges of at least two cycles in $C_1, C_2, \ldots, C_l$, that is, the exterior face is an induced face, by definition 2.11, we have $c_2(G) > |F(G)|$. By Claim 2, we have

$$c_1(G) \leqslant |F(G)| - 1.$$

By lemma 2.9, we have

$$c_1(G) \leqslant |E(G)| - |V(G)| + \omega(G).$$

In the following we want to prove that $c_1(G) = |F(G)| - 1$ if and only if G is the union of some generalized trees.

Suppose G is the union of some generalized trees, $c_1(G) = |F(G)| - 1$ holds obviously.

On the other hand, suppose $c_1(G) = |F(G)| - 1$ holds, we want to prove that G is the union of some generalized trees by mathematical induction.

(1) When $c_1(G) = 1$,we have $|F(G)| = 2$. Then, G is the union of one cycle and some isolated vertices. Therefore, G is the union of some generalized trees.

(2) Suppose the conclusion holds for any graph G'' with $c_1(G'') = k$ and $c_1(G'') = |F(G'')| - 1$. Let $c_1(G) = k + 1$ hold and the maximum cardinality cycle decomposition of G be $C_1, C_2, \ldots, C_k, C_{k+1}$. We have two cases to discuss.

Case 1: There exists a cycle C_i in $C_1, C_2, \ldots, C_{k+1}$ such that $|V(C_i) \cap V(C_j)| = 0$, where $j = 1, 2, \ldots k+1$ and $i \neq j$. Let $G^{(3)} = G - V(C_i)$. It is easy to see that the maximum cardinality cycle decomposition of $G^{(3)}$ is k. Since $c_1(G) = |F(G)| - 1$ we have

$$c_1(G^{(3)}) = |F(G^{(3)})| - 1.$$

Thus, $G^{(3)}$ is the union of some generalized trees, hence, G is the union of some generalized trees.

Case 2: For any C_i, there exists C_j such that $|V(C_i) \cap V(C_j)| \geqslant 1$, where $i \neq j$, $i = 1, 2, \ldots, k + 1$.

Claim 3: If for any C_i there exists C_j such that $|V(C_i) \cap V(C_j)| \geqslant 1$ holds, we have

$$|V(C_i) \cap V(C_j)| = 1.$$

Otherwise, for any C_i there exists C_j such that $|V(C_i) \cap V(C_j)| \geqslant 2$, we have $|F(G)| \geqslant c_1(G) + 2$, which contradicts with $|F(G)| = c_1(G) + 1$. Thus, for any C_i and C_j we have

$$|V(C_i) \cap V(C_j)| \leqslant 1.$$

We obtain a new graph $G^{(4)} = G - E(C_i)$, where C_i is a fixed cycle in $C_1, C_2, \ldots, C_{k+1}$. Obviously, $c_1(G^{(4)}) = k$ and $c_1(G^{(4)}) = |F(G^{(4)})| - 1$. By hypothesis in (2), we know that $G^{(4)}$ is the union of some generalized trees, hence, G is the union of some generalized trees. Hence, $c_1(G) = |F(G)| - 1$ if and only if G is the union of some generalized trees.

In the following, we use mathematical induction to prove that $\frac{1}{2}|F(G)| \leqslant c_1(G)$. By Claim 2, we want to prove that $\frac{1}{2}|F(G)| \leqslant c_2(G)$.

(3) When $c_2(G) = 1$, we have $|F(G)| = 2$, the conclusion follows.

(4) Suppose the conclusion holds for any plane even graph $G^{(5)}$ with $c_2(G^{(5)}) = m$, where $m \geqslant 1$. Let G be a plane even graph with $c_2(G) = m + 1$ and let the non-intersecting cycle decomposition of G be $C_1, C_2, \ldots, C_m, C_{m+1}$. Let $G^{(6)} = G - E(C_{m+1})$, by hypothesis in (4) we have $\frac{1}{2}|F(G^{(6)})| \leqslant c_2(G^{(6)})$, where $|F(G^{(6)})|$ is the number of faces in $G^{(6)}$. By definition 2.10, we have $c_2(G) = c_1(G^{(6)}) + 1$. When we add C_{m+1} to $G^{(6)}$ to form G, by definition 2.11, C_{m+1} constitutes a face of G and C_{m+1} may cut the embedding face of $G^{(6)}$ into at most two new faces. In fact, when we add C_{m+1} to $G^{(6)}$ to form G and $|F(G)| - |F(G^{(6)})| \geqslant 3$, we can decompose G again as we do in Claim 2 and obtain a new non-intersecting cycle decomposition of G and find a new cycle C_{m+1} such that $|F(G)| - |F(G^{(6)}| \leqslant 2$. Thus, we have

$$|F(G)| \leqslant |F(G^{(6)})| + 2.$$

Hence, $\frac{1}{2}|F(G)| \leqslant c_2(G)$.

Claim 4: When $c_1(G) = \frac{1}{2}|F(G)|$, G contains exactly one non-trivial component without cut vertex.

In fact, let the components of G be $G_1, G_2, \ldots, G_l, \ldots, G_k$, and $G_1, G_2, \ldots, G_l$ be the non-trivial components of G, where $k \geqslant l$, $l \geqslant 2$.

Obviously,

$$c_1(G) = c_1(G_1) + \cdots + c_1(G_l).$$

We have proved that

$$\frac{1}{2}|F(H)| \leqslant c_1(H),$$

where H is any plane even graph, especially, we have

$$c_1(G_i) \geqslant \frac{1}{2}|F(G_i)|,$$

where $|F(G_i)|$ denotes the face number of G_i, $i = 1, 2, \ldots, l$.

Hence,

$$c_1(G_1) + \cdots + c_1(G_l) \geqslant \frac{1}{2}(|F(G_1)| + \cdots + |F(G_l)|).$$

Because $G_1, \ldots, G_l$ share a common exterior face, we have

$$|F(G)| = |F(G_1)| + \cdots + |F(G_l)| - (l - 1).$$

Because $c_1(G) = \frac{1}{2}|F(G)|$ we have

$$c_1(G_1) + \cdots + c_1(G_l) = \frac{1}{2}[|F(G_1)| + \cdots + |F(G_l)| - (l - 1)].$$

Therefore, we obtain

$$\frac{1}{2}[|F(G_1)| + \cdots + |F(G_l)| - (l - 1)] \geqslant \frac{1}{2}(|F(G_1)| + \cdots + |F(G_l)|),$$

which is a contradiction.

In the following we want to prove that the unique non-trivial component contains no cut vertex. Without loss of generality, we suppose G has no isolated vertex. Otherwise, we consider the unique non-trivial component of G.

Suppose G has a cut vertex v. By the definition of cut vertex we have $\omega(G - v) > \omega(G)$. For each non-trivial component of $G - v$, we use one edge to connect two vertices whose degrees are odd. In this way, at last we obtain $G^{(7)}$. In the process we can assure $G^{(7)}$ is a plane graph. In fact, when we use one edge to connect two vertices whose degrees are odd, we can choose this adding edge as if it had been two original edges of G without v. Since G is a plane graph we have $G^{(7)}$ is a plane graph. Let the non-trivial component number of $G^{(7)}$ be r, where $r \geqslant 2$.

$|F(G^{(7)})|$ denotes the face number of $G^{(7)}$, $c_1(G^{(7)})$ denotes the maximum cardinality cycle decomposition of $G^{(7)}$. It is easy to see that $G^{(7)}$ is a plane even graph and

$$|F(G^{(7)})| = |F(G)| + (r - 1),$$

$$c_1(G^{(7)}) = c_1(G).$$

Since $\frac{1}{2}|F(G)| = c_1(G)$, we have

$$\frac{1}{2}[|F(G^{(7)})| - (r - 1)] = c_1(G^{(7)}).$$

However, we have prove that

$$c_1(H) \geqslant \frac{1}{2}|F(H)|,$$

where H is any plane even graph, especially, we have $c_1(G^{(7)}) \geqslant \frac{1}{2}|F(G^{(7)})|$, which is a contradiction.

From the above argument, Claim 1 follows.

Let n denote the number of vertices with degree 4 in $G(\pi)$. By lemma 2.8, we have $|E(G(\pi))| = m + 2n$. Since $|E(G(\pi))| = 2b(\pi)$, we have

$$b(\pi) = \frac{1}{2}(m + 2n).$$

Claim 5:

$$d(\pi) \geqslant b(\pi) - c_1(\pi).$$

In fact, if an alternating cycle decomposition is not a cycle decomposition, we decompose it further. Hence, $c(\pi) \leqslant c_1(\pi)$. By lemma 2.4, the claim follows.

By Claims 1 and 5, the theorem follows.

4. Concluding remarks

Lemma 2.12 is a very useful lower bound in [10]. However, when m is big enough, n and $\omega(G(\pi))$ are small enough, it is easy to see that theorem 3.1 is much better than lemma 2.12. For example, let

$$\pi = 0\ 10\ 9\ 7\ 8\ 6\ 5\ 3\ 4\ 1\ 2\ 11\ 13\ 12\ 14.$$

By definition 2.1, we have $m = 14$, $n = 1$, $\omega(G(\pi)) = 2$. By lemma 2.12, we have $d(\pi) \geqslant 4$. By theorem 3.1, we have $d(\pi) \geqslant 5$.

Theorem 3.1 can be used conveniently in practice. First, $\omega(G(\pi))$ always equals 1. If you wanted to find it exactly, you could find it by computers in polynomial time (see [14]). Second, m and n can be counted easily, let alone by computers. Let

$$\pi = \pi_0\pi_1\pi_2 \ldots \pi_k\pi_{k+1},$$

where $\pi_0 = 0$, $\pi_{k+1} = k+1$. By definition 2.1, the degrees of vertices π_0 and π_{k+1} are 2. For vertex π_i, if $|\pi_i - \pi_{i-1}| = 1$ or $|\pi_i - \pi_{i+1}| = 1$, by definition 2.1, the degrees of π is 2. For example, let

$$\pi = 042135,$$

where $\pi_2 = 2$. Since $|\pi_2 - \pi_3| = 1$ we have the degree of vertex 2 is 2. At last, it is easy to see that $n = k + 2 - m$.

By lemma 2.4, we obtain $d(\pi) \geqslant b(\pi) - c(\pi)$. From Caprara [11, page 94], we know that in practical cases this bound turns out to be very tight, and is frequently equal to the optimum. By Claim 1 of theorem 3.1, we have $\frac{1}{2}|F(G(\pi))| \leqslant c_1(G(\pi))$. Therefore, we obtain an upper bound of $d(\pi)$ which is frequently holds:

$$d(\pi) \leqslant \lfloor \frac{1}{2}[|V(G(\pi))| - \omega(G(\pi)) - 1] \rfloor,$$

where $d(\pi)$ is the reversal distance of π, $|V(G(\pi))|$ is the vertex number of $G(\pi)$, $\omega(G(\pi))$ is the number of components of $G(\pi)$.

Can we delete the condition “if $G(\pi)$ is a plane graph " in theorem 3.1? That is, we conjecture that $G(\pi)$ must be a plane graph.

Acknowledgments

The Project Supported by Zhejiang Provincial Natural Science Foundation of China (102055), National Nature Science Foundation of China (No. 10471131) and The Foundation of Zhejiang Universities′ Youth Teachers.

References

[1] J.D. Palmer and L.A. Herbon, Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence, J. Mol. Evol. 27 (1988) 87–97.

[2] S. Karlin, E.S. Mocarski and G.A. Schachtel, Molecular evolution of herpesviruses: genomic and protein sequence comparisons, J. Virol. 68 (1994) 1886–1902.

[3] D.J. McGeoch, Molecular evolution of large DNA viruses of eukaryotes, Sem. Virol. 3 (1992) 399–408.

[4] V. Bafna and P. Pevzner, Sorting by reversals: genome rearrangements in plant organelles and evolutionary history of x chromosome, Mol. Biol. Evol. 12 (1995) 239–246.

[5] V. Bafna and P. Pevzner, Sorting permutations by transpositions, in: *Proceedings of 6th ACM−SIAM Ann. Symp. on Discrete Algorithms*, 1995, pp. 614–623.

[6] V. Bafna and P. Pevzner, Genome rearrangements and sorting by reversals, SIAM J. Comupt. 25(2) (1996) 272–289.

[7] N. Franklin, Conservation of genome form but not sequence in the transcription antitermination determinants of bacteriophages λ, ϕ21, and p22, J. Mol. Evol. 181 (1985) 75–94.

[8] S. Hannenhalli and P. Pevzner, Transforming cabbage into turnip (polynomial algorithm for sorting signed permutation by reversals), in: *Proceedings of 27th ACM Symp. on Theory of Computing* (STOC’95) 1995, pp. 178–189.

[9] D.L. Hartl and E.W. Jones, *Genetics: analysis of Genes and Genomes,* 5th ed. (Jones and Bartlett Publishers Inc., One Exert Plaza, Boston, MA) pp. 364–367.

[10] J. Setubal and J. Meidanis, *Introduction to Computational Molecular Biology* (PWS Publishing Company, Boston, MA, 1997) pp. 237–238.

[11] A. Caprara, Sorting permutations by reversals and eulerian cycle decompositions, SIAM J. Discr. Math. 12 (1999) 91–110.

[12] J. Kececioglu and D. Sankoff, Exact and approximation algorithms for sorting by reversals, with application to genome rearrangement, Algorithmica 13 (1995) 180–210.

[13] J.A. Bondy and U.S.R. Murty, *Graph Theory with Applications* (Macmillan Press Ltd, London, 1976).
[14] C. H. Papadimitrious and S. Steiglitz, *Combinatorial Optimization* (Prentice-Hall Inc., Englewood Cliffs, NJ, 1982) pp. 214.