Genome Rearrangements and Sorting by Reversals

Vineet Bafna Pavel A. Pevzner

Computer Science Department
The Pennsylvania State University
University Park, PA 16802
email: bafna,pevzner@cse.psu.edu

Abstract

Sequence comparison in molecular biology is in the beginning of a major paradigm shift - a shift from **gene** *comparison based on local mutations to* **chromosome** *comparison based on global rearrangements. In the simplest form the problem of gene rearrangements corresponds to* **sorting by reversals**, *i.e. sorting of an array using reversals of arbitrary fragments. Kececioglu and Sankoff gave the first approximation algorithm for sorting by reversals with guaranteed error bound 2 and identified open problems related to chromosome rearrangements. One of these problems is Gollan's conjecture on the reversal diameter of the symmetric group. We prove this conjecture and further study the problem of expected reversal distance between two random permutations. We demonstrate that the expected reversal distance is very close to the reversal diameter thereby indicating that reversal distance provides a good separation between related and non-related sequences. The gene rearrangement problem forces us to consider reversals of* **signed permutations**, *as the genes in DNA are oriented. Our approximation algorithm for signed permutation provides a 'performance guarantee' of $\frac{3}{2}$. Finally, we devise an approximation algorithm for sorting by reversals with a performance ratio of $\frac{7}{4}$.*

1 Introduction

Genus *Lobelia* comprises over 350 species that range from small, slender herbs to woody, giant-rosette plants. Fig. 1 presents the order of genes in *Tobacco* and *Lobelia fervens* chloroplast genomes with a hypothetical sequence of rearrangement events (Knox et al.,[KDP93]) during evolution of *Lobelia fervens* from a tobacco-like ancestral genome.

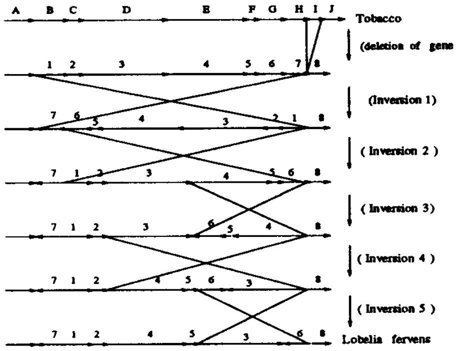

Figure 1: Evolution of *Lobelia fervens*

It is not so easy to verify that the evolutionary events presented in Fig. 1 represent the *shortest* series of reversal transforming the *Tobacco* permutation into the *Lobelia fervens* permutation. In fact, Theorem 2 of the paper indicates that the shortest sequence of rearrangement events contains just 4 reversals 71245368 → 71235468 → 71234568 → 76543218 → 12345678 (however, for the case of *signed* permutations (see below) the evolutionary events presented in Fig. 1 do present the shortest series of reversals).

With the advent of large-scale DNA mapping and sequencing, the number of biological problems similar to one presented in Fig. 1 is rapidly growing in different areas including evolution of chloroplast [RJ92] and mitochondrial genomes [S92], virology [K93] and *Drosophila* genetics [WPFJ89]. On the other hand, there are almost no computer science results allowing a biologist to analyse gene rearrangements. Recently, Kececioglu and Sankoff found an algorithm for reversal distance with guaranteed error bound 2 and raised a spectrum of open problem motivated by gene rearrangements [KS93]. The present paper solves some of them.

0272-5428/93 $03.00 © 1993 IEEE

In the problem we consider, the order of genes in two organisms is represented by permutations $\pi = (\pi_1\pi_2\dots\pi_n)$ and $\sigma = (\sigma_1\sigma_2\dots\sigma_n)$. A *reversal* ρ of an interval $[i,j]$ is the permutation

$$\rho = (1,2,\dots,i-1,\mathbf{j},\mathbf{j}-\mathbf{1},\dots,\mathbf{i}+\mathbf{1},\mathbf{i},j+1,\dots,n)$$

Clearly $\pi \cdot \rho$ has the effect of reversing genes $\pi_i, \pi_{i+1}, \dots \pi_j$. Given permutations π and σ, the *reversal distance problem* is to find a series of reversals $\rho_1, \rho_2, \dots, \rho_t$ such that $\pi \cdot \rho_1 \cdot \rho_2 \cdots \rho_t = \sigma$ and t is minimum. We call t the *reversal distance* between π and σ. Note that reversal distance between π and σ equals the reversal distance between the *identity* permutation $\sigma^{-1}\pi$ and $\imath$. *Sorting π by reversals* is the problem of finding reversal distance $d(\pi)$, between π and $\imath$.

Reversals generate the *symmetric* group S_n. Given an arbitrary permutation π from S_n we seek a shortest products of *generators* $\rho_1 \cdot \rho_2 \cdots \rho_t$ that equals π. Even and Goldreich [EG81] show that given a set of generators of a permutation group G and a permutation π, determining the shortest product of generators that equals π is NP-hard. Jerrum [J85] proves that the problem is PSPACE-complete, and remains so, when restricted to two generators. In our problem, the generator set is fixed. However, Kececioglu and Sankoff [KS93] conjecture that sorting by reversals is NP-complete.

A related problem is that of *sorting by prefix reversals* (also known as *pancake flipping problem*). [GP79, GT78, HS93, CB93] find bounds on the *prefix reversal diameter* of the symmetric group. Aigner and West [AW87] consider the diameter of sorting when the operation is reinsertion of the first element, and Amato et. al. [ABSR89], consider a variation inspired by reversing trains. Kececioglu and Sankoff [KS93] have found an approximation algorithm for sorting by reversals with performance guarantee 2. They also devise efficient bounds, allowing them to solve the reversal distance problem optimally or almost optimally for n ranging from 30 to 50. This range covers the biologically important case of mitochondrial genomes.

Define $d(n) = \max_{\pi \in S_n} d(\pi)$ to be the *reversal diameter* of the symmetric group of order n. For the problem of sorting by reversals, Gollan conjectured that $d(n) = n-1$ and that only one permutation γ_n, and its inverse, γ_n^{-1} require $n-1$ reversals to be sorted(see [KS93] for details). The *Gollan* permutation, in cycle notation, is defined as follows:

$$\gamma_n = \begin{cases} (1,3,5,7,\dots,n-1,n,\dots,8,6,4,2), & n \text{ even} \\ (1,3,5,7,\dots,n,n-1,\dots,8,6,4,2), & n \text{ odd} \end{cases}$$

For $n \leq 11$, Gollan verified this conjecture using extensive computations. Kececioglu and Sankoff [KS93] developed lower bounds for reversal distance allowing them to verify Gollan's conjecture for $n \leq 200$ for $n \bmod 3 = 1$. In the present paper we introduce the notion of ***breakpoint graph*** of a permutation and establish the links between reversal distance and ***maximum cycle decomposition*** of this graph. This construction allows us to prove Gollan's conjecture. Further, we study the problem of reversal distance between two random permutations. We demonstrate that reversal distance between two random permutations is very close to the reversal diameter, thereby indicating that reversal distance provides a good separation between related and non-related sequences in molecular evolution studies.

Afterwards, we study reversals of ***signed*** permutations. The Lobelia fervens permutation (Fig. 1) corresponds to the signed permutation $(-7,+1,+2,+4,+5,+3,-6,+8)$. In the biologically more relevant signed case, every reversal of fragment $[i,j]$ changes the signs of the elements within that fragment. For a signed permutation π, ***reversal distance*** is the minimum number of reversals required to transform π into $(+1,+2,\dots,+n)$. We devise an approximation algorithm for sorting signed permutations by reversals with guaranteed error bound $\frac{3}{2}$. Finally, we use signed permutations to get a performance guarantee of $\frac{7}{4}$ for (unsigned) sorting by reversals, thereby improving on the factor of 2 due to Kececioglu and Sankoff[KS93].

2 Breakpoint Graph and Reversal Distance

Let $i \sim j$ if $|i-j| = 1$. Extend a permutation $\pi = \pi_1\pi_2\dots\pi_n$ by adding $\pi_0 = 0$ and $\pi_{n+1} = n+1$. We call a pair of consecutive elements π_i and π_{i+1}, $0 \leq i \leq n$, of π an *adjacency* if $\pi_i \sim \pi_{i+1}$, and a *breakpoint* if $\pi_i \not\sim \pi_{i+1}$. Define an edge-coloured graph $G(\pi)$ with $n+2$ vertices $0,1,\dots,n,n+1$. We join vertices i and j by a ***black*** edge if (i,j) is a ***breakpoint*** of π. We join vertices i and j by a ***gray*** edge if $i \sim j$ and i,j are not consecutive in π. The graph $G(\gamma_6)$ corresponding to the Gollan permutation $\gamma_6 = 315264$ (in one line notation) is shown in Fig. 2.

A sequence of vertices, $x_1x_2\dots x_m = x_1$ is called a *cycle* in a graph $G(V,E)$ if $(x_i,x_{i+1})\epsilon E$ for $1 \leq i \leq m-1$. A cycle in an edge-coloured graph G is called *alternating* if the colours of every two consecutive edges of this cycle are distinct. In the following,

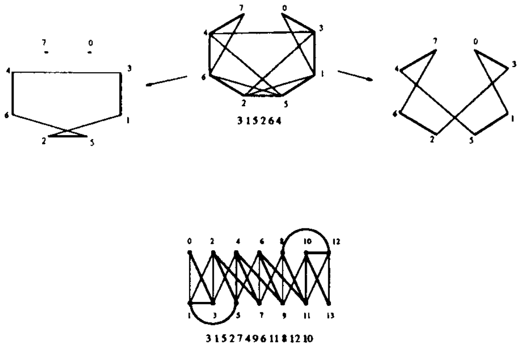

Figure 2: breakpoint graphs for $G(\gamma_6)$ and $G(\gamma_{12})$

we consider *cycle decompositions* of $G(\pi)$ into the maximum number $c(\pi)$ of edge-disjoint alternating cycles. A maximum cycle decomposition of $G(\gamma_6)$ into 2 cycles is shown in Fig. 2. (Black edges are shown by thick lines, while gray edges are shown by thin ones).

A vertex v is called *balanced* if the number of black edges incident to v equals the number of gray edges incident to v. A *balanced graph* is a graph in which every vertex is balanced. Note that $G(\pi)$ is a balanced graph for every π;therefore, it contains an alternating Eulerian cycle in every connected component, and $c(\pi) \geq 1$ for every π. For characterization of alternating Eulerian cycles in edge-colored graphs, see Kotzig [K68], and Pevzner [P93].

Cycle decompositions play an important role in estimating the reversal distance. When we apply a reversal to a permutation, there might be a change in the number of breakpoints, as well is in the number of cycles in a maximum decomposition. In theorem 1, we show that there is a strong correlation between these two changes. This idea allows us to bound the reversal distance in terms of the size of the maximum cycle decomposition.

Denote the number of black edges in $G(\pi)$ (breakpoints in π) as $b = b(\pi)$. Given an arbitrary reversal ρ denote $G' = G(\pi\rho)$, $b' = b(\pi\rho)$, the number of breakpoints in $\pi\rho$ and $c' = c(\pi\rho)$, the number of cycles in a maximum cycle decomposition of G'. Denote $\Delta b = \Delta b(\pi, \rho) = b - b'$ (decrease of breakpoints) and $\Delta c = \Delta c(\pi, \rho) = c' - c$ (increase of the number of cycles in a maximum decomposition).

Theorem 1 *For every permutation π and reversal ρ, $\Delta b(\pi, \rho) + \Delta c(\pi, \rho) \leq 1$.*

Proof (sketch): Every reversal removes/adds atmost 2 breakpoints. We will consider all 5 potential values of Δb in a case-by-case fashion. In the following we assume that ρ reverses a fragment of π starting from i and ending in j.

Case $\Delta b = 2$:

Every ρ with $\Delta b = 2$ reverses a permutation

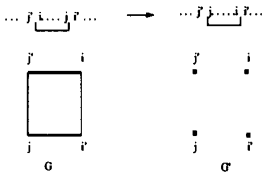

Figure 3: $\Delta b = 2$

$\pi = \ldots, j', i, \ldots, j, i', \ldots$ with $j' \sim j$ and $i' \sim i$. Breakpoints j', i and j, i' correspond to an alternating 4-cycle in π shown in Fig. 3 Reversal ρ simply removes this cycle from G. Therefore, every cycle decomposition of G' into c' cycles induces a cycle decomposition of G into $c' + 1$ cycle by adding the 4-cycle shown in Fig. 3. Therefore $c \geq c' + 1$ implying $\Delta c \leq -1$ and $\Delta b + \Delta c \leq 1$.

Case $\Delta b = 1$:

Every ρ with $\Delta b = 1$ either (i) creates one new adjacency, i.e. π has a generic form $\pi = \ldots, k, i, \ldots, j, i', \ldots$ with $i \sim i'$ or (ii) creates 2 new adjacencies and simultaneously destroys an adjacency, i.e. π has a generic form $\pi = \ldots, i-1, i, \ldots, j = i-2, i+1, \ldots.$

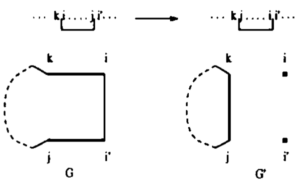

Figure 4: $\Delta b = 1(i)$

In the subcase (i) the graphs G and G' are shown in Fig. 4. Note that every alternating cycle in G' containing the edge (k, j) induces an alternating cycle in G containing the edges (k, i), (i, i') and (i', j'). Therefore, every cycle decomposition of G' into c' cycles induces a cycle decomposition of G into c' cycles.

In the subcase (ii) the graphs G and G' are shown in Fig. 5. Note that every alternating cycle in G' containing the edge $(i-1, i)$ induces an alternating cycle in G containing the edges $(i-1, i-2)$, $(i-2, i+1)$ and $(i+1, i)$. Therefore, every cycle

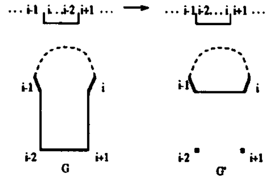

Figure 5: $\Delta b = 1(ii)$

decomposition of G' into c' cycles induces a cycle decomposition of G into c' cycles.

Therefore for both subcases (i) and (ii) $c \geq c'$ implying $\Delta c \leq 0$ and $\Delta b + \Delta c \leq 1$.

The other cases can be proved in a similar fashion. □ Theorem 1 immediately gives us a new lower bound for the reversal distance.

Theorem 2 *For every permutation π, $d(\pi) \geq b(\pi) - c(\pi)$.*

Proof : Let $\rho_t, \ldots, \rho_1$ be a shortest series of reversals transforming $\pi = \pi_t$ into the identity permutation π_0. Denote $\pi_{i-1} = \pi_i \rho_i$ for $i = 1, \ldots, t$, and apply theorem 1 for π_i and ρ_i.

$$\begin{aligned} d(\pi_i) &= d(\pi_{i-1}) + 1 \\ &\geq d(\pi_{i-1}) + \Delta b(\pi_i, \rho_i) + \Delta c(\pi_i, \rho_i) \\ &= d(\pi_{i-1}) + (b(\pi_i) - b(\pi_{i-1})) \\ &\quad +(c(\pi_{i-1}) - c(\pi_i)) \end{aligned}$$

Recalling that $d(\pi_0) = b(\pi_0) = c(\pi_0) = 0$, we get $d(\pi_i) - (b(\pi_i) - c(\pi_i)) \geq d(\pi_{i-1}) - (b(\pi_{i-1}) - c(\pi_{i-1})) \ldots \geq d(\pi_0) - (b(\pi_0) - c(\pi_0)) = 0$.
Substituting $i = t$, we prove the theorem. □

3 Reversal Diameter of the Symmetric Group

Now we have a characterization of the reversal distance of a permutation in terms of the maximum cycle decomposition of the breakpoint graph. Next we show that the graph corresponding to the Gollan permutation, γ_n, has atmost two disjoint alternating cycles.

Lemma 1 *Every alternating cycle in $G(\gamma_n)$ contains the vertex 1 or 3.*

Proof : Omitted.

Theorem 3 *(Gollan conjecture) $\forall n$, $d(\gamma_n) = d(\gamma_n^{-1}) = n - 1$.*

Proof : For $n \leq 2$, the claim is trivial. For $n > 2$, partition the vertex set of $G(\gamma_n)$ into $V_l = \{0, 1, 3\}$, and V_r. From lemma 1 and the fact that there is no cycle contained in V_l (see figure 2), we see that every alternating cycle must contain at least 2 edges from the cut (V_l, V_r). As the cut (V_l, V_r) contains 4 edges $((1,2), (1,5), (3,2), (3,4))$, the maximum number of edge disjoint alternating cycles in a cycle decomposition of $G(\gamma_n)$ is at most $\frac{4}{2} = 2$.

From theorem 2, $d(\gamma_n) \geq b(\gamma_n) - c(\gamma_n) \geq n+1-2 = n-1$. On the other hand, $d(\gamma_n) \leq n-1$ ([WEHM82]). Finally, note that $d(\gamma_n^{-1}) = d(\gamma_n)$. □

Before we prove that γ_n and γ_n^{-1} are the only permutations in S_n with a reversal distance of $n - 1$ (*strong Gollan conjecture*), we need to extend the concept of sorting permutations by reversals as follows: For any permutation of $\{1 \ldots n\}$, $\pi = \pi_1 \pi_2 \ldots \pi_n$, let $\hat{\pi} = \hat{\pi}_1 \hat{\pi}_2 \ldots \hat{\pi}_n$, with $\hat{\pi}_i = \pi_i + 1$, be a permutation of $\{2, 3, \ldots, n+1\}$. Define $d(\hat{\pi})$ as the minimum number of reversals required to transform $\hat{\pi}$ to $234 \ldots n+1$. Clearly, $d(\hat{\pi}) = d(\pi)$.

Theorem 4 *(strong Gollan conjecture) For every n, γ_n and γ_n^{-1} are the only permutations that require $n-1$ reversals to be sorted.*

Proof : Define $\mathcal{P}_n \equiv \{\pi | \pi \in S_n \text{ and } d(\pi) = n - 1\}$. We have seen that $\mathcal{P}_n \supseteq \{\gamma_n, \gamma_n^{-1}\}$. In what follows, we inductively prove that $\mathcal{P}_n \subseteq \{\gamma_n, \gamma_n^{-1}\}$.

For $n \leq 2$, the claim is trivial. Assume that the claim is true upto $n - 1$. Consider $\pi \in \mathcal{P}_n$. Let ρ_π be the reversal that brings n to the right end, that is, $\pi \cdot \rho_\pi = \pi' n$, where π' is a permutation of $\{1 \ldots n-1\}$.

It follows that, $d(\pi) \leq 1 + d(\pi')$. Then, $d(\pi') \geq d(\pi) - 1 = n - 2$. By induction, π' is either γ_{n-1} or γ_{n-1}^{-1}. Define, $\mathcal{A} = \{\pi | \pi \cdot \rho_\pi = \gamma_{n-1} n\}$, and $\mathcal{B} = \{\pi | \pi \cdot \rho_\pi = \gamma_{n-1}^{-1} n\}$, where γn denotes the concatenation of a permutation γ with element n. Obviously, $\mathcal{P}_n \subseteq \mathcal{A} \cup \mathcal{B}$.

Likewise, define ρ'_π as the reversal that brings 1 to the left end, that is, $\pi \rho'_\pi = 1 \pi'$, where π' is a permutation of $\{2, 3, \ldots, n+1\}$. As before, $d(\pi') \geq n - 2$, which implies that π is $\hat{\gamma}_{n-1}$ or $\hat{\gamma}_{n-1}^{-1}$. Define $\mathcal{C} = \{\pi | \pi \cdot \rho'_\pi = 1 \hat{\gamma}_{n-1}\}$, and $\mathcal{D} = \{\pi | \pi \cdot \rho'_\pi = 1 \hat{\gamma}_{n-1}^{-1}\}$. Then, $\mathcal{P}_n \subseteq \mathcal{C} \cup \mathcal{D}$, and therefore, $\mathcal{P}_n \subseteq \{\mathcal{A} \cup \mathcal{B}\} \cap \{\mathcal{C} \cup \mathcal{D}\}$. We state the following, without proof:

$$\mathcal{A} \cap \mathcal{C} = \phi, \quad \mathcal{A} \cap \mathcal{D} = \{\gamma_n\},$$
$$\mathcal{B} \cap \mathcal{C} = \{\gamma_n^{-1}\}, \quad \mathcal{B} \cap \mathcal{D} = \phi$$

It follows that $\mathcal{P}_n \subseteq \{\gamma_n, \gamma_n^{-1}\}$. □

4 Expected Reversal Distance

For any permutation $\pi \in S_n$, consider the set of cycles that form a maximum decomposition, and partition them according to size. Denote the number of alternating cycles of length i in the maximum decomposition by $c_i(\pi)$. Then, $c(\pi) = \sum_{i=4}^{2(n+1)} c_i(\pi)$.
For $k \leq 2(n+1)$, let us consider cycles in the decomposition whose size is atleast k. The number of such cycles is $c(\pi) - \sum_{i=4}^{k-1} c_i(\pi)$. Now, the breakpoint graph of π has exactly $2b(\pi)$ edges. From this, and the fact that the cycles are edge disjoint, we have $\forall k \leq 2(n+1)$,

$$\begin{aligned} c(\pi) &\leq \sum_{i=4}^{k-1} c_i(\pi) + \frac{1}{k}\left(2b(\pi) - \sum_{i=4}^{k-1} ic_i(\pi)\right) &(1)\\ &\leq \frac{1}{k}\left(2b(\pi) + \sum_{i=4}^{k-1}(k-i)c_i(\pi)\right) &(2)\\ \Rightarrow d(\pi) &\geq \left(1 - \frac{2}{k}\right) b(\pi) - \left(\sum_{i=4}^{k-1} c_i(\pi)\right) &(3) \end{aligned}$$

Lemma 2 provides an upper bound on the expected number of cycles of an arbitrary fixed length, in a random permutation of order n. Somewhat surprisingly, this bound is independent of n.

Lemma 2 $E(c_{2i}(\pi)) \leq \frac{2^{2i}}{2i}$

Proof : A cycle of length $2i$ is a set of i breakpoints (unordered pairs of vertices) of the form

$$\{(x'_i, x_1), (x'_1, x_2), (x'_2, x_3), \ldots, (x'_{i-1}, x_i)\},$$
$$\text{with } x_j \sim x'_j.$$

Consider the set, $x_1, x_2, \ldots, x_i$. First, we claim that in every maximum cycle decomposition, $x_1, x_2, \ldots, x_i$ are all distinct. To see this, consider the case $x_k = x_l$, for some $1 \leq k < l \leq i$. Then, $(x'_k, x_{k+1}), (x'_{k+1}, x_{k+2}), \ldots, (x'_{l-1}, x_l = x_k)$ form an alternating cycle, which can be detached to give a larger decomposition.

We have $\frac{n!}{(n-i)!}$ ways of selecting the ordered set, $x_1, x_2, \ldots, x_i$. Once this is fixed, we have a choice of atmost 2 elements for each of the x'_j, giving a bound of $2^i \frac{n!}{(n-i)!}$. Note that we count each $(2i)$-cycle $2i$ times, so a tighter bound is $\frac{2^i}{2i} \frac{n!}{(n-i)!}$.

Choose an arbitrary $(2i)$-cycle. The number of permutations in which this cycle can occur is no more than the number of ways of permuting the remaining $n - 2i$ elements plus the i pairs that form the cycle.

n	70	80	90	100
Theoretical	47.08	54.58	62.08	69.58
Experimental (matching)	48.73	55.80	63.33	70.49
Experimental (linear program)	51.7	58.9	67.6	74.2

Table 1: Comparison of theoretical and experimental lower bounds on expected diameter

Additionally, each pair can be flipped to give a different order, which gives atmost $2^i(n-i)!$ permutations. Therefore, the overall number of $(2i)$-cycles in all permutations is atmost $\frac{2^i}{2i}\frac{n!}{(n-i)!} \cdot 2^i(n-i)!$, and $E(c_{2i}(\pi)) \leq \frac{2^{2i}}{2i}$ □
Use lemma 2 and $E(b) = n + 1 - E(a) = n - 1$ to get a bound on the expected diameter:

Theorem 5 $E(d) \geq \left(1 - \frac{4}{\log n}\right) n$

Proof : Lemma 2 and inequality (3) imply for all $k \leq 2(n+1)$,

$$\begin{aligned} E(d) &\geq \left(1 - \frac{2}{k}\right) E(b) - \sum_{i=4}^{k-1} E(c_i) \\ &\geq \left(1 - \frac{2}{k}\right)(n-1) - \sum_{i=4}^{k-1} 2^i \\ &\geq n - \frac{2n}{k} - 2^k \end{aligned}$$

Choose $k = \log \frac{n}{\log n}$. Then $2^k \leq \frac{n}{k}$, and $E(d) \geq \left(1 - \frac{3}{\log \frac{n}{\log n}}\right) n$. This implies that $E(d) \geq \left(1 - \frac{4}{\log n}\right) n$ for $n \geq 2^{16}$. For $n \leq 2^{16}$ the claim is verified by a case by case analysis. □

Although the bound provided by Theorem 5 is good asymptotically, it is weak for small values of n. However, the bound given by inequality (2) is tight if we select a k that gives the minimum value. The derived bounds (Table 1) are comparable to the experimental bounds based on maximum matching and linear programming [KS93]. This indicates that, in the case of signed reversals (see below) where the cycle decomposition is unique, theorem 2 provides a computationally feasible way to prune the branch and bound tree used to solve the problem exactly.

5 Short Cycles and Approximating Reversal Distance

Starting from this section we discuss approximation algorithms for sorting by reversals. Define a ***strip*** of π as an interval $[i,j]$ such that $(i-1,i)$ and $(j,j+1)$ are breakpoints, and no breakpoint lies between them. A strip is ***increasing*** if $\pi_i < \pi_j$, otherwise it is ***decreasing***. A strip of 1 element is either increasing or decreasing, except for π_0 and π_{n+1}, which are always increasing.

A reversal can remove atmost 2 breakpoints; therefore $d(\pi) \geq \frac{b(\pi)}{2}$. Define an i-reversal, $i \in \{0,1,2\}$, as one that removes i breakpoints. For an upper bound on the number of reversals, [KS93] give a greedy procedure (Fig. 6) and prove the following lemma.

```
Procedure KS(π)
while π contains a breakpoint do
    ρ = Greedy(π)
    π = π · ρ
endwhile
Procedure Greedy(π)
begin
    Return a reversal that removes the most breakpoints
    of π, resolving ties in favor of reversals
    that leave a decreasing strip.
end
```

Figure 6: The greedy algorithm

Lemma 3 *If π is a permutation with a decreasing strip then, (i) π allows a 1− or 2-reversal. (ii)If every reversal that removes a breakpoint of π leaves a permutation with no decreasing strips, then π has a 2-reversal.*

Partition a sequence of reversals in *KS* into *rounds*, so that each round (except, perhaps, the first one) begins with a 0-reversal and has no other 0-reversals. Lemma 3 implies that each round ends in a 2-reversal, thereby proving that every 0-reversal can be amortized against a 2-reversal and, ***on the average***, we need atmost one reversal to remove a breakpoint. Comparison of the upper bound of $b(\pi)$ reversals against the lower bound $\frac{b(\pi)}{2}$ provides a performance guarantee of 2. Can we do better?

Theorem 2 gives a stronger lower bound, $d(\pi) \geq b(\pi) - c(\pi)$. Note that breakpoints correspond to black edges in a breakpoint graph and every cycle has atleast 2 black edges. Therefore, the lower bound of $\frac{b(\pi)}{2}$ is a simple corollary of theorem 2.

Let $c_4(\pi)$ be the number of 4-cycles in a maximum cycle decomposition of $G(\pi)$. For the lower bound we have

$$\begin{aligned} d(\pi) &\geq b(\pi) - c_4(\pi) - (c(\pi) - c_4(\pi)) \\ &\geq b(\pi) - c_4(\pi) - \frac{b(\pi) - 2c_4(\pi)}{3} \\ &= \frac{2}{3}b(\pi) - \frac{1}{3}c_4(\pi) \end{aligned}$$

In the following sections we devise algorithms that sort π in atmost $b(\pi) - \epsilon c_4(\pi)$ steps, for some $\epsilon > 0$. Then, the performance ratio of our algorithms is

$$\begin{aligned} \mathcal{A} &= \max_{0 \leq c_4(\pi) \leq \frac{b(\pi)}{2}} \left\{ \frac{b(\pi) - \epsilon c_4(\pi)}{\frac{2}{3}b(\pi) - \frac{1}{3}c_4(\pi)} \right\} \\ &= \begin{cases} 2 - \epsilon & \epsilon \leq \frac{1}{2} \\ \frac{3}{2} & \text{otherwise} \end{cases} \end{aligned} \qquad (4)$$

5.1 Approximation Algorithm for Signed Permutations

It is interesting to note that while the problem of sorting signed permutations is easier to handle, it is more relevant from a biological point of view. This is because genes are directed fragments of DNA sequences (Fig 1). Below, we devise an algorithm that sorts signed permutations in $b(\pi) - \frac{1}{2}c_4(\pi)$ steps, thereby achieving a ratio of $\frac{3}{2}$. Later, we will use signed permutations to improve the ratio for (unsigned) sorting by reversals.

We note that the concept of breakpoint graph as well as strips extends naturally to signed permutations. Define a transformation from a signed permutation π of order n to an unsigned permutation $\pi' \in S_{2n}$ as follows: replace $+i$ by $2i-1, 2i$ and $-i$ by $2i, 2i-1$. We observe that the identity signed permutation maps to the identity (unsigned) permutation, and the effect of a reversal on π can be mimicked by a reversal on π'. Therefore, any lower bound on π' is a lower bound on π. In particular, theorem 2 holds.

For the upper bound, we shall perform reversals only across breakpoints so that any reversal on the unsigned permutation can be mimicked by the signed permutation. It follows that for our purpose, the two permutations are equivalent and in the following discussion, whenever we refer to the breakpoint graph/strips of a signed permutation, it is implied that we refer to the breakpoint graph/strips of the transformed unsigned permutation.

Given a cycle decomposition of $G(\pi)$ and a reversal $\rho = [i,j]$, we call ρ a ***reversal on a cycle*** if the breakpoints (π_{i-1}, π_i) and (π_j, π_{j+1}) belong to the *same* cycle. A cycle is *oriented* if there exists a 1 or 2-reversal on it. Two cycles are *crossing* if some of the breakpoints corresponding to their black edges are *interleaved* in the permutation. For example, for γ_6, ρ : 315264 $\longrightarrow$ 315624 is a reversal on the cycle C = 134652 (Fig. 2), as the breakpoints $(5,2)$ and $(6,4)$ belong to C. C is oriented as ρ is a 1-reversal. Cycles C and C' in Fig. 2 are crossing as the breakpoints ($(3,1)$ and $(6,4)$ in C, $(1,5)$ and $(4,7)$ in C') are interleaved in the permutation 315264.

Observe that in the breakpoint graph of a signed permutation, every vertex has degree atmost 2. Therefore the cycle decomposition is unique, thus making the case of signed permutations easier. In order to sort a signed permutation π in less than $b(\pi)$ steps, we need 2-reversals that ***do not have to be amortized*** against 0-reversals. In the breakpoint graph of signed permutations, 2-reversals correspond to elimination of 4-cycles, while 1-reversals correspond to shortening of longer cycles. However, the breakpoints might not be oriented correctly so that 1− and 2-reversals are infeasible. Note that a reversal on a cycle can orient an unoriented crossing cycle. The following lemma shows that we can use reversals on a cycle to orient 4-cycles.

Lemma 4 *(proof omitted) Any* 4*-cycle that is not oriented has a crossing cycle* C*. Also, there exists a reversal on* C *which will orient the* 4*-cycle.*

Lemma 4 motivates the algorithm *SignedSort* for sorting signed permutations (Fig. 7).

```
Procedure SignedSort($\pi$)
 1. while $\pi$ contains a breakpoint do
 2.     if $\pi$ has no decreasing strips
 3.         if any 4-cycle $C$ remains in $G(\pi)$
 4.             Find a cycle $C'$ which crosses $C$.
 5.             Do a 0-reversal on $C'$ so that the
                4-cycle $C$ is oriented.
 6.             Do a 2-reversal on the 4-cycle $C$.
 7.         else
 8.             Do a 0-reversal on an arbitrary cycle.
 9.     else
10.         $\rho = Greedy(\pi)$ (see Fig. 6)
11.         $\pi = \pi \cdot \rho$
12. endwhile
```

Figure 7: Algorithm for signed permutations

Lemma 5 *(proof omitted) After step (6) in SignedSort, some decreasing strips remain.*

Lemma 6 *If there is a* 4*-cycle in* $G(\pi)$ *at the beginning of any round of SignedSort(except, perhaps, the first one), then there are atleast two* 2*-reversals in that round.*

Proof : If there is a 4-cycle in $G(\pi)$ at the beginning of any round, that round begins with a 0-reversal followed by a 2-reversal. Also, from lemma 5 some decreasing strips remain after this 2-reversal. On the other hand, if the graph has decreasing strips we call *Greedy*. Then, lemma 3 applies and every round of *SignedSort* also ends in a 2-reversal. □

Theorem 6 *The algorithm SignedSort sorts a signed permutation* π *in atmost* $b(\pi) - \frac{1}{2}c_4(\pi)$ *reversals and provides an approximation ratio of* $\frac{3}{2}$.

Proof : As each reversal is along a cycle, the number of cycles, and 4-cycles in particular, decreases only in 2-reversals. From lemma 6 we know that atmost half of the 2-reversals need to be amortized against 0-reversals. Therefore we can sort a signed permutation π in atmost $b(\pi) - \frac{1}{2}c_4(\pi)$ reversals. The bound on performance follows from equation (4). □

5.2 Approximation Algorithm for Sorting by Reversals

In general, finding a maximum cycle decomposition is not straightforward. In this section, we concentrate only on finding a cycle decomposition (not necessarily maximum) with a large number of 4-cycles, as such a decomposition will provide an improved performance ratio for sorting by reversals (eq. 4).

Consider the set of 4-cycles in a breakpoint graph $G(\pi)$. Any two 4-cycles can share atmost two edges. Two 4-cycles are 2-*overlapping* if they share two edges.

Lemma 7 *If two* 4*-cycles of* $G(\pi)$ *are* 2*-overlapping, then one of them is oriented (i.e. a* 2*-reversal is possible on it).*

Proof : Omitted.

Define the 4-*cycle graph* $H(\pi)$ of a permutation π, as one in which each node corresponds to a 4-cycle of $G(\pi)$, and two nodes are connected if the corresponding 4-cycles share an edge in $G(\pi)$. An *independent set* in $H(\pi)$ corresponds to a set of edge disjoint 4-cycles. Furthermore, this graph has bounded degree, and we can find a reasonable approximation to the maximum independent set problem in bounded degree graphs.We call a graph *strongly d-bounded* if degree of every vertex in the graph is bounded by d, and the degree of atleast one vertex in every connected component is less than d.

Lemma 8 *If* $G(\pi)$ *has no* 2*-overlapping cycles, then the* 4*-cycle graph* $H(\pi)$ *is strongly* 4*-bounded.*

Proof (Hint): A 4-cycle containing the maximal element of π among all 4-cycles in a given connected component of $H(\pi)$ overlaps with atmost 3 cycles. □

Lemma 9 *In a strongly d-bounded graph $G(V, E)$, an independent set of size atleast $\frac{|V|}{d}$ can be computed in $O(E)$ time.*

Proof : Omitted.

Lemma 10 *In a strongly d-bounded graph $G(V, E)$, a $\frac{2}{(d+1)}$ approximation to a maximum independent set can be computed in $O(E)$ time.*

Proof : Let the size of the maximum independent set be $c \cdot n$, for some c, where $|V| = n$. Consequently, the minimum vertex cover for G has a size $= (1-c) \cdot n$. We can find a vertex cover V' of size atmost $\min\{2(1-c) \cdot n, n\}$. Then, $I_1 = V \backslash V'$ is an independent set for G, of size atleast $\max\{(2c-1) \cdot n, 0\}$. Another independent set I_2 is given by Lemma 9 and is of size $\frac{n}{d}$. Obviously, we select the larger of the two sets. There are two cases in analyzing the performance of our approximation.
If $\max\{(2c-1)n, 0\} = 0$, then $c \leq \frac{1}{2}$. In this case we select I_2, and performance is $\frac{1}{cd} \geq \frac{2}{d}$.
Otherwise, we select the larger of the two sets, I_1 and I_2. The performance is $\max\left\{\frac{1}{dc}, \frac{(2c-1)}{c}\right\}$, with $c > \frac{1}{2}$. In the worst case, $\frac{1}{cd} = \frac{(2c-1)}{c}$, which implies that $c = \frac{(d+1)}{2d}$, and performance is $\frac{2}{(d+1)}$. □
Lemmas 7, 8 and 10 motivate the algorithm *ReversalSort* (Fig. 8).

Algorithm *ReversalSort(π)*

1. Starting with the permutation π, perform 2-reversals on $G(\pi)$ until no 2-overlapping cycles remain (lemma 7). Denote the resulting permutation by σ.
2. Use the Independent Set approximation in $H(\sigma)$ to find a set of 4-cycles of size atleast $\frac{2}{5}c_4(\sigma)$, in the breakpoint graph $G(\sigma)$ (lemmas 8, 10). Find an arbitrary cycle decomposition of the remaining edges.
3. Transform $G(\sigma)$ by splitting vertices of degree 4, so that the cycles are vertex disjoint. In terms of strips, replace single elements by strips, oriented appropriately, resulting in a signed permutation σ'.
4. Call *SignedSort(σ')* to sort σ'. Sorting of σ' mimics sorting of σ.

Figure 8: Algorithm for sorting by reversals

Theorem 7 *The algorithm ReversalSort achieves an approximation ratio of $\frac{9}{5}$.*

Proof : Every reversal on the signed permutation σ' in step 4 of *ReversalSort* can be simulated on the unsigned permutation σ. Therefore, an upper bound on the number of reversals in *SignedSort(σ')* is an upper bound on the number of reversals in step 4 of *ReversalSort*. If the number of 4-cycles found in the cycle decomposition in step 2 is $c_4'(\sigma)$, then Theorem 6 implies $d(\sigma) \leq b(\sigma) - \frac{1}{2}c_4'(\sigma)$ for the number of reversals in step 4 of *ReversalSort*. On the other hand $b(\pi) - b(\sigma) = 2x$, where x is the number of reversals in step 1. Therefore, *ReversalSort* requires no more than $b(\sigma) + x - \frac{1}{2}c_4'(\sigma) = b(\pi) - x - \frac{1}{2}c_4'(\sigma)$ reversals to sort π.

Let $i(\delta)$ be the size of the maximum independent set in a 4-cycle graph $H(\delta)$. Lemmas 8 and 10 guarantee that $c_4'(\sigma) \geq \frac{2}{5}i(\sigma)$. On the other hand, $i(\pi) - i(\sigma) \leq 4x$, as every reversal in step 1 'destroys' at most four of the $i(\pi)$ vertices of the maximum independent set in $H(\pi)$ (there are atmost 4 non-overlapping cycles sharing edges with a 4-cycle). Therefore,

$$\begin{aligned} d(\pi) &\leq b(\pi) - x - \frac{1}{5}i(\sigma) \\ &\leq b(\pi) - x - \frac{1}{5}(i(\pi) - 4x) \\ &\leq b(\pi) - \frac{1}{5}c_4(\pi) \end{aligned}$$

The bound on performance follows from (4). □

5.3 Improved Approximation for Sorting by Reversals

In this section we modify *ReversalSort* to improve the performance ratio. Recall that in step (1) of *ReversalSort*(π), we perform 2-reversals to transform π into a permutation σ, which has the property that $H(\sigma)$ is strongly 4-bounded. This allows us to find a set of 4-cycles of size atleast $\frac{2}{5}c_4(\sigma)$. In this section, we transform π using 2-reversals into a permutation σ, so that $H(\sigma)$ is bipartite. Consequently, we can find a maximum set of non-overlapping 4-cycles in σ, which leads to improved performance.

The problem of transforming π into ν is equivalent to sorting $\nu^{-1}\pi$, and, for convenience, we shall switch between the two notations. Denote $G(\pi, \nu) \equiv G(\nu^{-1}\pi)$, $b(\pi, \nu) \equiv b(\nu^{-1}\pi)$, $i(\pi, \nu) \equiv i(\nu^{-1}\pi)$. Observe that $G(\pi, \nu)$ and $G(\nu, \pi)$ coincide, but have reversed colors, i.e., black (gray) edges in $G(\pi, \nu)$ are gray (black) in $G(\nu, \pi)$. It follows that $b(\pi, \nu) = b(\nu, \pi)$ and $i(\pi, \nu) = i(\nu, \pi)$.

Fig. 9 describes the procedure *Transform*. Let $\rho_1\rho_2 \ldots \rho_x$ be the sequence of reversals in *Transform* that transforms π to $\bar{\pi}$ ($\pi\rho_1\rho_2 \ldots \rho_x = \bar{\pi}$), and

```
Procedure Transform($\pi$)
begin
    $\bar{\pi} = \pi$
    $\bar{\nu} = \imath$          /* $\imath$ is the identity permutation*/
    while $\bar{\nu}^{-1} \cdot \bar{\pi}$ or $\bar{\pi}^{-1} \cdot \bar{\nu}$ has a 2-reversal, $\rho$
        if $\rho$ is a 2-reversal on $\bar{\nu}^{-1} \cdot \bar{\pi}$
            $\bar{\pi} = \bar{\pi} \cdot \rho$
        else $\bar{\nu} = \bar{\nu} \cdot \rho$
    endwhile
    return $\bar{\nu}^{-1} \cdot \bar{\pi}$
end
```

Figure 9: Algorithm for preprocessing permutation π using 2-reversals

$\varrho_1 \varrho_2 \ldots \varrho_y$, the sequence of reversals that transforms $\imath$ to $\bar{\nu}$ ($\imath \varrho_1 \varrho_2 \ldots \varrho_y = \bar{\nu}$). The permutation $\sigma = \bar{\nu}^{-1} \cdot \bar{\rho}$ has the following properties:

Lemma 11
(i) $H(\sigma)$ is bipartite,
(ii) σ can be sorted in $b(\sigma) - \frac{1}{2}i(\sigma)$ reversals,
(iii) $b(\pi) - b(\sigma) = 2(x + y)$,
(iv) $i(\pi) - i(\sigma) \leq 4(x + y)$.

Proof : (i) Consider a 4-cycle in $G(\sigma)$, formed by the vertices $\pi_i, \pi_{i+1}, \pi_j, \pi_{j+1}$, with $i + 1 < j$. As no 4-cycle in $G(\sigma)$ is oriented, we have the relation $\pi_i \sim \pi_{j+1}$ and $\pi_j \sim \pi_{i+1}$. Now, observe that these vertices form a 4-cycle in $G(\sigma^{-1})$ also, with the color on the edges reversed. If $\pi_j - \pi_{i+1} = \pi_{j+1} - \pi_i$, then the 4-cycle in $G(\sigma^{-1})$ is oriented. Therefore, $\pi_j - \pi_{i+1} = -(\pi_{j+1} - \pi_i)$.

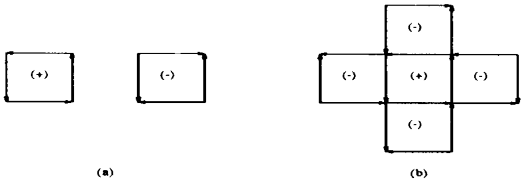

Figure 10: (+) and (-) cycles in $G(\pi)$

To interpret this graphically, direct black edges of $G(\sigma)$ from π_k to π_{k+1} and gray edges from k to $k + 1$. Note that the directions do not change in $G(\sigma^{-1})$. If a cycle is not oriented in $G(\sigma)$, then each gray edge must connect the tail of a black edge with the head of another. Likewise, if it is not oriented in $G(\sigma^{-1})$, then each black edge connects the tail of a gray edge with the head of another. Only two such 4-cycles (denoted (+) and (−)) are possible (Fig. 10a). Further, observe that all cycles that share an edge with a cycle of type (+), must be of type (−), and vice-versa (Fig. 10b). This implies that $H(\sigma)$ is bipartite.

(ii) As $H(\sigma)$ is a bipartite graph (lemma 11) with bounded degree, we can find a maximum independent set in $H(\sigma)$ in $O(n^{\frac{3}{2}})$ time. This implies that we can find a cycle decomposition of G(σ) in which the number of 4-cycles is $i(\sigma)$. Rest of the proof follows from theorem 6.

(iii,iv) In each iteration of the **while** loop in *Transform*, $(\bar{\pi}, \bar{\nu})$ is transformed into (π', ν'), in one of two ways. Either $\pi' = \bar{\pi}$, $\nu' = \bar{\nu} \cdot \rho$, or $\pi' = \bar{\pi} \cdot \rho$, $\nu' = \bar{\nu}$. In the first case, ρ is a 2-reversal on $\bar{\pi}^{-1}\bar{\nu}$ which implies that $2 = b(\bar{\pi}^{-1}\bar{\nu}) - b((\pi')^{-1}\nu') = b(\bar{\nu}, \bar{\pi}) - b(\nu', \pi') = b(\bar{\pi}, \bar{\nu}) - b(\pi', \nu')$. A similar argument holds for second case, implying that $b(\pi, \nu) - b(\pi', \nu') = 2$. $i(\bar{\pi}, \bar{\nu}) - i(\pi', \nu') \leq 4$ follows from the fact that every 2-reversal can destroy atmost four non overlapping 4-cycles in the breakpoint graph.

The proof follows from the fact that each reversal in *Transform* belongs either to the sequence $\rho_1 \rho_2 \ldots \rho_x$ or the sequence $\varrho_1 \varrho_2 \ldots \varrho_y$, implying a total of $x + y$ reversals. □

ImprovedSort (Fig. 11) exploits the structure of the permutation σ that *Transform*(π) returns, to sort π more efficiently. Theorem 8 analyzes the performance of this improved algorithm.

Algorithm *ImprovedSort(π)*

1. Call *Transform*(π) to find a sequence of reversals $\rho_1 \rho_2 \ldots \rho_x$ and $\varrho_1 \varrho_2 \ldots \varrho_y$ such that $\pi \rho_1 \rho_2 \ldots \rho_x = \bar{\pi}$ and $\imath \varrho_1 \varrho_2 \ldots \varrho_y = \bar{\nu}$. Denote $\sigma = \bar{\nu}^{-1}\bar{\pi}$.
2. Find a maximum set of non-overlapping 4-cycles in $G(\sigma)$ (see proof of lemma 11(ii)). Find an arbitrary cycle decomposition of the remaining edges.
3. Split vertices of degree 4 in $G(\sigma)$ according to the cycle decomposition found in step (2), so that the cycles are vertex disjoint. In terms of strips, replace single elements by strips, oriented appropriately, resulting in a signed permutation σ'.
4. Call *SignedSort(σ')* to find a sequence of reversals that sort σ'. Note that this sequence is mimicked by a sequence of reversals $\varphi_1 \varphi_2 \ldots \varphi_z$ that sorts σ.
5. Apply the sequence of reversals $\rho_1 \rho_2 \ldots \rho_x \varphi_1 \varphi_2 \ldots \varphi_z \varrho_y \ldots \varrho_2 \varrho_1$ to sort π.

Figure 11: Improved algorithm for sorting by reversals

Theorem 8 *ImprovedSort sorts π in $b(\pi) - \frac{1}{4}c_4(\pi)$*

steps, and therefore, achieves an approximation ratio of $\frac{7}{4}$.

Proof : From step (5) of *ImprovedSort*, $d(\pi) \leq x + y + z$. Lemma 11 provides an upper bound:

$$\begin{aligned} x+y+z &\leq x+y+b(\sigma)-\tfrac{1}{2}i(\sigma) \\ &= b(\pi)-(x+y)-\tfrac{1}{2}i(\sigma) \\ &\leq b(\pi)-(x+y) \\ &\quad -\tfrac{1}{2}\max\{0, i(\pi)-4(x+y)\} \\ &\leq b(\pi)-\tfrac{1}{4}i(\pi) \\ &\leq b(\pi)-\tfrac{1}{4}c_4(\pi) \end{aligned}$$

□

5.4 Running Time

We show that all our algorithms have a time complexity of $O(n^2)$. Consider lines $4-6$ of *SignedSort*. If we maintain both π and π^{-1}, it takes $O(n)$ time to find a 4-cycle, as well as a crossing cycle. Kececioglu and Sankoff give an $O(n)$ implementation of *Greedy*. Finally, no more than $n-1$ reversals are required to sort a permutation π, which gives an upper bound of $O(n^2)$ for *SignedSort*. The bounds for *ReversalSort* and *ImprovedSort* are the same as the preprocessing steps require atmost $O(n^{\frac{3}{2}})$ time (see proof of lemma 11(ii)).

Acknowledgements: We would like to thank John Kececioglu and David Sankoff for sending us the manuscript of their paper and for useful discussions. We are also grateful to Piotr Berman for useful comments on approximating maximum Independent Sets of bounded degree graphs.

References

[AW87] Aigner M, West D.B. (1987) Sorting by insertion of leading element. *Journal of Combinatorial Theory (Series A)*, **45**, 306-309.

[ABSR89] Amato, N., Blum, M., Irani, S. and Rubinfeld, R. (1989) Reversing trains: A turn of the century sorting problem. *Journal of Algorithms*, **10**, 413-428.

[CB93] Cohen, D., Blum, M., *Improved Bounds for Sorting Pancakes Under a Conjecture*, (submitted to Discrete Math.).

[EG81] Even, S. and Goldreich, O.,(1981), The minimum-length generator sequence problem is NP-hard, *Journal of Algorithms*, **2**, 311-313.

[GP79] Gates W.H., Papadimitriou C.H., (1979), Bounds for sorting by prefix reversals. *Discrete Mathematics*, **27**, 47-57.

[GT78] Gyori, E., Turan, E., (1978), Stack of Pancakes, *Studia-Sci.-Math. -Hungar.*, **13**, pp. 133-137.

[HS93] Heydari, M., Sudborough, I.H., *On Sorting by Prefix Reversals and the diameter of Pancake Networks* (manuscript).

[J85] Jerrum, M. R.,(1985), The complexity of finding minimum-length generator sequences. *Theoretical Computer Science* **36**, 265-289.

[KS93] Kececioglu J., Sankoff D. (1993), Exact and approximation algorithms for the reversal distance between two permutations. *Algorithmica* (to appear).

[KDP93] Knox E. B., Downie S.R., Palmer J.D. (1993), Chloroplast genome rearrangements and evolution of giant lobelias from herbaceous ancestors, *Mol. Biol. Evol.*, **10**, 414-430.

[K93] Koonin, E. V., Dolja, V. V., (1993), Evolution and Taxonomy of Positive-strand RNA viruses: Implications of comparative analysis of amino acid sequences, *Crit. Rev. Biochem. Molec. Biol.*(to appear)

[K68] Kotzig A., *Moves without forbidden transitions in a graph*, Matematicky casopis, **18**(1968), pp. 76-80.

[P93] Pevzner P., *DNA physical mapping and alternating Eulerian cycles in colored graphs*, Algorithmica, (to appear).

[RJ92] Raubenson L.A., Jansen R.K. (1992), Chloroplast DNA evidence on the ancient evolutionary split in vascular land plants. *Science*, **255**, 1697-1699.

[S92] Sankoff D., Leduc G., Antoine N., Paquin B., Lang B.F. and Cedergren, R.(1992), Gene order comparisons for phylogenetic inference: Evolution of the mitochondrial genome, *Proc. Natl. Acad. Sci. USA*, **89**, 6575-6579.

[WPFJ89] Whiting, J., Pliley,M., Farmer, J., Jeffery, D., (1989). In situ Hybridization Analysis of Chromosomal Homologies in Drosophila melanogaster and Drosophila virilis, *Genetics*, **122**, 99-109.

[WEHM82] Watterson, G.A., Ewens, W.J., Hall, T.E. and Morgan, A. (1982), The Chromosome Inversion Problem, *Journal of Theoretical Biology*, **99**, 1-7.