

Estendendo a Teoria de Bafna e Pevzner para o Problema da Ordenação por Transposições

M.E.M.T. WALTER¹, E.T.G. de OLIVEIRA², Departamento de Ciência da Computação, Instituto de Ciências Exatas, Universidade de Brasília, Campus Universitário Darcy Ribeiro, 70910-900 Brasília, DF, Brasil.

Resumo. Rearranjos de Genomas em Biologia Computacional é uma área na qual basicamente se visa resolver um problema combinatorial de encontrar uma série mínima de eventos mutacionais, afetando porções grandes do genoma, que transformam um genoma em outro. Neste artigo, estudamos o evento de transposição, que move um bloco de genes de um local para outro, dentro de um único cromossomo. Este evento leva ao problema da ordenação por transposições, que visa encontrar a menor seqüência de transposições que transforma um genoma em outro. Bafna e Pevzner [1] propuseram uma teoria baseada em propriedades de uma estrutura denominada grafo de ciclos, não inteiramente detalhada. Portanto, nossa contribuição é completar a teoria de Bafna e Pevzner [1]. Com base nestes resultados, apresentamos uma implementação e diversos experimentos.

1. Introdução

Rearranjos de Genomas, em Biologia Computacional, é uma área que basicamente visa computar uma distância entre genomas de dois organismos de espécies distintas, comparando blocos formados por um ou mais genes. De forma genérica, o *problema de distância de rearranjo* é encontrar a menor seqüência de eventos de rearranjo necessários para transformar um genoma em outro.

Neste trabalho, estudamos um evento particular ocorrendo em um único cromossomo, a *transposição*, que move genes de uma região para outra dentro do genoma (Figura 1). O *problema da distância de transposição* consiste em encontrar a menor seqüência de transposições que transforma um cromossomo em outro. Este artigo contribui para descobrir a complexidade deste problema, ainda em aberto.

Meidanis, Walter e Dias [5], e de forma independente Christie [2], calcularam a distância de transposição entre uma permutação e sua inversa. Sem restringir a forma das transposições, Bafna e Pevzner [1], Christie [2], Walter [8] e Walter, Dias e Meidanis [9] apresentaram algoritmos de aproximação para o problema. Vergara [7] e Guyer, Heath e Vergara [3] propuseram e implementaram diversos algoritmos

¹mia@cic.unb.br

²thereza@pgr.mpf.gov.br

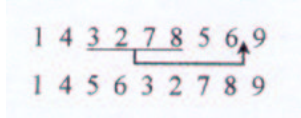


Figura 1: Exemplo do evento transposição, que move o bloco de genes indicado pelo traço para o local indicado pela seta.

heurísticos. Para o caso especial, em que os blocos da transposição não são adjacentes, citamos o trabalho de Irving e Christie [4], que apresentaram um algoritmo polinomial para o problema.

O objetivo deste trabalho é estender a teoria proposta por Bafna e Pevzner [1] para resolver o problema da ordenação por transposições. Com base nestes resultados, desenvolvemos uma implementação e realizamos diversos experimentos.

A teoria de Bafna e Pevzner [1] para o problema da ordenação por transposições é apresentada na Seção 2. O detalhamento desta teoria, além dos experimentos realizados, são mostrados na Seção 3. Finalmente, concluímos este trabalho e propomos novas direções de pesquisa na Seção 4.

2. O algoritmo de Bafna e Pevzner

Inicialmente modelaremos a seqüência de blocos de genes por uma permutação linear. Assim, a ordem dos blocos de genes é representada pela permutação $\pi = \pi_1 \pi_2 \pi_3 \dots \pi_n$, onde cada π_i é um inteiro sem sinais. Por exemplo, na Figura 1, a seqüência de blocos de genes 1 4 3 2 7 8 5 6 9 é representada por $\pi = (143278569)$.

A *transposição* “corta” certos blocos de genes do cromossomo e “cola” estes blocos de genes em um outro local no mesmo cromossomo. Formalmente, definimos uma transposição atuando em uma permutação π como se segue. Para toda permutação π , a transposição $\rho(i, j, k)$, $1 \leq j \leq n + 1$, $1 \leq k \leq n + 1$ tal que $k \notin [i, j]$, aplicada em π , “corta” os elementos π_i a π_{j-1} e os “cola” entre π_{k-1} e π_k :

$$\pi \cdot \rho(i, j, k) = (\pi_1 \dots \pi_{i-1} \pi_j \dots \pi_{k-1} \pi_i \dots \dots \dots \pi_{j-1} \pi_k \dots \pi_n).$$

Dadas as permutações π e σ , o *problema da distância de transposição* é encontrar uma série de transposições $\rho_1, \rho_2, \dots, \rho_t$ tal que $\pi \cdot \rho_t \cdot \dots \cdot \rho_2 \cdot \rho_1 = \sigma$, e t é mínimo. Chamamos t de *distância de transposição* entre π e σ , denotada por $d(\pi, \sigma)$. Note que a distância de transposição entre π e σ é igual a distância de transposição entre $\pi \cdot \sigma^{-1}$ e a permutação identidade $\iota = (1 \ 2 \ \dots \ n)$. Assim, o *problema da ordenação por transposições* é encontrar a distância de transposição, $d(\pi, \iota)$, entre π e ι , que será denotada apenas por $d(\pi)$.

Para resolver este problema, Bafna e Pevzner [1] propuseram uma estrutura denominada grafo de ciclos. Primeiramente estendemos a permutação inicial com mais dois elementos, $\pi_0 = 0$ e $\pi_{n+1} = (n + 1)$, e tomaremos como permutação inicial esta permutação estendida. Definimos o *grafo de ciclos* de π , denotado por $G(\pi)$, por um conjunto de vértices $\{+0, -1, +1, -2, +2, \dots, -n, +n, -(n + 1)\}$ e

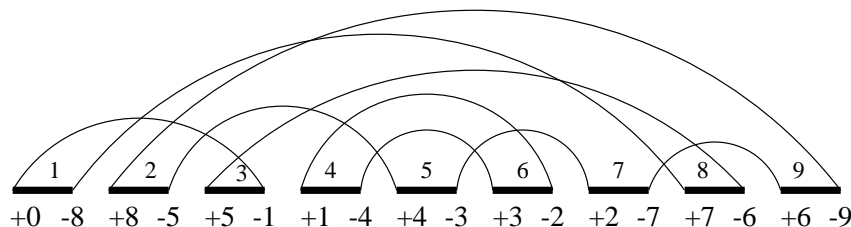


Figura 2: Exemplo de construção de um grafo de ciclos $G(\pi)$ para a permutação $\pi = (8\ 5\ 1\ 4\ 3\ 2\ 7\ 6)$.

um conjunto de arestas direcionadas bi-coloridas, cinzas e pretas. As arestas cinzas são direcionadas de $+0$ para -1 , de $+(i - 1)$ para $-i$, com $1 \leq i \leq n$, e de $+n$ para $-(n + 1)$. As arestas pretas são direcionadas de $-\pi_1$ para $+\pi_0$, de $-\pi_i$ para $+\pi_{i-1}$ com $1 \leq i \leq n$ e de $-\pi_{n+1}$ para $+\pi_n$ (Figura 2). Quando as arestas pretas tornam-se iguais às cinzas em todos os rótulos, temos $\pi = \iota$ e $d(\pi) = d(\iota) = 0$. Então o objetivo é aplicar transposições tais que as arestas pretas tornem-se iguais às arestas cinzas.

As definições e resultados apresentados nesta seção foram propostos por Bafna e Pevzner [1]. Um *ciclo alternado* de $G(\pi)$ é um ciclo direcionado no qual as arestas alternam suas cores. Observe que existe uma única aresta incidente e uma única aresta saindo de cada vértice, ambas com cores opostas, implicando numa única decomposição em ciclos para o conjunto de arestas de $G(\pi)$. Deste ponto em diante denominaremos ciclos alternados apenas por ciclos, e ciclo alternado de tamanhos $2k$ por k -ciclo. Um k -ciclo é *longo* se $k > 2$, e *curto*, caso contrário. Existe um total de $2(n + 1)$ arestas e no máximo $(n + 1)$ ciclos no grafo $G(\pi)$. A permutação identidade é a única com $n + 1$ ciclos. Denotamos o número de ciclos de $G(\pi)$ como $c(\pi)$. Logo, a seqüência de transposições que ordenam π devem aumentar o número de ciclos de $c(\pi)$ para $n + 1$. Para a permutação π e a transposição ρ , denotamos $\Delta(\rho) = c(\pi\rho) - c(\pi)$ como a variação no número de ciclos devido à transposição ρ . Um *ciclo é ímpar* em $G(\pi)$ se ele possui um número ímpar de arestas pretas. Definimos $c_{imp}(\pi)$ como sendo o número de ciclos ímpares em $G(\pi)$. Para a permutação π , e a transposição ρ , denotamos $\Delta c_{imp}(\rho) = c_{imp}(\pi\rho) - c_{imp}(\pi)$, como a variação no número de ciclos ímpares devido à transposição ρ . O lema a seguir estabelece a variação de $c_{imp}(\pi)$.

Lema 2.1. $\Delta c_{imp}(\rho) \in \{2, 0, -2\}$.

Como a permutação identidade tem $n + 1$ ciclos ímpares, Bafna e Pevzner [1] deduziram o limite abaixo a partir do Lema 2.1.

Teorema 2.1.

$$d(\pi) \geq \frac{n + 1 - c_{imp}(\pi)}{2},$$

Para $x \in \{2, 0, -2\}$, definimos um x -movimento em π como a transposição ρ tal que $\Delta c(\rho) = x$. Para ordenar o mais rapidamente possível seria interessante usar apenas 2-movimentos. Nesta seção apresentaremos os ciclos nos quais podem ser aplicados 2-movimentos. Numeramos as arestas pretas do grafo de ciclos $G(\pi)$ de 1 até $n+1$ associando um rótulo i para a aresta preta de π_i a π_{i-1} . Dizemos que a transposição $\rho(i, j, k)$ age nas arestas i, j e k . Dizemos também que a transposição $\rho(i, j, k)$ age no ciclo C se as arestas pretas i, j e k pertencerem a C . O Lema 2.1 pode ser usado para provar os seguintes resultados.

Lema 2.2. *Se a transposição ρ age em um ciclo em $G(\pi)$ e cria mais de um novo ciclo em $G(\pi\rho)$ então ρ é um 2-movimento.*

Lema 2.3. *Se a transposição ρ age em arestas pertencentes a exatamente dois ciclos diferentes então ρ é um 0-movimento.*

Podemos classificar ciclos como *não-orientados*, para os quais nenhum 2-movimento é possível, e *orientados*, para os quais um 2-movimento é possível. Mais precisamente, consideremos um k -ciclo C visitando, na ordem imposta pelo ciclo, as arestas pretas $i_1 \dots i_k$. Um ciclo C pode ser escrito de k formas possíveis dependendo da escolha da primeira aresta preta. Porém vamos fixar uma *representante canônica de um ciclo C* , tomando como aresta preta inicial i_1 a aresta mais à direita de C em π , isto é, $i_1 = \max_{1 \leq t \leq k} i_t$. No grafo de ciclos da Figura 2 temos três ciclos, com representantes canônicos $c_1 = [9, 7, 5, 2]$, $c_2 = [8, 1, 3]$ e $c_3 = [6, 4]$. Para todo $k > 1$, o ciclo $C = (i_1 \dots i_k)$ é *não-orientado* se $i_1 \dots i_k$ é uma seqüência decrescente, caso contrário C é um *ciclo orientado*.

Lema 2.4. *Se C é um ciclo orientado então existe um 2-movimento agindo em C . Se C é um ciclo não-orientado não existe nenhum 2-movimento agindo em C .*

Considere a tripla (x, y, z) de arestas pretas pertencentes a um mesmo ciclo C em $G(\pi)$. O ciclo C força uma ordem cíclica em x, y, z e temos três representações possíveis desta ordem. Escolheremos como *representante canônica de uma tripla (x, y, z)* aquela iniciando na aresta preta mais à direita, isto é, a aresta preta rotulada por $\max\{x, y, z\}$. Uma tripla na ordem canônica é *não-orientada* se $x > y > z$, e *orientada* se $y < z < x$. No grafo de ciclos da Figura 2 temos as seguintes triplas não-orientadas: $(9, 7, 5)$, $(9, 7, 2)$ e $(7, 5, 2)$; e a tripla orientada $(8, 1, 3)$. Duas seqüências orientadas de inteiros $V = (v_1, \dots, v_k)$, onde $v_1 < \dots < v_k$, e $W = (w_1, \dots, w_k)$, onde $w_1 < \dots < w_k$, são ditas *seqüências entrelaçadas* se $v_1 < w_1 < v_2 < w_2 < \dots < v_k < w_k$ ou $w_1 < v_1 < w_2 < v_2 < \dots < w_k < v_k$. Os conjuntos de inteiros V e W são *conjuntos entrelaçados* se possuírem seqüências entrelaçadas. Uma transposição $\rho(i, j, k)$ é uma *transposição embaralhada* com respeito à tripla (x, y, z) se os conjuntos (i, j, k) e (x, y, z) formarem seqüências entrelaçadas. Dois ciclos C e C' são *entrelaçados* se existirem uma tripla orientada em C e uma tripla não-orientada em C' que sejam entrelaçadas. Dizemos que uma transposição ρ é *válida* se $\Delta c(\pi) = \Delta c_{imp}(\rho)$.

Lema 2.5. *Se existe um ciclo orientado em $G(\pi)$, então temos um 2-movimento válido ou um 0-movimento válido seguido de dois 2-movimentos válidos consecutivos.*

Bafna e Pevzner [1], na prova do lema anterior, apresentam uma seqüência de transposições que garantem a criação de quatro ciclos ímpares em três movimentos consecutivos.

Inicialmente, eles observaram que existem ciclos orientados que não admitem 2-movimentos. Assim, caracterizaram *ciclos fortemente orientados*, e provaram que estes ciclos permitem 2-movimentos válidos. Caracterizaram ainda ciclos não-orientados da seguinte forma. Seja $C = (i_1, \dots, i_k)$ um ciclo em $G(\pi)$ e $C^* = (j_1 = i_1 > \dots > j_k)$ a seqüência das arestas pretas de C em ordem decrescente. As seqüências C e C^* coincidem para um ciclo não-orientado e são diferentes caso contrário.

Os dois lemas seguintes mostram como os ciclos fortemente orientados surgem de ciclos não-orientados.

Lema 2.6. *Se ρ é uma transposição embaralhada com relação a um ciclo não-orientado C , então ρ transforma C em um ciclo fortemente orientado em $G(\pi\rho)$.*

Lema 2.7. *Sejam $D(x, \dots, y)$ e $E(x', \dots, y')$ dois ciclos não-orientados em $G(\pi)$ sem nenhuma tripla entrelaçada e seja ρ a transposição agindo em três das quatro arestas pretas x, y, x', y' . Então ρ cria um ciclo fortemente orientado se, e somente se, D e E tiverem pares de arestas entrelaçadas.*

Cada ciclo fortemente orientado tem exatamente duas arestas direcionadas à direita, denotadas por (r, i_1) e (s, t) . Bafna e Pevzner [1] definiram dois tipos para os ciclos fortemente orientados. Para um ciclo C do primeiro tipo tomemos:

$$r' = \max_{i \in \text{esquerda}} i \quad \text{e} \quad t' = \min_{i \in \text{direita}} i$$

e consideremos os intervalos $I_1(C) = [r', r]$, $I_2(C) = [t, t']$ e $I_3 = [0, s]U[i_1, n + 1]$.

Para um ciclo C fortemente orientados de segundo tipo tomemos:

$$s' = \max_{i \in \text{esquerda}} i \quad t' = \min_{i \in \text{direita}} i \quad a = \max_{i \in \text{centro}'} i \quad \text{e} \quad a' = \min_{i \in \text{centro}''} i$$

e consideremos os três intervalos $I_1(C) = [s', s]$, $I_2(C) = [t, t']$ e $I_3 = [a, a']$.

Um ciclo fortemente orientado C e um ciclo não-orientado $C' = (i_1, \dots, i_k)$ são *fortemente cruzados* se existir uma aresta preta x em C' tal que cada um dos conjuntos $I_1(C)$, $I_2(C)$ e $I_3(C)$ contém exatamente um elemento da tripla (i_1, x, i_k) . Note que o 2-movimento para C descrito na prova do Lema 2.5 forma uma transposição embaralhada, sem perda de generalidade, (i_1, x, i_k) com relação a C' . Esta observação e o fato de ciclos fortemente orientados admitirem 2-movimentos válidos permitiram provar o lema abaixo.

Lema 2.8. *Se $G(\pi)$ tem ciclos fortemente cruzados então existem dois 2-movimentos válidos consecutivos em $G(\pi)$.*

Finalmente, são usados os Lemas 2.5 até 2.8, além de outros resultados, para provar o Teorema 2.2.

Teorema 2.2. *Se existe um ciclo longo em $G(\pi)$, então um 2-movimento válido ou um 0-movimento válido seguido de dois 2-movimentos válidos consecutivos são possíveis em π .*

Bafna e Pevzner [1] afirmam que o Teorema 2.2 tem uma prova construtiva que demonstra a correção do algoritmo proposto por eles e fornece uma complexidade de $O(n^2)$ para o problema da ordenação por transposições. De maneira análoga, os Teoremas 2.1 e 2.2 permitem estabelecer o seguinte corolário.

Corolário 2.1. *O Algoritmo TransSort ordena π em no máximo $3/4 \cdot (n + 1 - c_{imp}(\pi))$ transposições, garantindo uma razão de 1.5.*

3. Detalhamento teórico

Os resultados seguintes estendem a teoria de Bafna e Pevzner [1]. Deve-se notar que Guyer, Heath e Vergara [3] e Christie [2] comentaram que este algoritmo não teria sido implementado por falta de detalhes técnicos.

Como dito na seção anterior, os ciclos fortemente orientados possuem dois tipos, Tipo 1 e Tipo 2. Para verificar se um ciclo C fortemente orientado é fortemente cruzado com um ciclo C' não-orientado, primeiramente precisamos descobrir sua classificação entre Tipo 1 ou Tipo 2 pois os intervalos $I_1(C)$, $I_2(C)$ e $I_3(C)$ são diferenciados para cada tipo de ciclo fortemente orientado. Nestes intervalos verificamos se existe uma tripla não-orientada de um ciclo C' não-orientado, para que posteriormente possamos classificar C como fortemente cruzado com C' .

O primeiro passo é procurar as arestas i_1 , s , r e t do ciclo fortemente orientado em $G(\pi)$. Observamos que para os ciclos fortemente orientados do Tipo 1, a posição da aresta r é maior que a posição da aresta s , sendo o contrário para o Tipo 2. Desta forma, a relação entre as arestas r e s permite classificar os ciclos como Tipo 1 ou Tipo 2. Para um ciclo C do Tipo 1, procuramos as demais arestas delimitadoras dos intervalos I_1, I_2 e I_3 , r' e t' , onde $I_1(C) = [r', r]$, $I_2(C) = [t, t']$ e $I_3(C) = [0, s] \cup [i_1, n + 1]$, e verificamos se existe uma tripla não-orientada, de um ciclo C' não-orientado, tal que cada elemento da tripla esteja contido em um dos intervalos encontrados. Se o ciclo C for fortemente cruzado com um ciclo C' não-orientado, o próximo passo é verificar se as arestas r' e t são consecutivas ou não. Para um ciclo C do Tipo 2, procuramos as demais arestas delimitadoras dos intervalos I_1, I_2 e I_3 , s' , a , a' e t' , onde $I_1(C) = [s', s]$, $I_2(C) = [t, t']$ e $I_3(C) = [a, a']$, e verificamos se existe uma tripla não-orientada, de um ciclo C' não-orientado, tal que cada elemento da tripla esteja contido em um dos intervalos encontrados.

Para os ciclos do Tipo 1, com arestas não consecutivas, e do Tipo 2, Bafna e Pevzner [1] indicam duas transposições possíveis para obter 2-movimentos válidos. Estudamos a paridade das distâncias entre as arestas destes intervalos para decidir qual das duas transposições indicadas geravam 2-movimentos válidos. Em seguida,

apresentamos estes resultados, observando-se que ambas as provas são construtivas (Oliveira [6]).

Lema 3.1. *Dado um ciclo C fortemente orientado do Tipo 1, com arestas não consecutivas, e fortemente cruzado com um ciclo C' não-orientado, existem dois 2-movimentos válidos.*

Lema 3.2. *Dado um ciclo C fortemente orientado do Tipo 2 e fortemente cruzado com um ciclo C' não-orientado, existem dois 2-movimentos válidos.*

Bafna e Pevzner [1] provaram que todos os ciclos fortemente orientados têm uma tripla orientada (x, y, z) e que uma transposição agindo nas arestas y, z e x transforma C em três ciclos contendo trechos de tamanhos $d(x, y)$, $d(y, z)$ e $d(z, x)$, respectivamente. Provaram também que se $d(y, z)$ e $d(z, x)$ fossem ambas pares, os trechos de y a z e de z a x contêm no mínimo duas arestas.

Com base nestas provas, para o caso do ciclo C fortemente orientado mas não fortemente cruzado com nenhum outro ciclo C' não-orientado, o algoritmo *TransSort* procura uma tripla orientada (x, y, z) , tal que $d(x, y)$ é ímpar. Se pelo menos uma das distâncias $d(y, z)$ ou $d(z, x)$ for ímpar aplicamos a transposição $\rho(y, z, x)$ e garantimos um 2-movimento válido. Porém, se ambas as distâncias $d(y, z)$ e $d(z, x)$ forem pares, procuramos duas arestas b e a para aplicarmos a transposição $\rho(b, a, z)$ e garantirmos um 2-movimento válido. Observamos que a prova do lema seguinte é construtiva, conforme detalhamento feito em Oliveira [6].

Lema 3.3. *Dado um ciclo C fortemente orientado, mas não fortemente cruzado com nenhum outro ciclo C' não-orientado, pode-se encontrar a tripla orientada (x, y, z) tal que, $d(x, y)$ é ímpar e $d(y, z)$ ou $d(z, x)$ é ímpar, ou podem-se encontrar arestas pretas rotuladas por a e b tais que o ciclo $C = (x \dots y \dots azb \dots)$, $z > a$, $d(a, z) = 1$ e $d(z, b) = 1$. Em ambas as condições obtemos 2-movimentos válidos.*

Para encontrar um 2-movimento válido, se C possui uma tripla orientada (i, j, k) então a transposição $\rho(j, k, i) = \rho(k, i, j)$ gera 2-movimento. Para garantir o 2-movimento válido, a tripla orientada (i, j, k) , $i < j < k$, $k \notin [i, j]$, deve ser escolhida de tal forma que $d(i, j)$ é ímpar, $d(k, i)$ ou $d(j, k)$ é ímpar. Se o ciclo não permitir 2-movimento válido, isto é, não possuir tripla orientada com distâncias ímpares como descrito anteriormente, deve-se verificar se o ciclo possui uma aresta z , que gera um 0-movimento válido seguido de dois 2-movimentos válidos (Teorema 2.2). Para isto deve-se verificar se no ciclo orientado existe uma aresta z tal que a aresta saindo de z é b , a aresta chegando em z é a , e $a < b < z$. Se existirem as arestas a, b , e z nestas condições, o próximo passo é encontrar duas arestas x e y tais que $x > z$, $b < y < z$ e $d(x, y)$ seja ímpar. Se forem encontradas as arestas a, b, z, x, y , nas condições acima, é necessário verificar a ordem das arestas no ciclo, que deve ser a, z, b, x , e y . Estes passos foram descritos na prova do Lema 3.4, que é longa e foi descrita detalhadamente em Oliveira [6].

Lema 3.4. *Dado um ciclo orientado C com formato $C = (\dots azb \dots x \dots y \dots)$ podemos aplicar 0 – 2 – 2-movimentos válidos.*

Se não existirem triplas embaralhadas em $G(\pi)$, é possível encontrar dois ciclos não-orientados C e D com pares entrelaçados (Lema 2.7).

Bafna e Pevzner [1] mostram os $0-2-2$ -movimentos válidos a partir de ciclos com pares entrelaçados, mas nas suas provas são descritas apenas as transposições para o caso em que $y > i_k$, tendo indicado que a rotina para $y < i_k$ é análoga. Para a implementação fez-se necessário estudar detalhadamente os casos onde $y < i_k$.

O lema abaixo detalha o caso de dois ciclos C e D não-orientados, que não têm triplas entrelaçadas, estendendo o Lema 2.7 (Oliveira[6]).

Lema 3.5. *Sejam um ciclo longo não-orientado $C = (i_1 \dots i_k)$, não coberto por nenhum outro ciclo longo, e um ciclo $D = (x \dots c \dots d \dots y)$ não-orientado, tais que (c, d) e (i_1, i_k) se entrelacem, e suponhamos $y < i_k$. Então, existem $0-2-2$ -movimentos válidos.*

Os resultados anteriores estendem a teoria de Bafna e Pevzner [1] permitindo a implementação do algoritmo *TransSort*, conforme descrito na Figura 3.

Algoritmo *TransSort*

entrada: π
saída: $\rho_1, \rho_2, \dots, \rho_u, u$, tal que $\rho_u \dots \rho_2 \cdot \rho_1 \cdot \pi = \iota$

1. **enquanto** a permutação π for diferente da permutação identidade ι , **faça**
 - se** existirem ciclos longos em $G(\pi)$
 - C recebe o ciclo longo
 - 2. **se** o ciclo C for orientado
 - 3. **se** o ciclo C for fortemente orientado (Lema 3.3)
 - 4. Classifique C como Tipo 1 ou Tipo 2 (Lema 3.3)
 - 5. **se** existir um ciclo C' , não-orientado, fortemente cruzado com C
 - 6. Verifique a paridade entre os intervalos do ciclo C (lemas 3.1 e 3.2)
 - 7. **faça** dois 2-movimentos válidos (lemas 2.8, 3.1 e 3.2)
 - senão** $\{C$ apenas orientado, mas não fortemente orientado $\}$
 - 8. Verifique a paridade das distâncias dos intervalos do ciclo C (Lema 3.3)
 - 9. **faça** um 2-movimento válido (Lema 3.3)
 - senão**
 - 10. **se** encontramos em C uma tripla orientada (i, j, k) tal que duas das três distâncias, $d(k, i)$, $d(j, k)$ ou $d(i, j)$, são ímpares
 - 11. **faça** um 2-movimento válido (Lema 2.5)
 - senão** C pode ser particionado em cinco intervalos ou mais
 - 12. Encontre z no ciclo $C = (\dots azb \dots x \dots y \dots)$ (Lema 3.4)
 - 13. **faça** um 0-movimento válido seguido de dois 2-movimentos válidos (Lema 3.4)
 - senão** $\{C$ não-orientado ($C = C^*$) $\}$
 - 14. **se** existir transposição embaralhada (Lema 2.6)
 - 15. **faça** um 0-movimento seguido de dois 2-movimentos válidos (lemas 2.6 e 2.8)
 - senão**
 - 16. Encontre pares entrelaçados (Bafna e Pevzner [1])
 - 17. **faça** um 0-movimento seguido de dois 2-movimentos válidos (Bafna e Pevzner [1] e Lema 3.5)
 - senão** $\{$ existem apenas ciclos curtos em $G(\pi)$ $\}$
 - 18. Encontre dois ciclos entrelaçados (Bafna e Pevzner [1])
 - 19. **faça** dois 2-movimentos válidos (Bafna e Pevzner [1])

Figura 3: Detalhamento do algoritmo *TransSort*.

Executamos este programa para todas as permutações de tamanho n , com n variando de 2 a 9. Os resultados foram comparados com os valores exatos da distância de transposição. Na tabela abaixo, *Diferenças* indica em quantas permutações o resultado exato diferiu do encontrado pelo algoritmo, observando-se que o número de permutações onde há diferenças é bastante pequeno e ocorre a partir de $n = 7$, isto é, *Diferenças* permanece 0 para $n = 2$ a 6, ou seja, para $2 \leq n \leq 6$, os algoritmos de aproximação e exato apresentaram exatamente os mesmos resultados.

| n | Diferenças | Percentual das diferenças |
|---|------------|---------------------------|
| 7 | 1 | 0.02 |
| 8 | 135 | 0.33 |
| 9 | 4361 | 1.20 |

Tabela 1: Resultados obtidos da implementação.

4. Conclusões e Trabalhos Futuros

Uma investigação poderia ser feita na forma de aplicar transposições 0-movimento nas transposições embaralhadas e nos pares entrelaçados. A questão aqui é se as transposições propostas por Bafna e Pevzner [1] poderiam ser melhoradas de forma que a partir do primeiro 0-movimento não tivéssemos mais outros 0-movimentos. Os resultados experimentais devem ser estendidos para permutações maiores que formalizem porções de DNA obtidos em laboratório. Outro estudo importante é comparar os resultados obtidos pelo algoritmo 1.5 de Bafna e Pevzner [1] com outros algoritmos encontrados na literatura, inicialmente, o algoritmo 1.5 de Christie [2], o 2.25 de Walter, Dias e Meidanis [9] e os algoritmos heurísticos propostos por Guyer, Heath e Vergara [3]. Isto possibilitaria descobrir quais são as permutações que levam a diferenças entre os resultados, para tentar descobrir propriedades interessantes. Melhorar o limite inferior poderia também indicar novos parâmetros que auxiliassem na descoberta da complexidade do problema da ordenação por transposições.

Abstract. In the field of genome rearrangements in computational biology, we study a combinatorial problem of finding a minimal series of mutational events, affecting large portions of a genome, transforming this genome into one another. In this work, we study an specific event, transposition, that moves a block of genes from one place of a cromossome to another one. This event leads to the problem of transposition sorting, which is to find a minimal sequence of transposition events transforming a genome into another. Bafna and Pevzner [1] proposed a theory, based on properties of an structure named cycle graph, not entirely detailed. Our main contribution is to complete the theory of Bafna and Pevzner [1]. Based on these results, we present an implementation and many experiments.

Referências

- [1] V. Bafna and P. Pevzner, Sorting by transpositions, *SIAM Journal on Discrete Mathematics*, **11**, No. 2, (1998), 224-240.
- [2] D.A. Christie, “Genome Rearrangements Problems”, PhD thesis, Glasgow University, Scotland, 1998.
- [3] S.A. Guyer, L.S. Heath and J.P.C. Vergara, Subsequences and run heuristics for sorting by transpositions, in “4th Dimacs International Algorithm Implementation Challenge”, 1995.

- [4] R.W. Irving and D.A. Christie, Sorting permutations by block interchanges, *Information Processing Letters*, **60** (1996), 165-169.
- [5] J. Meidanis, M.E.M.T. Walter and Z. Dias, Transposition distance between a permutation and its reverse, in “Anais do IV South American workshop on string processing - WSP’97” (R. Balza Yates, ed.), pp. 70-79, Carleton University Press, Valparaiso, Chile, 1997.
- [6] E.T.G. Oliveira, “Implementação de Algoritmos para o Problema da Ordenação por Transposições”, Master’s thesis, Departamento de Ciência da Computação, Universidade de Brasília, 2001.
- [7] J.P. Vergara, “Sorting by Bounded Permutations”, PhD thesis, Virginia Polytechnic Institute and State University, 1997.
- [8] M.E.M.T. Walter, “Algoritmos para Problemas em Rearranjos de Genomas”, PhD thesis, Universidade de Campinas, 1999.
- [9] M.E.M.T. Walter, Z. Dias and J. Meidanis, A new approach for approximating the transposition distance, in “Proc. String Processing and Information Retrieval - SPIRE’2000”, pp. 199-208, 2000.