

# MO640 - Biologia Computacional

## Rearranjo de Genomas

André Atanasio Maranhão Almeida

Ata de aula: 06/12/2004

### Resumo

Nesta aula foi discutida a Seção 3, *Selected Results from the Hannenhalli-Pevzner Theory*, do trabalho desenvolvido pela Profa. Anne Bergeron [1], *A Very Elementary Presentation of the Hannenhalli-Pevzner Theory*. Bergeron apresenta uma série de simplificações no trabalho apresentado por Hannenhalli e Pevzner em 1995 [2], o primeiro algoritmo polinomial para o problema da distância de reversão orientada.

Dada um genoma  $\pi$  de tamanho  $n$  e emoldurado por 0 e  $n + 1$ , o diagrama de pontos de quebra é construído da seguinte forma:

1. Gera-se a permutação  $\pi'$  substituindo-se cada elemento positivo  $x$  em  $\pi$  por  $2x - 1$  em  $\pi'$  e cada elemento negativo  $-x$  em  $\pi$  por  $2x$  em  $\pi'$ . O elemento 0 de  $\pi$  mantém-se em  $\pi'$  e o elemento  $n + 1$  em  $\pi$  é substituído por  $2n + 1$  em  $\pi'$ .
2. Toma-se os elementos da permutação  $\pi'$  como vértices. Totalizando  $2n + 2$  vértices.
3. As arestas são definidas e classificadas da seguinte forma:

Arestas retas: elementos consecutivos<sup>1</sup> de  $\pi'$  par sim, par não. Começando do 0.

Arestas curvas (ou arcos): inteiros consecutivos de  $\pi'$  par sim, par não. Começando de 0.

Por exemplo,

$$\pi = (0 \ 3 \ -2 \ 1 \ 4 \ 6 \ 5 \ 7)$$

torna-se:

$$\pi' = (0 \ 5 \ 6 \ 4 \ 3 \ 1 \ 2 \ 7 \ 8 \ 11 \ 12 \ 9 \ 10 \ 13)$$

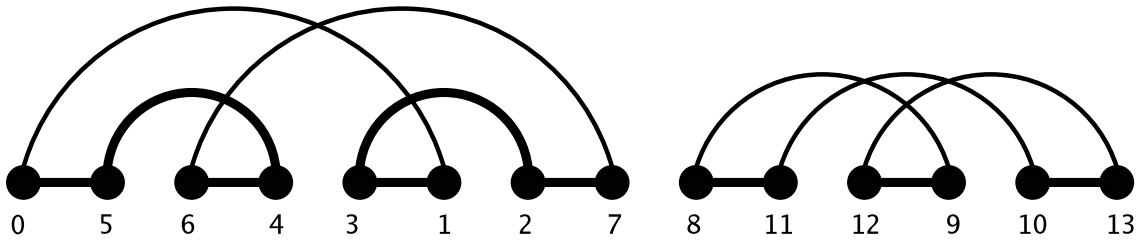


Figura 1: Diagrama de pontos de quebra de  $\pi$  em relação a identidade.

A Figura 1 apresenta o diagrama de pontos de quebra de  $\pi = (0\ 3\ -2\ 1\ 4\ 6\ 5\ 7)$  em relação a identidade.

Para definir as arestas retas precisamos olhar para o genoma origem ( $\pi$ ) e para definir as arestas curvas precisamos olhar para o genoma destino ( $\sigma$ ). Alguns autores enunciam que arestas retas apresentam o estado atual e as arestas curvas apresentam o estado desejado.

Um diagrama de pontos de quebra de um genoma  $\pi$ , de tamanho  $n$ , em relação a um genoma  $\sigma$  possui  $2n + 2$  arestas, todos os seus vértices têm grau 2 e é composto por ciclos de tamanho par. Os ciclos, que compõem o diagrama, são formados por um mesmo número de arestas retas e curvas. Neste texto estamos sempre considerando como genoma destino ( $\sigma$ ) a identidade.

O suporte de um arco é dado pelo intervalo de elementos de  $\pi'$  entre, e incluindo, seus terminais. Por exemplo, no diagrama da Figura 1, o suporte do arco cujos terminais são 0 e 1 é dado pelo intervalo  $[0..1] = \{0, 5, 6, 4, 3, 1\}$ . Um arco é orientado caso o tamanho de seu suporte seja ímpar, caso contrário o arco é não orientado. No diagrama de pontos de quebra é fácil de verificar se um dado arco é orientado, basta verificar se seus terminais possuem mesmo terminal nas arestas retas. No diagrama da Figura 1 apenas os arcos (4, 5) (suporte com tamanho 3) e (2, 3) (suporte com tamanho 3) são orientados. Perceba que no arco (4, 5), ambos os terminais estão no terminal direito de suas respectivas arestas retas e no arco (2, 3) eles estão no terminal esquerdo.

Um arco é orientado se, e apenas se, seus terminais são imagens de um par orientado da permutação original ( $\pi$ ). Perceba que o arco orientado (4, 5) tem origem no par orientado (3, -2) e que o arco orientado (2, 3) tem origem no par orientado (-2, 1).

Dois arcos estão sobrepostos se a intersecção dos suportes não é vazia e um deles não está contido propriamente no outro. O grafo de sobreposição de arcos é definido da seguinte forma:

- Vértices: arcos do diagrama de pontos de quebra. Os vértices podem ser orientados (pretos), caso corresponda a um arco orientado, ou não orientados (brancos), caso corresponda a um arco não orientado.
- Arestas: entre arcos sobrepostos.

---

<sup>1</sup>Dado o elemento  $x = \pi'(i)$ , seu consecutivo é  $y = \pi'(i + 1)$ .

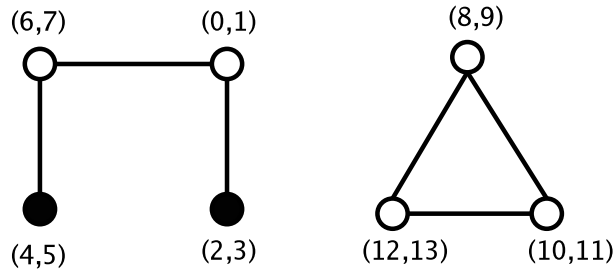


Figura 2: Grafo de sobreposição de arcos.

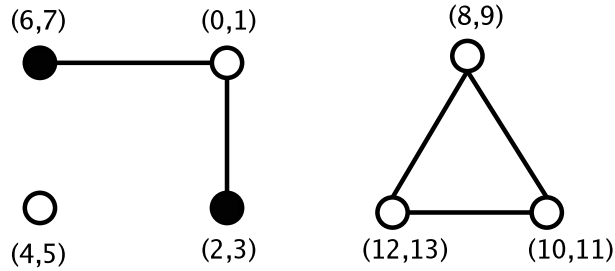


Figura 3: Grafo de sobreposição de arcos resultante da aplicação da reversão induzida pelo vértice  $(4, 5)$ .

A Figura 2 apresentado o grafo de sobreposição de arcos correspondente ao diagrama de pontos de quebra da Figura 1.

Um componente conexo do grafo de sobreposição de arcos é orientado se pelo menos um de seus vértices for orientado, caso contrário o componente é não orientado. É fácil perceber que um vértice (do grafo de sobreposição de arcos) tem um grau ímpar se, e somente se, ele é orientado. O suporte de um arco orientado é ímpar. Se retiramos do suporte seus terminais continuamos com um número ímpar. Se para cada arco, cujo suporte está contido no suporte do arco orientado, retirarmos seus terminais ainda continuaremos com um número ímpar. Este número resultante indica o total de arcos sobrepostos.

Uma reversão induzida por um vértice é a reversão induzida pelo par orientado correspondente na permutação original  $(\pi)$ . Só tem sentido falar em reversões induzidas por vértices quando estes são orientados.

No Fato 2, Bergeron mostra que ao aplicar uma reversão correspondente a um vértice orientado  $v$ , o efeito no grafo de sobreposição de arcos será o complemento do subgrafo de  $v$  e seus vértices adjacentes. A Figura 3 apresenta o grafo de sobreposição de arcos depois de aplicada a reversão induzida pelo arco  $v = (4, 5)$ . Perceba que o subgrafo ao qual o fato 2 se refere restringe-se aos vértices  $(4, 5)$  e  $(6, 7)$  e a aresta que os conecta. Na Figura 4 podemos visualizar o que ocorre no diagrama de pontos de quebra. Uma reversão  $\rho(i, j)$  em  $\pi$  é simulada por uma reversão  $\rho(2i - 1, 2j)$  em  $\pi'$ . Tal reversão isola o arco que induz a reversão e inverte a ordem dos elementos que estão entre os terminais do arco em  $\pi'$ .

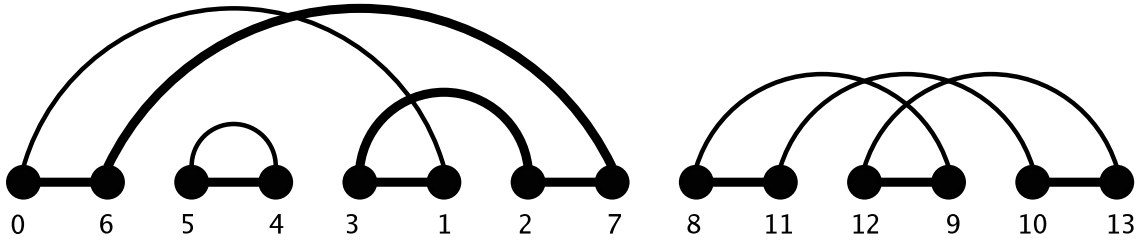


Figura 4: Diagrama de pontos de quebra resultante da aplicação da reversão induzida pelo vértice (4,5).

No Fato 3, Bergeron mostra que esta reversão ainda altera a orientação dos vértices adjacentes a  $v$ . Tal alteração ocorre por causa do isolamento de  $v$ , que faz com que os tamanhos do suporte dos arcos adjacentes a  $v$  sejam alterados de 1.

No Fato 4, Bergeron mostra como pontuar, Equação 1, uma reversão correspondente a um vértice orientado  $v$  através do grafo de sobreposição de arcos.

$$T + U - O - 1 \tag{1}$$

Na equação,  $T$  é o número total de vértice orientados,  $U$  é o número de vértices não orientados adjacentes a  $v$  e  $O$  é o número de vértices orientados adjacentes a  $v$ . A subtração  $T - O$  representa o número de vértices orientados não afetados pela reversão,  $U$  é o número de novos vértices orientados e  $-1$  é o vértice orientado sobre o qual aplica-se a reversão, o qual irá se tornar um componente de apenas um vértice e não orientado. No exemplo da Figura 2, se tomarmos  $v = (4, 5)$  temos  $T = 2$ ,  $U = 1$ ,  $O = 0$  e  $T + U - O - 1 = 2 + 1 - 0 - 1 = 2$ .

Um reversão é segura caso não crie novas componetes não orientadas, excetuando-se os vértices isolados. Hannenhalli e Pevzner mostraram em 1995 que qualquer seqüência de reversões seguras é ótima, ou seja, que a cada passo a distância é diminuída.

Bergeron prova que uma reversão orientada de pontuação máxima é segura. Este é um dos pontos chave de seu algoritmo para o cálculo da distância de reversão orientada.

A cobertura de um conjunto de vértices  $X$  no grafo de sobreposição de arcos é o intervalo mínimo que contém, na ordem circular, todos os intervalos de vértices em  $X$ . Por exemplo, considere o genoma  $\pi = (0\ 2\ 4\ 6\ 5\ 7\ 3\ 8\ 1\ 9)$ . A Figura 5 apresenta o diagrama de pontos de quebra de  $\pi$  e a Figura 6 apresenta o grafo de sobreposição de arcos. Os componentes conexos de seu grafo de sobreposição de arcos têm coberturas:

$$\begin{aligned} [4, 15] &= [4\ 7\ 8\ 11\ 12\ 9\ 10\ 13\ 14\ 5\ 6\ 15] \\ [8, 13] &= [8\ 11\ 12\ 9\ 10\ 13] \\ [16, 3] &= [16\ 1\ 2\ 17\ 0\ 3] \end{aligned}$$

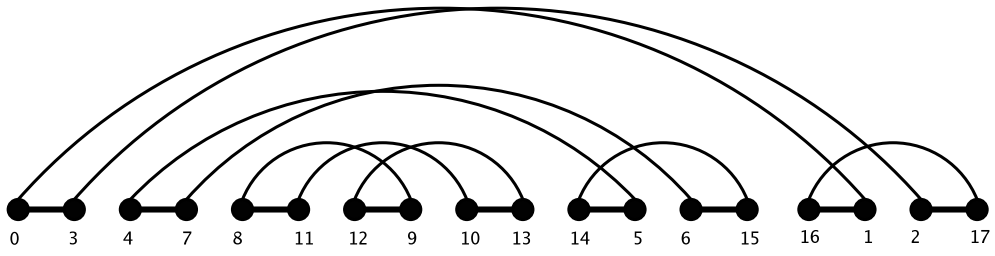


Figura 5: Diagrama de pontos de quebra de  $\pi$  em relação a identidade.

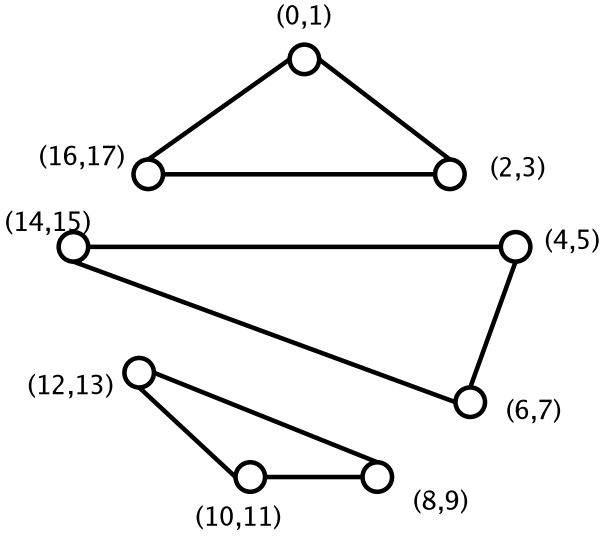


Figura 6: Grafo de sobreposição de arcos.

## Referências

- [1] Anne Bergeron. A very elementary presentation of the Hannenhalli-Pevzner theory. In *Proceedings of the 12th Annual Symposium of the Combinatorial Pattern Matching (CPM'2001)*, volume 2089 of *Lecture Notes in Computer Science*, pages 106–117, Berlin, Germany, September 2001. Springer-Verlag.
- [2] S. Hannenhalli and P. A. Pevzner. Transforming cabbage into turnip (polynomial algorithm for sorting signed permutations by reversals). In *Proceedings of the Twenty-Seventh Annual ACM Symposium on the Theory of Computing*, pages 178–189, Las Vegas, USA, May 1995.