

# Limites para distância e diâmetro em rearranjo de genomas por transposições

Luís Felipe Ignácio Cunha<sup>1</sup>

Orientadores: Celina Miraglia Herrera de Figueiredo<sup>1</sup>,  
Luis Antonio Brasil Kowada<sup>2</sup>

<sup>1</sup>COPPE/PESC – Universidade Federal do Rio de Janeiro  
Rio de Janeiro – RJ – Brasil

<sup>2</sup>IME – Universidade Federal Fluminense  
Niterói – RJ – Brasil.

lfignacio, celina@cos.ufrj.br, luis@vm.uff.br

**Abstract.** *Sorting by transpositions is a classic problem proposed in genome rearrangement and settled as NP-hard. The historical approach has been to study permutations candidates to be diametral. The transposition diameter is a related challenging open problem. We advance the study of both problems. We present tighter bounds for the distance of lonely permutations, we set the current lower bound for the diameter, and also show families that reach such bound. Furthermore, we propose a strategy to increase the lower bound.*

**Resumo.** *Ordenação por transposições é um clássico problema em rearranjo de genomas, provado ser NP-difícil. Historicamente tem sido considerado o estudo de permutações candidatas a serem diametraes. O diâmetro de transposição é assim relacionado ao problema de ordenação, em aberto, bastante desafiador. Avançamos em ambos problemas. Apresentamos limites justos para distância de famílias de permutações solitárias, estabelecemos o limite inferior para o diâmetro, e apresentamos famílias que por uniões de permutações solitárias atingem este limite. Além disso apresentamos uma estratégia que possivelmente aumentará o limite inferior.*

## 1. Introdução

Pelo princípio da evolução molecular, genomas dão origem a outros. Moléculas de DNA são responsáveis por toda informação genética dos seres vivos, e em proteínas estes conteúdos são quase similares, porém suas organizações são bem diferentes. Rearranjo de genomas é um conjunto de eventos mutacionais que afetam a organização do DNA. Distância de rearranjo trata de determinar o menor número de eventos de rearranjo para transformar um genoma em outro, também definida por determinar o menor número de operações necessárias para transformar uma permutação de inteiros em outra.

Para que seja melhor entendida a filogenia dentro das hipóteses evolutivas, diversos modelos matemáticos foram propostos, dentre os quais destacamos *ordenação por transposições*. Os interesses computacionais e combinatórios tornam-se cada vez mais desafiadores. Ordenação por transposições, foi introduzido por Bafna e

Pevzner [Bafna e Pevzner 1998] e foi provado ser  $NP$ -difícil [Bulteau et al. 2012]. Assim, busca-se por algoritmos exatos ou limites justos para classes de permutações. Apresentamos neste trabalho propriedades e limites para distância da classe das *permutações solitárias* e também limites algébricos para distância de *uniões* de permutações.

Bulteau *et al.* [Bulteau et al. 2012] provaram ser difícil determinar a distância para permutações muito próximas da identidade e estabeleceu uma separação entre dois problemas que historicamente eram pesquisados juntos: ordenação por transposições e diâmetro de transposição. O *diâmetro de transposição* permanece em aberto para  $n \geq 16$  e são conhecidos apenas limites para este valor, com melhor limite inferior conhecido de  $\lfloor \frac{n+1}{2} \rfloor + 1$  [Elias e Hartman 2006] e limite superior de  $\lfloor \frac{2n-2}{3} \rfloor$  [Eriksson et al. 2001].

Lu e Yang [Lu e Yang 2010] propuseram um novo limite inferior de  $\frac{17n+1}{33}$  para o diâmetro através de uniões das então definidas *permutações super-ruins*, como também descrito no livro *Combinatorics of Permutations* [Bóna 2012]. Invalidamos esta proposta demonstrando a não existência de permutações super-ruins, reestabelecendo assim o limite inferior  $\lfloor \frac{n+1}{2} \rfloor + 1$ . Apresentamos também famílias com distância igual ao limite inferior corrente, e propomos uma nova estratégia para aumentar o limite inferior do diâmetro.

Resultados foram estabelecidos durante o mestrado iniciado em setembro de 2011 e dissertação defendida em 27 de fevereiro de 2013. Tivemos as seguintes divulgações:

- Publicação no *SIAM Journal on Discrete Mathematics* [Cunha et al. 2013];
- Apresentação de artigo no *Brazilian Symposium on Bioinformatics 2012* e, publicação no *Lecture Notes in Bioinformatics* [Cunha et al. 2012];
- Publicação na revista *Matemática Contemporânea* [Cunha e Kowada 2010];
- Apresentação de pôster no *RECOMB Comparative Genomics 2012*.

Além disso, divulgações motivacionais ocorreram neste período, destacamos:

- Apresentação na *Escola Latino-iberoamericana de Pesquisa Operacional 2013*;
- Minicurso na *VI Semana da Matemática da UFF 2012*.

Descreveremos resumidamente alguns conceitos básicos e os principais resultados obtidos, de modo que cada resultado é indicado com a mesma numeração da dissertação. E para mais detalhes recomendamos o texto completo.

## 2. Ordenação por transposições

Um genoma é representado por uma *permutação*  $\pi_{[n]} = [\pi_1 \pi_2 \dots \pi_n]$ , tal que,  $1 \leq \pi_i \leq n$  para  $i = 1, \dots, n$ , com todos elementos distintos. Uma *transposição* atua em  $\pi$  pela troca de dois blocos contíguos de elementos. A *distância de transposição*,  $d(\pi)$ , é o menor número de transposições que transforma  $\pi$  em  $\iota = [1 2 \dots n]$ , a permutação identidade. O *diâmetro de transposição*  $TD(n)$  é o valor da maior distância de transposição dentre todas permutações com  $n$  elementos. Ou seja,  $TD(n) = \max_{\pi_{[n]}} \{d(\pi)\}$ .

Permutações são agrupadas em classes de equivalência tórica e aquelas que são toricamente equivalentes possuem mesma distância. Uma classe que possui uma única permutação, é chamada de *classe tórica unitária* e a permutação correspondente é uma *permutação solitária*  $u_{n,\ell} = [\bar{\ell} \ 2\bar{\ell} \ 3\bar{\ell} \ \dots \ n\bar{\ell}]$  onde  $\bar{x}$  é o resto da divisão de  $x$  por  $n+1$ . Uma condição suficiente e necessária para existência de  $u_{n,\ell}$  é que  $\gcd(n+1, \ell) = 1$ . Dada

a permutação  $u_{n,\ell}$ , a permutação  $u_{n,\ell^{-1}}$  é sua permutação inversa, onde  $\ell^{-1}$  é o inverso multiplicativo de  $\ell$  modulo  $n + 1$ . Além disso  $d(u_{n,\ell}) = d(u_{n,\ell^{-1}})$ .

Através do *diagrama de realidade e desejo*, limites não triviais para distância foram apresentados. Dada uma permutação  $\pi$ ,  $G(\pi)$  é o grafo  $G = (V, R \cup D)$ , onde  $V = \{0, -1, +1, -2, +2, \dots, -n, +n, -(n + 1)\}$ , as arestas são particionadas em dois conjuntos: as arestas de realidade  $R = \{(+\pi_i, -\pi_{i+1}) \mid i = 1, \dots, n - 1\} \cup \{(0, -\pi_1), (+\pi_n, -(n + 1))\}$ ; e as arestas de desejo  $D = \{(+i, -(i + 1)) \mid i = 0, \dots, n\}$ . Dizemos que um ciclo em  $G(\pi)$  tem tamanho  $k$ , ou que é um  $k$ -ciclo, se há exatamente  $k$  arestas de realidade. Se  $G(\pi)$  possui somente um ciclo, então  $\pi$  é *unicíclica*.

O número de ciclos ímpares em  $G(\pi)$ ,  $c_{\text{ímpar}}(\pi)$ , é alterado após uma transposição  $t$  da seguinte forma:  $c_{\text{ímpar}}(\pi.t) = c_{\text{ímpar}}(\pi) + x$ , onde  $x = \{-2, 0, 2\}$ , e dizemos que  $t$  é um  $x$ -movimento. Se  $\pi$  é unicíclica, então  $G(\pi)$  consiste de apenas um ciclo ímpar, e  $\pi$  contém um número par de elementos. Uma sequência composta somente por 2-movimentos nos dá o limite inferior de  $d(\pi) \geq \left\lceil \frac{(n+1) - c_{\text{ímpar}}(\pi)}{2} \right\rceil$ .

Estabelecemos regularidades de  $G(u_{n,\ell})$ , e com isto podemos estimar limites de distâncias de famílias de permutações solitárias. Dada  $u_{n,\ell}$ , temos que: todo ciclo possui mesmo tamanho  $k = \frac{n+1}{\text{mdc}(n+1,\ell-1)}$ , ou seja  $u_{n,\ell}$  é um  $k$ -ciclo; a sequência dos elementos não negativos dos ciclos é  $+i, +i + \ell - 1, +i + 2(\ell - 1), \dots, +i + (k - 1)(\ell - 1)$  para  $i = 0, \dots, |C(u_{n,\ell})| - 1$ ; e a posição do elemento  $i + y(\ell - 1)$  é  $i\ell^{-1} + y(1 - \ell^{-1})$ , onde pela ordem das posições destes elementos caracterizamos a existência de um  $x$ -movimento.

### 2.1. Limites de distância para permutações solitárias

Vista a dificuldade em determinar a distância de transposição de uma permutação, surge o desafio de determinar distâncias exatas ou limites justos para classes de permutações. Mostramos distâncias e limites para famílias de permutações solitárias. Das regularidades dos ciclos encontramos reduções entre permutações solitárias e limitamos as distâncias.

**Teorema 2.13.**  $d(u_{kq-2,k}) \leq d(u_{k(q-2),k}) + k - 1$ .

**Teorema 2.14.**  $d(u_{kq,k}) \leq d(u_{q(k-2)-2,q(k-3)-1}) + q + 2$ .

A Tabela 1 apresenta alguns limites encontrados, onde omitimos as permutações inversas, pois apresentam igual distância. Estes limites são justos, como verificados para todo  $n \leq 20$ . Limites superiores são obtidos através de reduções de permutações solitárias, e os limites inferiores pelas regularidades dos ciclos do diagrama de realidade e desejo.

**Tabela 1. Limites para permutações solitárias, suas inversas são omitidas.**

$n$	$\ell$	$d(u_{n,\ell})$		Referência
		limite inferior	limite superior	
$2q$	$2q - 1$	$q - 1$	$q + 1$	<b>Teorema 2.10</b>
$3q$	3	$\lfloor 3q/2 \rfloor$	$\lfloor 3q/2 \rfloor + 1$	<b>Teorema 2.12</b>
$3q + 1$	3	$\lfloor 3q/2 \rfloor$	$\lfloor 3q/2 \rfloor + 1$	<b>Teorema 2.12</b>
$2q$	4	$q - 1$	$q + 3$	<b>Corolário 2.15</b>

**Permutação nó  $u_{n,\ell^*}$**  Uma *permutações nó* é aquela única unicíclica que não admite um 2-movimento [Christie 1999]. Christie definiu essa permutação por  $\omega_{[n]}$ , unicíclica

com sequência de posições dos elementos não negativos do ciclo  $\langle 0, \frac{n}{2}, n, \frac{n}{2} - 1, n - 1, \dots, 1, n - (\frac{n}{2} - 1) \rangle$ . Esta é uma importante permutação solitária, as permutações não desempenham o papel central no estudo do diâmetro de transposição. Mostramos que as estas são as únicas candidatas a serem permutações super-ruins [Lu e Yang 2010].

**Corolário 2.5.** As permutações  $\omega_{[n]}$  são justamente as permutações nó  $u_{n,\ell^*}$  onde:

$$\ell^* = \begin{cases} 2q + 1 & \text{para } n = 6q, \\ 4q & \text{para } n = 6q - 2. \end{cases}$$

Pelo fato de não haver 2-movimento em  $u_{n,\ell^*}$  temos que  $d(u_{n,\ell^*}) \geq \frac{n}{2} + 1$ . Além disso, há uma sequência de  $\frac{n}{2} + 1$  transposições que ordena  $u_{n,\ell^*}$ , assim  $d(u_{n,\ell^*}) = \frac{n}{2} + 1$ .

## 2.2. Limites algébricos de uniões de permutações

A união de permutações consiste da justaposição de permutações, foi primeiramente abordada no contexto do problema do diâmetro de transposição. Mostramos propriedades algébricas da união de permutações, e mais particularmente união de permutações solitárias. Assim estimamos limites e valores exatos para a distância de mais famílias de permutações. Estes resultados são importantes no estudo do diâmetro de transposição.

Dadas as permutações  $\pi_{[r]}$  e  $\sigma_{[s]}$ , a *união* de  $\pi$  e  $\sigma$  é a permutação  $(\pi \uplus \sigma)_{[n]}$ , onde  $n = r + s + 1$  e:  $\pi \uplus \sigma = [\pi_1 \dots \pi_{r-1} \pi_r (r+1) (r+1+\sigma_1) \dots (r+1+\sigma_{s-1}) (r+1+\sigma_s)]$ .

Pela estrutura dos ciclos de  $G(\pi \uplus \sigma)$ , temos o seguinte limite inferior:  $d(\pi_{[r]} \uplus \sigma_{[s]}) \geq \frac{r+1-c_{\text{impar}}(\pi)}{2} + \frac{s+1-c_{\text{impar}}(\sigma)}{2}$ . Mostramos também algumas propriedades a respeito da operação de união, a saber:

**Proposição 3.3.** *Comutatividade.*  $\pi \uplus \sigma$  é toricamente equivalente a  $\sigma \uplus \pi$ . Assim,  $d(\pi \uplus \sigma) = d(\sigma \uplus \pi)$ .

**Proposição 3.6.** *Desigualdade triangular.*  $d(\pi \uplus \sigma) \leq d(\pi) + d(\sigma)$ .

**Proposição 3.8.** *Distributividade da inversa.*  $(\pi \uplus \sigma)^{-1} = \pi^{-1} \uplus \sigma^{-1}$ .

A partir destas propriedades temos os seguintes resultados:

**Proposição 3.9.** Dadas duas permutações unicíclicas  $\pi_{[r]}$  e  $\sigma_{[s]}$ , onde  $d(\pi) = \frac{r}{2}$ , temos:

$$d(\pi \uplus \sigma) = \begin{cases} \frac{r+s}{2} & \text{se } d(\sigma) = \frac{s}{2}, \\ \frac{r+s}{2} + 1 & \text{se } d(\sigma) = \frac{s}{2} + 1. \end{cases}$$

**Corolário 3.10.** Se  $\pi$  é unicíclica e  $d(\pi) = \frac{r}{2}$ , e  $s$  é par, então  $d(\pi \uplus \rho_{[s]}) = \frac{r+s}{2} + 1$ .

**Corolário 3.11.**  $d(u_{n,2} \uplus u_{n,\frac{n}{2}+1}) = n$ .

**Proposição 3.12.** Dadas duas permutações unicíclicas  $\pi_{[r]}$  e  $\sigma_{[s]}$ , onde  $d(\pi) = \frac{r}{2} + 1$  e  $d(\sigma) = \frac{s}{2} + 1$ , temos que:  $\frac{r+s}{2} + 1 \leq d(\pi \uplus \sigma) \leq \frac{r+s}{2} + 2$ .

**Teorema 3.13.** A distância de  $\rho_{[r]} \uplus \rho_{[r]}$  satisfaz a seguinte desigualdade:  $d(\rho_{[r]} \uplus \rho_{[r]}) \leq r + 1$ . A igualdade segue se  $r$  é par ou se  $r \equiv 3 \pmod{4}$ .

**União de permutações nó** Pelos resultados anteriores, a união de permutações nó:

$$\frac{p+q}{2} + 1 \leq d(u_{p,\ell^*} \uplus u_{q,\ell'^*}) \leq \frac{p+q}{2} + 2. \quad (1)$$

Mostramos uniões de permutações nó que atingem o limite inferior de (1).

**Teorema 3.19.**  $d(u_{n,\ell^*} \uplus u_{n,\ell^*}) = n + 1$ .

**Teorema 3.21.**  $d(u_{n,\ell^*} \uplus (u_{n,\ell^*})^{-1}) = n + 1$ .

**Teorema 3.23.** Dada a permutação  $\zeta_{[n]} = u_{6q-2,4q} \uplus u_{6q,2q+1}$ , temos que  $d(\zeta) = \lfloor \frac{n}{2} \rfloor + 1$ .

Por um algoritmo força bruta encontramos uniões de permutações nó que atingem o limite superior de (1), como a união  $u_{4,4} \uplus u_{10,8}$ , que possui distância 9. Obter permutações que tenham distância igual ao limite superior de (1) é uma estratégia promissora para estabelecer novos limites inferiores do diâmetro.

### 3. Diâmetro de transposição

O melhor limite inferior do diâmetro para  $n$  ímpar e  $n \geq 13$  foi apresentado em [Elias e Hartman 2006], este resultado foi baseado em [Eriksson et al. 2001] que obtiveram permutações diametrais para  $n = 13$  e  $n = 15$ . Elias e Hartman consideraram para  $n \equiv 1 \pmod{4}$ , a permutação  $(\rho_{[4]} \uplus \rho_{[8]}) \uplus \rho_{[3]} \uplus \rho_{[3]} \dots \uplus \rho_{[3]}$ , e para  $n \equiv 3 \pmod{4}$ , a permutação  $(\rho_{[4]} \uplus \rho_{[10]}) \uplus \rho_{[3]} \uplus \rho_{[3]} \dots \uplus \rho_{[3]}$ . Para permutações com número par de elementos, até agora não se conhece alguma que sua distância seja maior que a da reversa.

Lu e Yang [Lu e Yang 2010] propuseram um melhor limite de  $TD(n) \geq \frac{17n+1}{33}$  pelas uniões de cópias de permutações super-ruins.  $\gamma$  é uma *permutação super-ruim* se: i)  $\gamma$  é unicíclica; ii) Não há 2-movimento a ser aplicado em  $\gamma$ ; iii)  $d(\gamma \uplus \gamma) \geq n + 2$ . Lu e Yang afirmaram que a permutação  $u_{10,8}$  seria super-ruim, a qual teria sido obtida através de um algoritmo BFS dentre todas as possíveis permutações.

**O atual limite inferior** Dos resultados obtidos sobre as permutações nó e sobre operações de uniões de permutações, mostramos:

**Teorema 4.6.** Não existem permutações super-ruins.

As permutações nó são as únicas que satisfazem as condições i) e ii) anteriores, então existiriam permutações super-ruins se a união de duas cópias da mesma permutação nó tivesse distância igual ao limite superior de (1), porém mostramos que  $d(u_{n,\ell^*} \uplus u_{n,\ell^*}) = n + 1$  (Teorema 3.19), invalidando as condições de existência de permutação super-ruim. Retornamos portanto, ao limite inferior provado por Elias e Hartman.

**Construção das famílias mais distantes** Mostramos condições suficientes para que uniões de algumas permutações nó não atinjam ao limite superior de (1), não sendo úteis portanto para construirmos permutações mais distantes. Entretanto, as uniões  $u_{4,4} \uplus u_{10,8}$  e  $u_{4,4} \uplus u_{12,5}$  atingem o limite superior. Assim construímos novas famílias que atingem o atual limite inferior do diâmetro, as permutações nó  $u_{10,8}$  e  $u_{12,5}$  se aplicam como raízes destas famílias, assim como  $\rho_{[10]}$  e  $\rho_{[8]}$  em [Elias e Hartman 2006].

Um limite inferior melhor para o diâmetro possivelmente será obtido através da obtenção de  $\xi_{[n]}$ , construída pela união de quatro permutações nó distintas  $\xi_{[n]} = \pi \uplus \sigma \uplus \delta \uplus \epsilon$ , tal que  $d(\xi)$  atinge o limite superior  $\frac{n+1}{2} + 2$ . Uma condição para obter  $\xi$  é que: cada um dos seis pares de uniões tomados dois a dois necessite de dois 0-movimentos; e cada uma das quatro uniões de três permutações necessite de três 0-movimentos.

Uma consequência do diâmetro limitado inferiormente por  $TD(n) \geq d(\xi) = \frac{n+1}{2} + 2$

é que para  $n$  par, a reversa não seria diametral, já que  $\xi_{[n]}$   $n$  é ímpar e para  $TD(n+1)$  sabemos que  $TD(n) \leq TD(n+1) \leq TD(n) + 1$  e  $\frac{n+1}{2} + 2 > d(\rho_{[n+1]}) = \frac{n+1}{2} + 1$ .

#### 4. Conclusões

Apresentamos neste trabalho importantes contribuições para os problemas de ordenação por transposições e diâmetro de transposição.

No problema de ordenação, exploramos a classe das permutações solitárias e encontramos valores exatos e limites justos de distância para algumas famílias. Estabelecemos também um estudo algébrico em torno da operação de união, já que esta operação tem se mostrado como uma estratégia promissora para a obtenção de famílias mais distantes.

Com isso, no problema de diâmetro encontramos uma invalidação da proposta de melhor limite inferior publicada por Lu e Yang e estabelecemos o atual limite. Encontramos também famílias alternativas que atingem as maiores distâncias conhecidas e apresentamos uma estratégia que possivelmente aumentará o limite inferior.

O ótimo desempenho no mestrado, realizado em um ano e meio, publicando numa das melhores revistas da área, além da apresentação no principal congresso de bioinformática do país permitiu aprovação em primeiro lugar para o doutorado em 2013 do PESC/COPPE (conceito CAPES 7). Atualmente avançamos na elaboração de algoritmos aproximativos para ordenação por transposições e outros problemas em rearranjo.

#### Referências

- Bafna, V. e Pevzner, P. A. (1998). Sorting by transpositions. *SIAM J. Discrete Math.*, 11(2):224–240.
- Bóna, M. (2012). *Combinatorics of Permutations*. The CRC Press.
- Bulteau, L., Fertin, G. e Rusu, I. (2012). Sorting by transpositions is difficult. *SIAM J. Discrete Math.*, 26(3):1148–1180.
- Christie, D. A. (1999). *Genome Rearrangement Problems*. Phd. thesis, University of Glasgow, UK.
- Cunha, L. F. I. e Kowada, L. A. B. (2010). Upper bounds and exact values on transposition distance of permutations. *Mat. Contemp.*, 39:77–84.
- Cunha, L. F. I., Kowada, L. A. B., Hausen, R. A. e de Figueiredo, C. M. H. (2012). Transposition diameter and lonely permutations. In *Advances in Bioinformatics and Computational Biology*, pages 1–12. LNBI, Springer-Verlag.
- Cunha, L. F. I., Kowada, L. A. B., Hausen, R. A. e de Figueiredo, C. M. H. (2013). Advancing the transposition distance and diameter through lonely permutations. *SIAM J. Discrete Math.*, 27(4):1682–1709.
- Elias, I. e Hartman, T. (2006). A 1.375-approximation algorithm for sorting by transpositions. *IEEE/ACM Trans. Comput. Biol. Bioinformatics*, 3(4):369–379.
- Eriksson, H., Eriksson, K., Karlander, J., Svensson, L. J. e Wastlund, J. (2001). Sorting a bridge hand. *Discrete Math.*, 241(1-3):289–300.
- Lu, L. e Yang, Y. (2010). A lower bound on the transposition diameter. *SIAM J. Discrete Math.*, 24(4):1242–1249.