
Instituto de Computação
Universidade Estadual de Campinas

Programação Linear Inteira e Programação Lógica por
Restrições para Problemas de Rearranjo de Genomas

Victor de Abreu Iizuka

Orientador: Zanoni Dias

Resumo

Rearranjo de Genomas é uma área que vem recebendo crescente atenção de pesquisadores no decorrer da última década. Isso se deve ao fato que os rearranjos são eventos que melhor caracterizam a distância evolutiva entre duas espécies do que o estudo de mutações pontuais. No modelo de rearranjo de genomas, um genoma é representado por uma seqüência de elementos inteiros, onde cada inteiro representa um gene ou um grupo de genes. Apareceram novas linhas de pesquisas que se afastam dos paradigmas clássicos, que seguem a linha de pensamento dos trabalhos de Pevzner [3, 4, 5, 18], na área de rearranjo de genomas. A primeira é a modelagem do problema com programação linear inteira por Dias e de Souza [13], que apresentaram um modelo de tamanho polinomial. A segunda é a modelagem do problema com programação lógica por restrições por Dias e Dias [12], mostrando uma forma alternativa ao modelo de programação linear inteira. O objetivo do projeto é aprimorar os modelos já existentes de programação linear inteira e programação lógica por restrições para o problema de rearranjo de genomas, utilizando técnicas e informações que ainda não foram usadas no desenvolvimento destes modelos. Pretendemos, com isso, utilizar contribuições de pesquisas já consolidadas nestas áreas para propiciar novos avanços na área de rearranjo de genomas.

1 Introdução

O processo evolutivo foi o principal responsável pela diferenciação entre os seres vivos e uma das teorias contemporâneas acerca do modo como ocorre esse processo afirma que, durante o curso da evolução, mudanças genéticas aconteceram criando diferentes espécies de seres-vivos.

Muitas dessas mudanças são devido a mutações pontuais que alteram a cadeia de DNA, impedindo que a informação seja expressa, ou que seja expressa de um modo diferente. Tais alterações debilitam, na maioria dos casos, o organismo portador ou proporcionam vantagens no processo de seleção natural.

A comparação de seqüências é o método mais usual de se caracterizar a ocorrência de mutações pontuais, sendo um dos problemas mais abordados em Biologia Computacional. O interesse em fazer tal comparação é encontrar a distância de edição [27], que é o número mínimo de operações de inserção, remoção e substituição necessárias que transformam uma seqüência em outra.

A distância de edição é uma medida capaz de estimar a distância evolutiva entre duas cadeias, mas não possui a informação de quais operações globais foram utilizadas para a transformação de uma seqüência em outra. Estas operações globais são os chamados Rearranjos de Genomas, que podem ser, por exemplo, reversões, transposições, fissões e fusões.

Um conceito de distância pode ser definido para qualquer classe de rearranjo como sendo o menor número de operações pertencentes a essa classe

que são necessários para transformar um genoma em outro. Por exemplo, chama-se a distância de reversão o menor número de reversões necessárias para transformar um genoma em outro [3] e a distância de transposição é o menor número de transposições [4].

Estudos recentes mostram que os rearranjos de genomas são mais apropriados que mutações pontuais quando se deseja comparar os genomas de duas espécies [26]. Nesse contexto, a distância evolutiva entre dois genomas pode ser estimada pelo conceito de distância para uma classe de rearranjo definido no parágrafo anterior.

Na literatura estão documentados vários avanços no que diz respeito a implementação de métodos computacionais para os vários tipos de rearranjos, a seguir apresentamos um breve histórico do que já foi estudado, evidenciando os pontos onde esforços são ainda necessários.

Para fins computacionais, um genoma é representado por uma n -tupla de genes, quando não há genes repetidos essa n -tupla é chamada de permutação. Uma permutação é representada como $\pi = (\pi_1 \pi_2 \dots \pi_n)$, onde π_i , para todo $1 \leq i \leq n$, representa um gene (ou um grupo de genes) e os vários eventos de rearranjo ρ são aplicados a π de uma maneira específica.

O evento de transposição ocorre quando dois blocos adjacentes no genoma trocam de posição. Representando-se um genoma pela permutação $\pi = (\pi_1 \pi_2 \dots \pi_n)$, uma transposição $\rho(i, j, k)$, onde $1 \leq i < j < k \leq n + 1$, faz com que um intervalo $[i, j - 1]$ de π seja inserido entre π_{k-1} e π_k , de outra

forma, a transposição $\rho(i, j, k)$ aplicada a permutação π é escrita como:

$$\pi \cdot \rho = (\pi_1, \pi_2, \dots, \pi_{i-1}, \pi_j, \dots, \pi_{k-1}, \pi_i, \pi_{i+1}, \dots, \pi_{j-2}, \pi_{j-1}, \pi_k, \dots, \pi_n).$$

O problema de encontrar a distância de transposição, que envolve encontrar o menor número de transposições necessárias para transformar um genoma em outro, foi estudado por Bafna e Pevzner [4], que apresentaram um algoritmo capaz de fornecer uma resposta aproximada na razão de 1.5, além de derivar um importante limitante inferior para o problema.

Entretanto, talvez a maior contribuição do trabalho de Bafna e Pevzner [4] resida nas heurísticas implementadas, que utilizaram o conceito de *break-points*, elementos adjacentes em um dos genomas, mas não no outro, e uma ferramenta que chamou de grafo de ciclos orientados com arestas de cores alternadas, que fora bastante utilizada em trabalhos posteriores. Foram apresentados várias questões em aberto como verificar a complexidade do problema da distância de transposições e o diâmetro, que é a maior distância possível entre duas permutações de tamanho n . O problema do diâmetro foi estudado por Meidanis, Walter e Dias [22].

Entretanto, atualmente não se conhece nenhuma prova de que o problema da distância de transposição pertence a classe dos problemas NP-Difícil, sendo que também não existem evidências da existência de um algoritmo polinomial, o que torna o estudo do problema interessante. Recentemente, Elias e Hartman [14] propuseram um novo algoritmo de aproximação na razão

de 1.375 e Labarre [19] apresentou novos limitantes, além de definir classes de permutações em que a distância de transposição pode ser calculada em tempo e espaço lineares. Labarre usou a idéia de compartimentalizar o problema em várias classes menores que podem ser resolvidas polinomialmente. Essa idéia fora usada anteriormente por outros pesquisadores, podendo citar Fortuna [15], que mostrou um subconjunto de permutações, conhecidas como permutações fáceis, pode ser resolvido em tempo polinomial.

Uma generalização do problema de transposição é a operação de troca de blocos, cuja solução pode auxiliar na resolução do problema da distância de transposição, já que a distância de troca de blocos entre dois genomas é um problema com solução polinomial para qualquer permutação [10].

Seja $\pi = (\pi_1 \pi_2 \dots \pi_n)$ uma permutação, uma troca de blocos $\rho(i, j, k, l)$, onde $1 \leq i < j < k < l \leq n + 1$ é uma operação em que dois blocos $[\pi_i \dots \pi_{j-1}]$ e $[\pi_k \dots \pi_{l-1}]$ são trocados de posição, de outra forma, a troca de blocos $\rho(i, j, k, l)$ aplicada ao genoma π é escrita como:

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

O caso especial da transposição ocorre quando $j = k$. Christie [10] apresentou um algoritmo polinomial para o problema e utilizou, na implementação do algoritmo, conceitos já conhecidos na literatura sobre transposições.

O evento de reversão ocorre quando um bloco do genoma é invertido. Seja $\rho(i, j)$ uma reversão de um intervalo $[i, j]$ e $\pi = (\pi_1 \pi_2 \dots \pi_n)$ uma

permutação, no caso em que a permutação é orientada a reversão ρ aplicada a permutação π é escrita como:

$$\pi \cdot \rho = (\pi_1, \pi_2, \dots, \pi_{i-1}, -\pi_j, -\pi_{j-1}, \dots, -\pi_{i+1}, -\pi_i, \pi_{j+1}, \dots, \pi_n)$$

No caso em que a permutação não é orientada a reversão ρ aplicada a permutação π é escrita como:

$$\pi \cdot \rho = (\pi_1, \pi_2, \dots, \pi_{i-1}, \pi_j, \pi_{j-1}, \dots, \pi_{i+1}, \pi_i, \pi_{j+1}, \dots, \pi_n)$$

Um estudo inicial sobre o problema da distância de reversão foi apresentado em 1993 por Bafna e Pevzner [3] que apresentaram um algoritmo de aproximação com razão de 1.5 quando a orientação dos genes é conhecida e 1.75 caso contrário.

Conhecer a orientação dos genes em um genoma é importante no problema de reversão, pois existem algoritmos polinomiais caso a orientação seja conhecida. Entretanto, se não se conhece a orientação dos genes o problema de encontrar a distância de reversão pertence a classe de problemas NP-Difícil [9].

O primeiro algoritmo polinomial para o problema de reversão com orientação conhecida foi criado por Hannenhalli e Pevzner [18] que fez uso de várias operações aplicadas a uma estrutura intermediária conhecida como grafo de *breakpoints*. Posteriormente a estratégia usada por Hannenhalli e Pevzner foi simplificada por Bergeron [7]. Atualmente já existe um algo-

ritmo com complexidade sub-quadrática [28] e, quando apenas a distância é necessária, um algoritmo linear pode ser usado [2].

Um resultado importante obtido por Meidanis, Walter e Dias [23], mostrou que toda teoria sobre reversões desenvolvida para genomas lineares pode ser adaptada facilmente para genomas circulares, que são comuns em seres inferiores como vírus e bactérias.

Quando a orientação dos genes não é conhecida existem algoritmos de aproximação que se seguiram ao de Bafna e Pevzner citado anteriormente como, por exemplo, o algoritmo implementado por Berman, Hannenhalli e Karpinski [8] com razão de aproximação de 1.375.

Na natureza um genoma não sofre apenas eventos de reversão ou de transposição, ele está exposto a diversos tipos de eventos diferentes. Portanto um estudo acerca da distância entre dois genomas sendo possíveis diferentes tipos de eventos é interessante.

No que diz a respeito apenas aos eventos de reversão e transposição existe o trabalho de Hannenhalli e co-autores [17], que analisaram a evolução de genomas por diferentes eventos, em especial reversões e transposições, e de Gu, Peng e Sudborough [16] que criaram um algoritmo de aproximação para computar a distância entre dois genomas com a orientação dos genes conhecida permitindo operações de reversão, transposição e reversão+transposição simultaneamente.

Em 1998, Walter, Dias e Meidanis [29] apresentaram um algoritmo de aproximação para a distância de reversão e transposição, além de limitantes

para o diâmetro de reversão e transposição em permutações orientadas que foram posteriormente melhorados [24].

É importante mencionar que os trabalhos na área de rearranjo de genomas apresentam várias abordagens indo desde a “clássica” que segue a linha de pensamento dos trabalhos de Pevzner [3, 4, 5, 18] ao utilizar ferramentas como o grafo de *breakpoints*, passando pelo formalismo algébrico de Meidanis e Dias [20], criado com o intuito de diminuir a dependência de representações visuais ao se expor argumentos, provas e teoremas em rearranjo de genomas.

Outra abordagem que difere da teoria “clássica” é a abordagem proposta por Bergeron [7]. Bergeron apresentou um algoritmo mais simples que o anterior de Hannenhall e Pevzner [18] e utiliza o conceito de pares e intervalos fechados, que não haviam sido implementados anteriormente.

Entretanto, existem novas linhas de pesquisas que realmente se afastam dos paradigmas já consolidados na área de rearranjo de genomas. A primeira é a modelagem do problema com Programação Linear Inteira realizada por Dias e de Souza [13], que criaram um modelo para distância de reversão e transposição. A segunda é a modelagem do problema com Programação Lógica por Restrições apresentada por Dias e Dias [12], que utiliza o conhecimento dos limitantes do problema. Ambos modelos serão tratados na Seção 2.

O Instituto de Computação da Universidade Estadual de Campinas, local onde este projeto de pesquisa será realizado, desenvolve trabalhos na área de Rearranjo de Genomas ao longo dos últimos anos com os projetos coordena-

dos pelos professores doutores João Meidanis e Zanoni Dias.

Em termos cronológicos, os primeiros trabalhos importantes foram apresentados em 1997 por Meidanis, Walter e Dias e estavam relacionados a distância de reversão em cromossomos circulares [21] e ao diâmetro de transposição entre permutações de um dado tamanho n [22]. No ano seguinte, houve a publicação de um estudo sobre a distância de reversão e transposição em cromossomos lineares [29].

Um ponto importante a ser mencionado é que o Instituto de Computação foi responsável pela elaboração de uma maneira alternativa de lidar com rearranjo de genomas, o chamado formalismo algébrico publicado em 2000 por Meidanis e Dias [20] e utilizados em trabalhos posteriores como, por exemplo, nas dissertações de mestrado apresentadas por Fortuna [15] e Almeida [1].

Esta proposta está organizada como segue: a Seção 2 apresentar o modelo de Programação Linear Inteira e o modelo de Programação Lógica por Restrições, a Seção 3 descreve o objetivo do trabalho, a Seção 4 apresenta o cronograma de atividades e o plano de trabalho a ser seguido no decorrer do projeto, a Seção 5 descreve a metodologia a ser usada no trabalho e discute como serão analisados os resultados obtidos durante a pesquisa.

2 Modelos

Nesta seção serão apresentados as linhas de pesquisas que realmente se afastam dos paradigmas já consolidados na área de rearranjo de genomas. Na

Seção 2.1 apresentaremos o modelo de Programação Linear Inteira e na Seção 2.2 o modelo de Programação Lógica por Restrições.

2.1 Programação Linear Inteira

A abordagem de programação linear inteira que vamos utilizar é apresentada no trabalho de Dias e de Souza [13]. Neste trabalho define-se um modelo para o problema da distância com tamanho polinomial. O modelo são específicos para os eventos de reversão, transposição ou reversão+transposição simultaneamente.

2.2 Programação Lógica por Restrição

A abordagem da programação lógica por restrições que vamos utilizar é apresentada no trabalho de Dias e Dias [12]. Neste trabalho o modelo foca nos eventos de transposições, usando os conhecimentos sobre os limitantes do problema da distância de transposição. Utiliza alguns resultados propostos por Christie [11], Bafna e Pevzner [6] e Labarre [19].

Duas formulações foram apresentadas, sendo que uma é baseada na teoria do Problema de Satisfação de Restrições (CSP) e a outra na teoria do Problema de Otimização com Restrições (COP).

Os modelos baseados em CSP precisam do valor da distância $d(\pi)$ para inicializar as restrições e variáveis que representam as permutações. Para resolver este problema escolhemos um candidato $T \in [L..U]$, onde L é um

valor conhecido do limitante inferior e U é um valor conhecido do limitante superior. Com isso tentamos encontrar uma combinação apropriada de T transposições. Caso o CSP falhar com o candidato T , escolhemos um outro candidato incrementando o valor de T , isto é feito usando uma estratégia *bottom-up* e por definição não se verifica um valor maior que qualquer limitante superior.

Os modelos baseados em COP precisam de um valor para o limitante superior e seus predicados diferem do modelo anterior. Utilizamos variáveis binárias para indicar quando uma transposição alterou a permutação. Os modelos COP foram usados para analisar os limitantes superiores definidos por Bafna e Pevzner [6] e por Labarre [19].

3 Objetivos

Um dos objetivos do projeto é procurar um método para melhorar o tempo de execução do modelo de programação linear inteira, desta forma tornando o uso desta abordagem viável para o uso na prática.

O modelo de programação linear inteira apresentado Dias e de Souza [13] possui um tempo de execução muito grande, por exemplo o programa retornou *timeout* para instâncias com permutações de tamanho 10, sendo que o limite de tempo era de 10 horas. A idéia é aplicar técnicas usadas em outros problemas de programação linear inteira para melhorar os limitantes do problema, desta forma reduzindo o tempo de execução do modelo.

O outro objetivo do projeto é aprimorar o modelo de programação lógica por restrições apresentado por Dias e Dias [12]. O modelo não está usando informações que poderiam melhorar os resultados. Por exemplo, podemos citar o grafo de ciclos da permutação que pode auxiliar na escolha de qual conjunto de transposições vai ser ramificado primeiro.

4 Plano de Trabalho

As Tabelas 1 e 2 descrevem a distribuição das atividades a serem realizadas durante a execução deste trabalho. As atividades a serem desenvolvidas são as seguintes:

1. Cumprimento das disciplinas obrigatórias do programa de mestrado.
2. Revisão Bibliográfica dos artigos clássicos sobre rearranjo de genomas.
3. Escrita do projeto de mestrado e exame de qualificação.
4. Estudo e desenvolvimento sobre o modelo de programação linear inteira para o problema de rearranjo de genomas.
5. Estudo e desenvolvimento sobre o modelo de programação lógica por restrições para o problema de rearranjo de genomas.
6. Testes dos modelos desenvolvidos.
7. Escrita da dissertação. A escrita começará a ser feita a partir dos primeiros resultados.
8. Revisão final do texto da dissertação.
9. Defesa da dissertação.

Ativ.	2009											2010	
	03	04	05	06	07	08	09	10	11	12	01	02	
1	x	x	x	x	x	x	x	x	x	x			
2	x	x	x	x	x	x	x	x	x	x			
3					x	x	x						
4											x	x	

Tabela 1: Cronograma de atividades (mar/2009 - fev/2010)

Ativ.	2010											2011	
	03	04	05	06	07	08	09	10	11	12	01	02	
4	x	x	x	x									
5					x	x	x	x	x	x			
6		x	x	x				x	x	x			
7		x	x	x				x	x	x			
8											x		
9												x	

Tabela 2: Cronograma de atividades (mar/2010 - fev/2011)

5 Metodologia e Formas de análise dos resultados

Nesta seção iremos apresentar quais as metodologias que serão usadas no projeto e como serão feitas as análises dos resultados.

Para programação linear inteira serão aplicadas técnicas para melhorar os limitantes, por exemplo a Relaxação Lagrangeana [30, 25], com a intenção de melhorar o modelo, causando a redução do tempo de execução deste método. Caso o resultado não melhorar significativamente, outras técnicas poderão ser usadas desenvolvendo um novo modelo e, por exemplo, aplicar a técnicas de

Geração de Colunas [30], caso o novo modelo possuir um número polinomial de restrições e um número exponencial de colunas (variáveis), ou aplicar a técnica de *Branch-and-Cut* [30, 25], caso o novo modelo possuir um número polinomial de variáveis e um número exponencial de restrição.

Para analisar os resultados do modelo de programação linear inteira, iremos verificar se a técnica aplicada realmente melhorou o modelo, se o novo modelo obteve o resultado correto usando um tempo de execução menor e se houve uma melhora significativa nos limitantes em relação ao modelo apresentado por Dias e de Souza [13].

Já para programação lógica por restrições, a idéia é usar ferramentas mencionadas em trabalhos na área para melhorar o modelo, tornando-o mais eficiente.

A análise dos resultados obtidos pelo modelo de programação lógica por restrições será feita simplesmente considerando a performance, se o tempo de execução for menor e se conseguiu reduzir a quantidade de recalculos feitos pelo modelo.

Referências

- [1] A. A. M. Almeida. Comparação algébrica de genomas: O caso da distância de reversão. Master's thesis, Institute of Computing, University of Campinas, 2007.

- [2] D. A. Bader, B. M. E. Moret, and M. Yan. A linear-time algorithm for computing inversion distance between signed permutations with an experimental study. In *Proceedings of the Seventh Workshop on Algorithms and Data Structures (WADS'01)*. Springer Verlag, 2001.
- [3] V. Bafna and P. A. Pevzner. Genome rearrangements and sorting by reversals. In IEEE, editor, *Proceedings of the 34th Annual Symposium on Foundations of Computer Science (FOCS'93)*, pages 148–157, Palo Alto, USA, November 1993. IEEE Computer Society Press.
- [4] V. Bafna and P. A. Pevzner. Sorting by transpositions. In *Proceedings of the Sixth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 614–623, San Francisco, USA, January 1995.
- [5] V. Bafna and P. A. Pevzner. Genome rearrangements and sorting by reversals. *SIAM Journal on Computing*, 25(2):272–289, 1996.
- [6] V. Bafna and P. A. Pevzner. Sorting by transpositions. *SIAM Journal on Discrete Mathematics*, 11(2):224–240, May 1998.
- [7] A. Bergeron. A very elementary presentation of the Hannenhalli-Pevzner theory. In *Proceedings of the 12th Annual Symposium of Combinatorial Pattern Matching (CPM'2001)*, volume 2089 of Lecture Notes in Computer Science, pages 106–117, Berlin, Germany, September 2001. Springer-Verlag.

- [8] P. Berman, S. Hannenhalli, and M. Karpinski. 1.375 approximation algorithm for sorting by reversals. In *Proceedings of the 10th European Symposium on Algorithms (ESA '2002)*, Lecture Notes in Computer Science, Rome, Italy, September 2002. Springer.
- [9] A. Caprara, G. Lancia, and S.-K. Ng. Sorting by reversals is difficult. In *Proceedings of the First International Conference on Computational Molecular Biology*, pages 75–83, New York, USA, 1997. ACM Press.
- [10] D. A. Christie. Sorting permutations by block-interchanges. *Information Processing letters*, 60(4):165–169, November 1996.
- [11] D. A. Christie. *Genome Rearrangements Problems*. PhD thesis, Glasgow University, 1998.
- [12] U. Dias and Z. Dias. Constraint programming models for transposition distance problem. In *Proceedings of the 4th Brazilian Symposium on Bioinformatics (BSB'2009)*, Lecture Notes in Computer Science, pages 13–23, Porto Alegre, Brazil, 2009. Springer-Verlag.
- [13] Z. Dias and C. de Souza. Polynomial-size ILP models for rearrangements distance problems. In *BSB2007 Poster Proceedings*, 2007.
- [14] I. Elias and T. Hartman. A 1.375-approximation algorithms for sorting by transpositions. *IEEE/ACM Trans. Comput. Biol. Bioinformatics*, 3(4):369–379, 2006.

- [15] V. J. Fortuna. Distâncias de transposição entre genomas. Master's thesis, Institute of Computing, University of Campinas, 2005.
- [16] Q.-P. Gu, S. Peng, and H. Sudborough. Approximating algorithms for genome rearrangements. In *Proceedings of the 7th Workshop on Genome Informatics (GIW'96)*, Tokyo, Japan, December 1996.
- [17] S. Hannenhalli, C. Chappey, E. V. Koonin, and P. A. Pevzner. Genome sequence comparison and scenarios for gene rearrangements: a test case. *Genomics*, 30:299–311, 1995.
- [18] 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.
- [19] A. Labarre. New bounds and tractable instances for the transposition distance. *IEEE/ACM Trans. Comput. Biol. Bioinformatics*, 3(4):380–394, 2006.
- [20] J. Meidanis and Z. Dias. An algebraic formalism for genome rearrangements. In D. Sankoff and J. H. Nadeau, editors, *Comparative Genomics: Empirical and Analytical Approaches to Gene Order Dynamics, Map Alignment and Evolution of Gene Families*, pages 213–223. Kluwer Academic Publishers, November 2000.

- [21] J. Meidanis, M. E. M. T. Walter, and Z. Dias. Distância de reversão de cromossomos circulares. In *Proceedings of the XXIV Seminário Integrado de Software e Hardware (SEMISH'97)*, pages 119–131, August 1997. In Portuguese.
- [22] J. Meidanis, M. E. M. T. Walter, and Z. Dias. Transposition distance between a permutation and its reverse. In R. Baeza-Yates, editor, *Proceedings of the 4th South American Workshop on String Processing (WSP'97)*, pages 70–79, Valparaiso, Chile, 1997. Carleton University Press.
- [23] J. Meidanis, M. E. M. T. Walter, and Z. Dias. Reversal distance of signed circular chromosomes. Technical Report IC-00-23, Institute of Computing, University of Campinas, December 2000.
- [24] J. Meidanis, M. E. M. T. Walter, and Z. Dias. A lower bound on the reversal and transposition diameter. *Journal of Computational Biology*, 9(5), 2002. Submitted: 14/January/2000. Accepted: 16/July/2002.
- [25] G. Nemhauser and L. Wolsey. *Integer and Combinatorial Optimization*. Wiley-Interscience, 1988.
- [26] J. D. Palmer and L. A. Herbon. Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. *Journal of Molecular Evolution*, 27:87–97, 1988.

- [27] J. C. Setubal and J. Meidanis. *Introduction to Computational Molecular Biology*. PWS Publishing Company, 1997.
- [28] E. Tannier and M.-F Sagot. Sorting by reversals in subquadratic time. In *Proceedings of CPM 2004, LNCS. 2004*, volume 3109, pages 1–13, 2004.
- [29] M. E. M. T. Walter, Z. Dias, and J. Meidanis. Reversal and transposition distance of linear chromosomes. In *Proceedings of the String Processing and Information Retrieval (SPIRE'98)*, 1998.
- [30] L. Wolsey. *Integer Programming*. Wiley-Interscience, 1998.