Comparação algébrica de genomas: o caso da distância de reversão

Este exemplar corresponde à redação final da Dissertação devidamente corrigida e defendida por André Atanasio Maranhão Almeida e aprovada pela Banca Examinadora.

Campinas, 11 de Abril de 2007.

Prof. Dr. João Meidanis (Orientador)

Dissertação apresentada ao Instituto de Computação, UNICAMP, como requisito parcial para a obtenção do título de Mestre em Ciência da Computação.

Substitua pela ficha catalográfica

(Esta página deve ser o verso da página anterior mesmo no caso em que não se imprime frente e verso, i.é., até 100 páginas.)

Substitua pela folha com as assinaturas da banca

Instituto de Computação
Universidade Estadual de Campinas

Comparação algébrica de genomas: o caso da distância de reversão

André Atanasio Maranhão Almeida[1]

23 de Fevereiro de 2007

Banca Examinadora:

- Prof. Dr. João Meidanis (Orientador)
- Profa. Dra. Maria Emília Machado Telles Walter
  Departamento de Ciência da Computação – CIC – UnB
- Prof. Dr. Zanoni Dias
  Instituto de Computação – IC – Unicamp
- Prof. Dr. Flávio Keidi Miyazawa (Suplente)
  Instituto de Computação – IC – Unicamp

[1]**Apoio financeiro do CNPq**

Resumo

Nas últimas décadas presenciamos grandes avanços na biologia molecular que levaram ao acúmulo de um grande volume de dados acerca de moléculas, tais como DNAs e proteínas, essenciais para a vida e para seu entendimento. O estágio atual é de busca por ferramentas que permitam extrair informações com relevância biológica destes dados. Neste contexto, a comparação de genomas surge como uma das ferramentas e nesta categoria incluímos rearranjo de genomas. Em rearranjo, o genoma é representado por uma seqüência de blocos conservados e, dados dois genomas e um conjunto de operações, busca-se pela seqüência mínima de operações que transformem um genoma no outro.

Em 1995, Hannenhalli e Pevzner apresentaram o primeiro algoritmo polinomial para o problema da ordenação por reversões orientadas. Tal algoritmo executa em tempo $O(n^4)$ e foi o primeiro algoritmo polinomial para um modelo realístico de rearranjo de genomas. Desde então, surgiram algoritmos que apresentam desempenho assintoticamente melhor. O melhor deles, apresentado por Tannier e Sagot em 2004, é capaz de executar em tempo $O(n\sqrt{n \log n})$. Há um algoritmo linear, desenvolvido por Bader e colegas [2], mas este não é capaz de determinar a seqüência de reversões, apenas calcula a distância.

Motivado pela carência de uma derivação algébrica mais formal da teoria desenvolvida em rearranjo de genomas, desenvolvemos uma solução formal para o problema da distância de reversão com sinal. Utilizamos, em tal solução, um formalismo algébrico para rearranjo de genomas que relaciona a recente teoria de rearranjo de genomas – basicamente fundamentada no trabalho de Hannenhalli e Pevzner – e a teoria de grupos de permutação de uma nova forma. Pretendemos criar a base para grandes avanços na área através de um formalismo algébrico forte.

Abstract

In the last decades we have seen a great progress in molecular biology. That lead to a large volume of data on molecules, DNA and proteins, essential for life. The current stage of research lies in the pursuit of tools to extract information with biological relevance from this data.

In this context, comparison of genomes is an important tool and genome rearrangements is a way of doing that comparison. In rearrangement analysis the genome is represented by a sequence of conserved blocks. The aim is to find a minimum sequence of operations that transform a genome into another given as input two genomes and a set of allowed operations.

In 1995, Hannenhalli and Pevzner presented the first polinomial algorithm for sorting signed permutations by reversals. This algorithm has complexity $O(n^4)$ in time and was the first polinomial algorithm for a realistic model of genome rearrangement. Since then, new algorithms with better asintotic performance had appeared. The fastest algorithm, with complexity $O(n\sqrt{n \log n})$, was developed by Tannier and Sagot in 2004.

Motivated by a lack of a more formal derivation in the genome rearrangement developed theory, we developed a formal solution for the signed reversal distance problem.

We use an algebraic formalism that relates the recent genome rearrangement theory – basically based on a work of Hannenhalli and Pevzner – to permutation group theory in a new form. We intend to build a solid theoretical base for further advances in the area through strong algebraic formalism.

Agradecimentos

Há muitas pessoas a quem devo meus sinceros agradecimentos.

Agradeço primeiramente à minha família, principalmente aos meus pais (Atanasio e Graça), irmãos (Diogo e Analice), avós (em memória) e tias Denilma, Vitória e Telma, pelo apoio e amor incondicionais, assim como, pela participação ativa em minha formação. Não esquencendo o cunhado, Felipe.

À Taciana, minha esposa, a quem amo e admiro cada vez mais. Obrigado por todo o apoio, carinho e compreensão.

À família que ganhei de presente da Taciana em especial, meus sogros (Zinho e Béba) e cunhado (João Paulo – o feinho). Muito obrigado pelo apoio, confiança, torcida e carinho.

Um agradecimento especial a meu orientador, João, pela grande contribuição na minha evolução acadêmica em todos os aspectos.

Aos membros da banca de defesa de dissertação, Maria Emília (CIC-UnB) e Zanoni, pelas críticas e sugestões que serão de grande utilidade para a evolução deste trabalho e para mim como aprendizado para futuros trabalhos.

Aos membros da banca de minha proposta de dissertação, Zanoni e Pedro Rezende, pelas preciosas sugestões. A todos os professores da graduação (UFAL), em especial a Eliana (minha orientadora de iniciação científica), Cid, Evandro, Chico, Alejandro, Agamenon, Roberta, Eduardo e Denise. A todos os professores do curso técnico (CEFET-AL), em especial a Adolfo, Stênio, Heitor, Tárcio, Mônica, Jailton, Edinaldo e Natalício.

Aos professores do IC com os quais tive a oportunidade de muito aprender (Pannain, Célia, Guido, Rubira, Meidanis e Bauzer).

Aos amigos que encontrei em Campinas. Aos do IC (Leonardo – Caipirinha – e Sânia, Neumar, Tiago Moronte e Graciela, Thiago Coelho e Daniele, Daniel – grande fã de Axé :D –, Bruno, Fernando, Marília, Wylber, André Vignatti, Luiz Bittencourt, Rafael, Allan, Anderson, Carlos Eduardo, Celso, Cleber, Augusto Devegili, Borin, Felipe Klein e Bartho), obrigado pelo muito que aprendi com vocês, pelo apoio, companherismo, conversas fiadas e momentos de lazer. Ao Cleber um agradecimento especial pelas contribuições em meu trabalho. Aos da Scylla (Zé Augusto, Christian, Miguel, Carol, Arnaldo, Carina, Kariza, Zanoni e João) obrigado pelo ambiente de trabalho agradável e pelo muito

que me ensinaram. Christian obrigado pelas contribuições em minha dissertação. Aos de república e convivência: Elvio, Mike, Lin, Sr. Adilson e Sra. Vera. Pela amizade, momentos de descontração e apoio em momentos críticos.

Aos amigos de Maceió: Elthon, Wallace, André (Tiririca), Klebson, Glória, Márcio, Talita, Alan, Bruno (Chukulate), Karl, Glauber (Galubê), Hyggo, Rodrigo (Mosca – um dos grandes responsáveis por eu estar aqui), Marcello (Stress), Leandro, Julian (o Argentino – este é bem diferente da imagem que tinha de Argentino), Xambinho (Alexandre), Fred, Milena, Moacy (ou será que é Atanasio?), Emerson (Memesso), Patrick (um grande amigo que me deu bastante apoio em Campinas), Rinaldo, Rodrigo, Augusto, Romero, Ítalo, Lauro, João Paulo, Jean, Sóstenes, Nilson, Catarine, Joana, Mônica, Gabriela, Fabrício (Satã), Leandro Cerqueira, Benísio, Fábio Maranhão, Lucas, Mário Jucá, Dalmo, Zulmira, Júnior, Diana, Celina e tia Marli.

Ao CNPq pelo apoio financeiro.

Finalmente, o meu muito obrigado a todos aqueles que trabalharam nos bastidores, cujo nome eu desconheço. Um agradecimento especial a todos aqueles que injustamente esqueci de citar o nome. Minhas sinceras desculpas e o meu agradecimento.

Conteúdo

Resumo **vii**

Abstract **ix**

Agradecimentos **xi**

1 Introdução **1**

2 Fundamentos **3**
2.1 Ácidos nucléicos 4
2.2 Proteínas 6
2.3 Genomas 9
2.4 Os mecanismos da genética molecular 13
2.5 Evolução de genomas 15
2.6 Grafos 16

3 Comparação de Genomas **21**
3.1 Diagrama de pontos de quebra 23
3.2 Grafo de entrelaçamento de arcos 26
3.3 Obstáculos 27

4 Grupos de Permutação **29**
4.1 Função 29
4.2 Permutação 31
4.2.1 Ciclos curtos 33
4.2.2 Novas operações 33

5 Um formalismo algébrico para rearranjo de genomas **39**
5.1 Genes e complementaridade 39
5.2 Genomas circulares 40

5.3 Operações . . . 41
5.4 Distância de Reversão . . . 42
5.5 Genomas Lineares . . . 44

6 Uma solução algébrica para o problema da Distância de Reversão 45
6.1 Quociente . . . 45
6.2 Diagrama de pontos de quebra . . . 49
6.2.1 Arcos: orientação e reversões . . . 51
6.2.2 Entrelaçamento de arcos . . . 54
6.3 Grafo de entrelaçamento de arcos $OV(P)$ . . . 58
6.4 Componentes . . . 61
6.4.1 Componentes e componentes conexos de $OV(P)$ . . . 67
6.4.2 Obstáculos . . . 69
6.5 Algoritmo algébrico . . . 73
6.5.1 Componentes Bons . . . 74
6.5.2 Componentes Ruins . . . 75

7 Conclusão 81

A Revisão Bibliográfica 83
A.1 Transforming Cabbage into Turnip: Polynomial Algorithm for Sorting Signed Permutation by Reversals . . . 83
A.1.1 Introdução . . . 83
A.1.2 Diagrama de pontos de quebra para permutações sinalizadas . . . 86
A.1.3 Grafo de entrelaçamento e obstáculos . . . 87
A.1.4 Transformações de equivalência em permutações . . . 89
A.1.5 Reversões seguras em componentes orientados . . . 91
A.1.6 Eliminando obstáculos . . . 92
A.1.7 Fortalezas . . . 93
A.1.8 Algoritmo Polinomial . . . 94
A.2 A Faster and Simpler Algorithm for Sorting Signed Permutations by Reversals 95
A.2.1 Introdução . . . 95
A.3 A Very Elementary Presentation of the Hannenhalli-Pevzner Theory . . . 96

Bibliografia 103

Lista de Tabelas

2.1 Lista de cromossomos humanos com o respectivo tamanho da molécula [21]. 8
2.2 Os vinte aminoácidos comumente encontrados nas proteínas [26]. 10
2.3 Lista de organismos juntamente com o número de cromossomos de seu genoma [8]. 11
2.4 Lista de organismos com o respectivo tamanho de seu genoma [8]. 12
2.5 O código genético padrão de mapeamento de códons em aminoácidos [26]. . 14

Lista de Figuras

2.1 Estrutura molecular esquemática de uma molécula de ribose [26]. 4
2.2 Estrutura molecular esquemática dos quatro tipos de bases nitrogenadas presentes nas moléculas de DNA, assim como a forma com que as bases nitrogenadas se unem para compor a dupla fita [26]. 5
2.3 Estrutura molecular esquemática da uracila [8]. 5
2.4 Estrutura molecular esquemática de uma molécula de 2′-desoxirribose [26]. 6
2.5 Figura ilustrativa de uma molécula de DNA. Tal ilustração foi extraída do livro *Genomes* [8] e editada por Christian Baudet. 7
2.6 Estrutura molecular esquemática de um aminoácido [26]. 9
2.7 Figura ilustrativa de uma proteína. Tal ilustração foi obtida através do *software* `Cn3D`, que é de autoria do NCBI [21]. 9
2.8 Fluxo da informação genética em uma célula: dogma central da biologia molecular [26]. 14
2.9 Um exemplo de grafo. 16
2.10 Exemplo de grafos complementares. 17
2.11 Exemplo de caminho com quatro vértices. 17
2.12 Exemplo de subgrafo do grafo apresentado na Figura 2.10 (b). 18
2.13 Subgrafo, do grafo apresentado na Figura 2.12, induzido por $T = \{a, b, c, e\}$. 19
2.14 Exemplo de grafo de cordas sobrepostas 19

3.1 Abstração de um genoma como uma seqüência de genes. 22
3.2 Seqüência mínima de operações que transforma π em σ. 22
3.3 Diagrama de pontos de quebra de uma permutação sem sinais. 24
3.4 Diagrama de pontos de quebra de uma permutação com sinais. 25
3.5 Grafo de entrelaçamento de arcos 26
3.6 Grafo de entrelaçamento de arcos depois de aplicada a reversão. 27
3.7 Diagrama de pontos quebra após a aplicação da reversão induzida pelo arco $(16, 17)$. 27

4.1 Função injetora $f : X \mapsto Y$. 30

4.2 Função sobrejetora $f : X \mapsto Y$. 30
4.3 Função bijetora $f : X \mapsto Y$. 31

5.1 Mapeamento linear $\mapsto$ circular. 44

6.1 Diagrama de pontos de quebra do problema P_2. 50
6.2 Exemplos de grafos de entrelaçamento de arcos. a) Apresenta o grafo de entrelaçamento de arcos do problema P_2. b) Apresenta o grafo de entrelaçamento de arcos do problema $P_2\rho_1$, onde $\rho_1 = \rho(+0, -1, \pi_2)$. c) Apresenta o grafo de entrelaçamento de arcos do problema $P_2\rho_1\rho_2$, onde $\rho_2 = \rho(+2, -3, \pi_2\rho_1)$. d) Apresenta o grafo de entrelaçamento de arcos de $P_2\rho_1\rho_2\rho_3$, onde $\rho_3 = \rho(+3, -0, \pi_2\rho_1\rho_2)$. 59
6.3 Grafo de entrelaçamento de arcos do problema $P_2\rho_4$, onde $\rho_4 = \rho(+1, -2, \pi_2)$. 60
6.4 Modelo associado ao grafo de cordas sobrepostas apresentado na Figura 6.2 (a). 60
6.5 Efeito da reversão $\rho = (a, b, \pi)$ no modelo associado ao grafo de cordas sobrepostas. 61
6.6 Ilustra um possível diagrama de decomposição em componentes de P_1 . . . 64
6.7 Estudo de caso para determinação dos possíveis posicionamentos de um componente ruim (R) em relação aos obstáculos (1 e 2) em um problema P com $h(P) = 2$. 72
6.8 Estudo de caso para determinação dos possíveis posicionamentos de um componente ruim (R) em relação aos obstáculos (1 e 2) em um problema P com $h(P) = 2$. 72
6.9 Estudo de caso para determinação dos possíveis posicionamentos de componentes ruins em relação aos obstáculos (1, 2 e 3) em um problema P com $h(P) = 3$. 76
6.10 Estudo de caso para determinação dos possíveis posicionamentos de um componente ruim (R1) em relação aos obstáculos (1, 2 e 3) em um problema P com $h(P) = 3$ e $\overline{h}(P) = 1$. 78
6.11 Estudo de caso para determinação dos possíveis posicionamentos de componentes ruins (R1, R2 e R3) em relação aos obstáculos (1, 2 e 3) em um problema P com $h(P) = 3$ e com dois ou três componentes ruins. 79

A.1 Diagrama de pontos de quebra de uma permutação não sinalizada. 85
A.2 Diagrama de pontos de quebra de uma permutação sinalizada. 86
A.3 Grafo de entrelacamento 88
A.4 Exemplo de obstáculo. Note que o componente não orientado U não é um obstáculo. 88

A.5 Efeito de uma reversão, que atua em um ciclo orientado, sob o diagrama de pontos de quebra. 91
A.6 Efeito de uma reversão, que atua em um ciclo orientado, sob o grafo de entrelaçamento. 92
A.7 Diagrama de pontos de quebra de π em relação a identidade. 98
A.8 Grafo de entrelaçamento de arcos. 99
A.9 Grafo de entrelaçamento de arcos resultante da aplicação da reversão induzida pelo vértice $(4, 5)$. 100
A.10 Diagrama de pontos de quebra resultante da aplicação da reversão induzida pelo vértice $(4, 5)$. 100
A.11 Diagrama de pontos de quebra de π em relação a identidade. 101
A.12 Grafo de entrelaçamento de arcos. 102

Capítulo 1

Introdução

Com a elucidação da estrutura do DNA, o decorrente desenvolvimento das técnicas de manipulação genética e, por conseguinte, das técnicas de seqüenciamento atuais, presenciou-se, e ainda é notório, um grande crescimento na quantidade de dados moleculares, principalmente em seqüências de DNA e proteína. Os esforços agora concentram-se na busca por informações biologicamente interessantes deste grande volume de dados.

Uma das formas de se buscar por estas informações é através da comparação de genomas, onde são analisadas semelhanças entre genomas de espécies distintas. Há diversas formas de comparar genomas, sendo a comparação a nível de seqüências de DNA [27] uma forma comumente utilizada.

Uma outra forma de compará-los é através de análises de rearranjo de genomas. Nesta, o genoma é visto como uma lista ordenada de blocos, onde um bloco representa uma seqüência conservada de DNA que pode estar associada a um gene ou um outro marcador qualquer. Dados dois genomas e um conjunto de operações para manipular os genomas, busca-se pela seqüência mínima de operações que transformem um genoma no outro.

Esta abordagem computacional baseada na comparação da ordem de genes foi inicialmente estudada por Sankoff [23, 24, 25] e ganhou destaque nas últimas décadas [2, 4, 5, 6, 7, 11, 13, 14, 17, 33]. Uma de suas possíveis aplicações está na construção de filogenias [26].

Há diversos tipos de operação, dentre eles podemos citar troca de blocos, transposição, reversão, fusão e fissão. Variando o conjunto de operações temos diversos problemas em rearranjo de genomas. Neste trabalho nos limitaremos às reversões.

Hannenhalli e Pevzner [11] apresentaram o primeiro algoritmo polinomial para o problema da ordenação por reversões orientadas. Tal algoritmo executa em tempo $O(n^4)$ e foi o primeiro algoritmo polinomial para um modelo realístico de rearranjo de genomas.

Desde então, surgiram algoritmos que apresentam desempenho assintoticamente melhor. Berman e Hannenhalli [7] conseguiram um algoritmo que executa em tempo $O(n^3)$.

Kaplan, Shamir e Tarjan [13] atingiram $O(n^2)$. Bader, Moret e Yan [2] apresentaram um algoritmo que executa em tempo linear. É importante ressaltar que este último calcula apenas a distância, não sendo capaz de determinar a seqüência de reversões. Recentemente, Tannier e Sagot [29] apresentaram um algoritmo que executa em tempo $O(n\sqrt{n \log n})$.

Este trabalho é motivado pela carência de uma derivação algébrica mais formal da teoria desenvolvida em rearranjo de genomas. Utilizando um formalismo algébrico para rearranjo de genomas, que relaciona a teoria clássica de rearranjo de genomas – basicamente fundamentada no trabalho de Hannenhalli e Pevzner [11] – e a teoria de grupos de permutação de uma nova forma, desenvolveremos uma solução formal para o problema da distância de reversão com sinal. A pretensão é criar a base para grandes avanços na área através de um formalismo algébrico forte.

Alcançamos diversos avanços em nosso trabalho. Fizemos alterações nas definições do formalismo objetivando compatibilizar as definições algébricas e clássicas. Alteramos o conceito de entrelaçamento de arcos, que o tornou de mais simples visualização e manipulação. Foram definidos vários conceitos. Definimos grafo de entrelaçamento de arcos, definimos decomposição de um problema. Apresentamos o esboço de uma demonstração que correlaciona componentes oriundos da decomposição e os componentes conexos do grafo de entrelaçamento de arcos. Apresentamos uma demonstração do efeito de uma reversão no grafo de entrelaçamento de arcos. Por fim apresentamos um algoritmo que resolve o problema da distância de reversão e um esboço da demonstração de sua corretude.

No Capítulo 2 apresentamos conceitos fundamentais no contexto deste trabalho. São apresentados conceitos na área de biologia computacional e de teoria dos grafos. A teoria clássica de rearranjo de genomas e, mais aprofundadamente, de distância de reversão é apresentada no Capítulo 3. No Capítulo 4 apresentamos conceitos e definições acerca de permutações, que serão a base do formalismo que utilizaremos na solução do problema. Tal formalismo é apresentado no Capítulo 5. No Capítulo 6 apresentamos o esboço da solução formal para o problema da distância de reversão. Finalmente, apresentamos a conclusão e trabalhos futuros no Capítulo 7. Ainda temos um apêndice, onde são resumidos os trabalhos utilizados na realização deste trabalho.

Capítulo 2

Fundamentos

Neste capítulo iremos apresentar os conceitos fundamentais da biologia molecular e da computação, que serão empregados ao longo deste trabalho. Maiores informações podem ser obtidas na seguinte bibliografia [8, 26].

Na natureza existem coisas vivas e não vivas. As coisas vivas são caracterizadas por uma participação ativa em seu ambiente. Já as coisas não vivas são caracterizadas por um estado de equilíbrio com o ambiente que as cerca. Um organismo vivo constantemente troca matéria e energia com outros organismos ao seu redor. Apesar de compostas – as coisas vivas ou não – por elementos semelhantes, no interior dos organismos vivos ocorrem diversas e complexas reações químicas, que os mantêm vivos.

Chamamos a atenção para duas formas de vida que podem permanecer completamente inativas por longos períodos de tempo e ainda sim estarem vivas. São elas as sementes e os vírus. Uma regra em biologia molecular é que não há regras sem exceções.

Todos os organismos, desde a mais simples forma de vida até a mais complexa, possuem uma química molecular – também conhecida por bioquímica – semelhante. Nesta química, nosso foco volta-se para dois tipos de moléculas: ácidos nucléicos e proteínas. Os ácidos nucléicos fornecem as informações necessárias para produzir as proteínas. Estes são, também, os responsáveis por passar estas informações para as gerações subseqüentes e, assim, dão origem a hereditariedade e aos mecanismos que permitem a evolução das espécies. As proteínas constituem a maioria das substâncias nos organismos vivos. Podem ser de diversos tipos e possuem diversas funções [26].

Na Seção 2.1 descrevemos os ácidos nucléicos. As proteínas são descritas na Seção 2.2. Na Seção 2.3 apresentamos como os ácidos nucléicos estão dispostos nos organismos. Descrevemos o mecanismo pelo qual as proteínas são construídas na Seção 2.4. Na Seção 2.5 descrevemos as maneiras pelas quais os genomas evoluem. Finalmente, as definições e conceitos relativos a grafos estão na Seção 2.6.

2.1 Ácidos nucléicos

Um dos tipos de ácido nucléico presente nos organismos vivos é o ácido ribonucléico, também conhecido por RNA. O RNA é uma molécula composta por uma cadeia de moléculas mais simples, chamadas nucleotídeos, que são formadas por uma molécula de açúcar, um resíduo fosfato e uma base nitrogenada. No RNA a molécula de açúcar é a ribose [26].

A ribose possui cinco átomos de carbono que são rotulados de $1'$ a $5'$ e dispostos conforme apresentamos na Figura 2.1. A cadeia de nucleotídeos é formada da seguinte forma: o carbono $3'$ liga-se ao resíduo fosfato e este, por sua vez, liga-se ao carbono $5'$ de outro nucleotídeo. Tal composição permite-nos definir uma orientação. Convencionou-se que a molécula de RNA inicia-se na extremidade que tem o carbono $5'$ livre e acaba na extremidade que tem o carbono $3'$ livre.

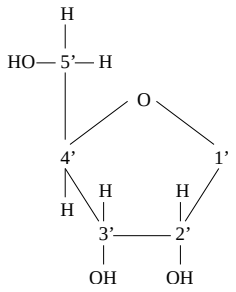


Figura 2.1: Estrutura molecular esquemática de uma molécula de ribose [26].

Na biologia molecular há cinco tipos de bases nitrogenadas, que chamaremos simplesmente de bases. São elas: adenina (A), guanina (G), citosina (C), uracila (U) e timina (T). Timina não aparece em moléculas de RNA. Nas Figuras 2.2 e 2.3 apresentamos a estrutura molecular esquemática de cada base. As bases A e G pertencem a um grupo de substâncias chamado purinas e as bases C, T e U pertencem ao grupo das pirimidinas. As bases nitrogenadas ligam-se ao açúcar no carbono $1'$.

Existem diversos tipos de RNA nos organismos vivos. Cada tipo possui uma função distinta. Os diversos tipos serão apresentados, juntamente com suas funções, na Seção 2.4.

O outro tipo de ácido nucléico presente nos organismos vivos é o ácido desoxirribonucléico, também conhecido por DNA. As moléculas de DNA são semelhantes às moléculas de RNA, porém observamos quatro diferenças.

A primeira das diferença está na molécula de açúcar. No DNA, o açúcar presente é o $2'$-desoxirribose, que tem sua estrutura molecular esquemática apresentada na Figura 2.4. A segunda diferença está na substituição das uracilas por timinas.

A terceira diferença está no fato de que, as moléculas de DNA, possuem dupla fita. As fitas são associadas de uma maneira que a molécula toma uma forma de dupla hélice

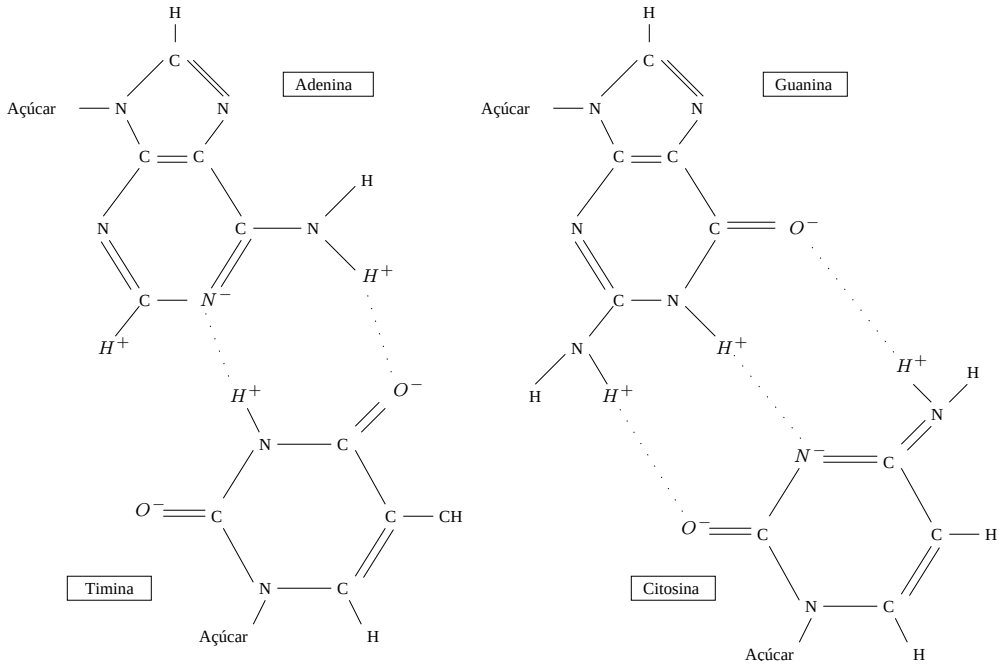


Figura 2.2: Estrutura molecular esquemática dos quatro tipos de bases nitrogenadas presentes nas moléculas de DNA, assim como a forma com que as bases nitrogenadas se unem para compor a dupla fita [26].

conforme apresentamos mais adinte na Figura 2.5. Possuindo uma das fitas é possível calcular a outra fita. Uma base A sempre é pareada com uma base T ou, em outras palavras, A é o complemento de T e vice-versa ou, ainda, A e T são bases complementares. As bases C e G também são complementares. Seja $f_1 = TACGAC$ uma das fitas de uma molécula de DNA. Podemos obter sua fita complementar f_2 através do complemento reverso de f_1: primeiro invertemos a ordem de f_1 e obtemos $f_1' = CAGCAT$, depois trocamos cada base por seu complemento obtendo assim $\overline{f_1} = GTCGTA$. Comumente uma molécula de DNA é simplesmente definida por sua seqüência de bases em uma das

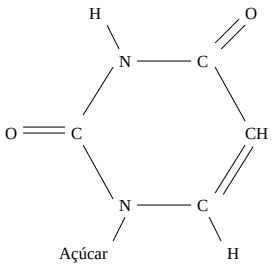


Figura 2.3: Estrutura molecular esquemática da uracila [8].

fitas na direção canônica $5' \mapsto 3'$. Na Figura 2.2 apresentamos uma esquema de como se dão as ligações, através de pontes de hidrogênio, entre as bases complementares. Na Figura 2.5 podemos visualizar a ilustração de uma molécula de DNA.

A última das diferenças está no fato de que o DNA possui, essencialmente, uma função, diferentemente do RNA, que possui diversas funções.

As diferenças apresentadas pelas moléculas de DNA, em relação às moléculas de RNA, as tornam mais estáveis e levam a uma menor variedade de estruturas tridimensionais.

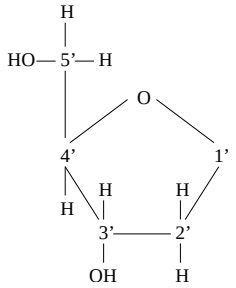


Figura 2.4: Estrutura molecular esquemática de uma molécula de $2'$-desoxirribose [26].

O tamanho das moléculas de DNA é normalmente definido em número de pares de bases, que denotamos por *bp*. Na natureza, moléculas de DNA são muito compridas. Em células humanas, as moléculas de DNA possuem dezenas ou centenas de milhares de pares de bases. Na Tabela 2.1 apresentamos a lista dos cromossomos humanos juntamente com o tamanho das moléculas.

Em organismos cujas células não possuem núcleo, o DNA é encontrado flutuando livremente dentro da célula. Nos outros organismos está delimitado pelo núcleo e por organelas celulares chamadas mitocôndrias – encontradas em animais e plantas – e cloroplastos – encontrados apenas em plantas.

2.2 Proteínas

As proteínas encontradas nos organismos vivos são de diversos tipos e podem assumir diversas funções. Elas podem, por exemplo, atuar como estruturas em diversas partes do organismo ou catalizadoras de reações químicas (enzimas).

Da mesma forma que os ácidos nucléicos, as proteínas são moléculas compostas por uma cadeia de moléculas mais simples. Os aminoácidos são estas moléculas mais simples presentes nas proteínas. Todo aminoácido é composto por um átomo central de carbono – conhecido como carbono alpha ou C_α – e associado a este temos um átomo de hidrogênio, um grupo amina (NH_2), um grupo carboxi ($COOH$) e uma cadeia lateral. Distinguimos

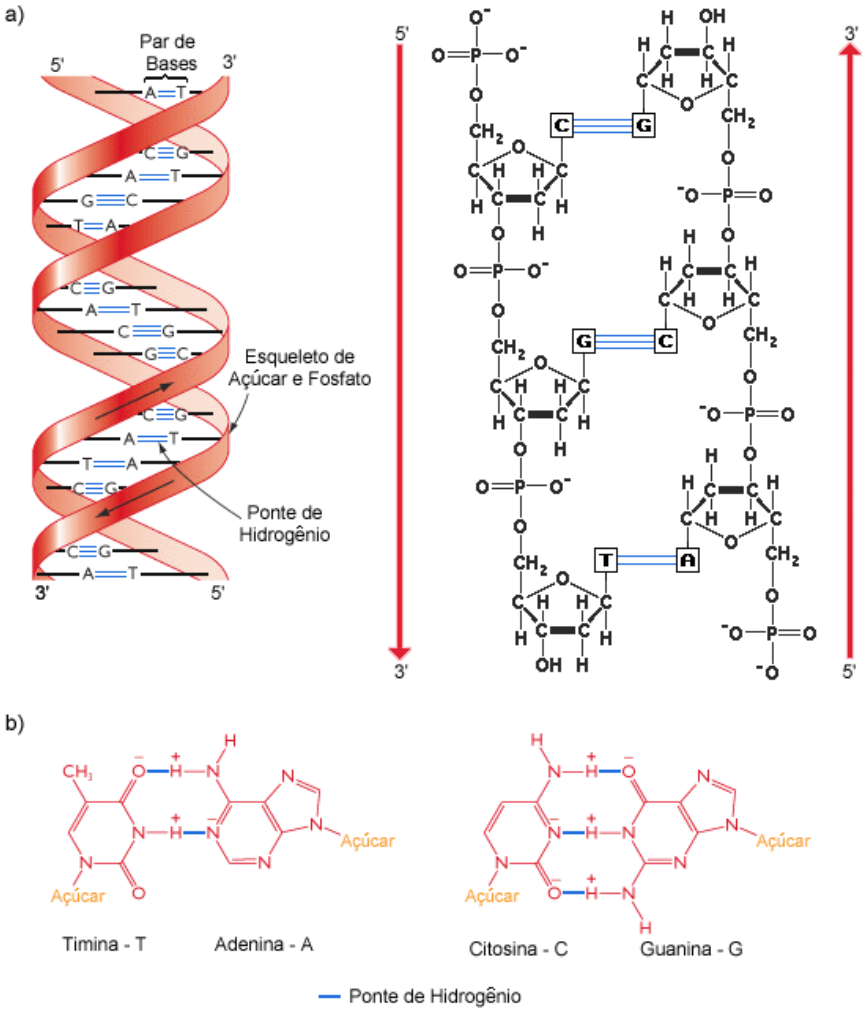


Figura 2.5: Figura ilustrativa de uma molécula de DNA. Tal ilustração foi extraída do livro *Genomes* [8] e editada por Christian Baudet.

um aminoácido dos outros por sua cadeia lateral [26]. Na Figura 2.6 apresentamos uma estrutura molecular esquemática de um aminoácido. Na natureza, encontramos comumente 20 diferentes aminoácidos. Listamos tais aminoácidos na Tabela 2.2.

Nas proteínas, os aminoácidos são conectados através de ligações peptídicas ou, em outras palavras, o átomo de carbono do grupo carboxi de um aminoácido liga-se ao átomo de nitrogênio do grupo amina de outro aminoácido e uma molécula de água é liberada [26]. Uma proteína é composta por cadeias polipeptídicas.

As ligações peptídicas apresentadas pelas proteínas formam um esqueleto com a repetição de um bloco básico $-N-C_\alpha-(CO)-$. Da mesma forma que nos ácidos nucléicos, podemos dizer que há uma orientação nas proteínas. Convencionou-se que a cadeia inicia

Cromossomo	**Tamanho (Mbp)**
1	247,0
2	243,0
3	200,0
4	191,0
5	181,0
6	171,0
7	159,0
8	146,0
9	140,0
10	135,0
11	134,0
12	132,0
13	114,0
14	106,0
15	100,0
16	89,0
17	79,0
18	76,0
19	64,0
20	62,0
21	46,9
22	50,0
X	155,0
Y	58,0

Tabela 2.1: Lista de cromossomos humanos com o respectivo tamanho da molécula [21].

no grupo amina livre e termina no grupo carboxi livre.

O tamanho de uma proteína pode ser mensurado pelo número de aminoácidos. Trezentos aminoácidos é o tamanho médio apresentado pelas proteínas, mas seu tamanho pode variar bastante podendo possuir 100 ou 5000 aminoácidos.

Quando estudamos uma proteína o foco concentra-se na descoberta de sua função. Para se atingir tal objetivo a determinação de sua forma pode ser uma etapa a ser executada. Partindo de sua seqüência de aminoácidos, que é conhecida como estrutura primária, observamos como esta cadeia dobra-se dando origem a uma estrutura tridimensional conhecida por estrutura secundária. Estas estruturas (secundárias), por sua vez, agrupam-se formando a estrutura terciária que, finalmente, agrupa-se a estruturas terciárias de diferentes proteínas e origina a estrutura quaternária. Na estrutura secundária há definição de estruturas tais como: hélices e folhas. Na estrutura terciária começamos a observar a posição tridimensional de cada átomo. Na estrutura quaternária observamos como se

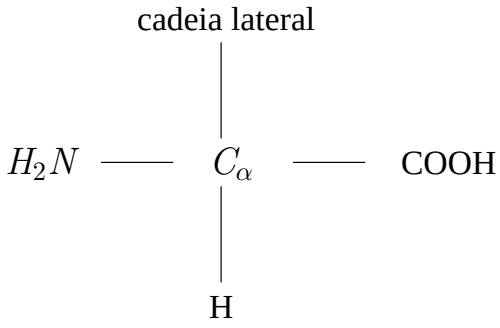


Figura 2.6: Estrutura molecular esquemática de um aminoácido [26].

ligam as subunidades, caso a proteína seja composta de mais que uma subunidade. Na Figura 2.7 apresentamos a ilustração de uma proteína.

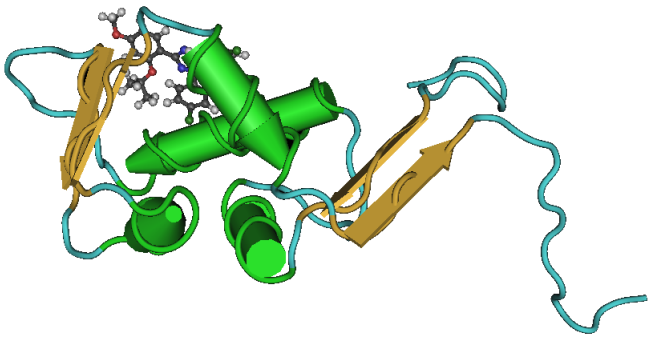


Figura 2.7: Figura ilustrativa de uma proteína. Tal ilustração foi obtida através do *software* `Cn3D`, que é de autoria do NCBI [21].

2.3 Genomas

O conjunto de moléculas de ácido nucléico que definem todas as proteínas de um organismo é conhecido como genoma. Cada molécula do genoma recebe o nome de cromossomo. O genoma da grande maioria das formas de vida é composto de moléculas de DNA. Genomas definidos através de moléculas de RNA só aparecem em vírus [8].

Podemos classificar os organismos vivos em procariotos ou eucariotos. Chamamos eucariotos os organismos cujas células possuem compartimentos limitados por membrana. Dentre estes compartimentos podemos citar um núcleo e organelas como mitocôndrias e cloroplastos (apenas em plantas). São classificados como eucariotos organismos dos reinos dos animais, plantas, fungos e protozoários. Já as células de organismos procariotos são caracterizadas pela ausência de compartimentos internos. Há dois distintos grupos de

Código uma-letra	**Código três-letras**	**Nome**
A	Ala	Alanina
C	Cys	Cisteina
D	Asp	Ácido aspártico
E	Glu	Ácido glutâmico
F	Phe	Fenilalanina
G	Gly	Glicina
H	His	Histidina
I	Ile	Isoleucina
K	Lys	Lisina
L	Leu	Leucina
M	Met	Metionina
N	Asn	Asparagina
P	Pro	Prolina
Q	Gln	Glutamina
R	Arg	Arginina
S	Ser	Serina
T	Thr	Treonina
V	Val	Valina
W	Trp	Triptofan
Y	Tyr	Tirosina

Tabela 2.2: Os vinte aminoácidos comumente encontrados nas proteínas [26].

procariotos, são eles as bactérias e arqueobactérias.

No grupo das bactérias estão os procariotos encontrados mais comumente, tais como *Escherichia coli*, *Bacillus subtilis* e cianobactéria.

Uma molécula de DNA pode ser linear ou circular. Todos os genomas nucleares de eucariotos que têm sido estudados são divididos em duas ou mais moléculas de DNA linear. Todo eucarioto também possui um genoma mitocondrial, que é menor que o nuclear e, normalmente, é circular. Em plantas e outros organismos fotossintéticos há um terceiro genoma localizado no cloroplasto [8]. Na Tabela 2.3 apresentamos uma lista de organismos juntamente com o número de cromossomos em seu genoma nuclear.

Doravante quando nos referirmos a um genoma estaremos na verdade nos referindo ao genoma nuclear, que é o mais estudado nas pesquisas e o que define a grande maioria das proteínas. A estrutura física básica de genomas de organismos eucariotos são semelhantes, porém os genomas variam bastante no tamanho. O menor genoma de um eucarioto possui tamanho de aproximadamente 10 Mbp e o maior é maior que 100.000 Mbp. Na Tabela 2.4

Organismo	Cromossomos
Saccharomyces cerevisiae	16
Drosophila melanogaster	4
Homem	23
Milho	10

Tabela 2.3: Lista de organismos juntamente com o número de cromossomos de seu genoma [8].

apresentamos uma lista de organismos juntamente com o tamanho de seu genoma.

Conforme a Tabela 2.4 podemos observar que, a grosso modo, o genoma cresce de acordo com a complexidade do genoma, os menores genomas sendo observados entre organismos do reino dos fungos e os maiores entre organismos superiores, tais como vertebrados e plantas com flores. Isto faz sentido se pensarmos que quanto mais complexo for o organismo mais proteínas precisa codificar. É importante observar que esta relação não segue um critério de proporcionalidade. Por exemplo, o genoma da levedura *Saccharomyces cerevisiae* possui um genoma 250 vezes menor que o genoma humano. Como o genoma humano possui aproximadamente 40.000 genes, era de se esperar que a levedura possuisse algo em torno de 160 genes. Na realidade, a *Saccharomyces cerevisiae* possui aproximadamente 6.000 genes. Tal fato ocorre porque nos genomas de organismos de menor complexidade há uma otimização do espaço. Os genes estão mais próximos uns dos outros e os genes estão dispostos de uma forma mais compacta.

Os genes nem sempre são compostos por um trecho contínuo de DNA. Podem existir trechos intercalados, que serão descartados para a produção da proteína. Os trechos que são utilizados para a produção da proteína são chamados de exons e os outros de íntrons. Quando dizemos que os genes estão dispostos de uma forma mais compacta queremos dizer que há uma menor presença de íntrons. Além dos genes, com seus exons e íntrons, há regiões reguladoras e regiões intergênicas. As regiões reguladoras são responsáveis por transmitir informações aos mecanismos celulares. As regiões intergênicas não possuem uma função bem conhecida. Dentre as informações que uma região reguladora transmite podemos citar "presença de um gene mais a frente". Quando dizemos que os genes estão mais próximos queremos dizer que há menor presença de regiões intergênicas. Menos de 10% do genoma humano está relacionado a geração de proteínas.

Os genomas em procariotos são bem diferentes dos genomas em eucariotos. Em regra geral genomas em procariotos são muito menores que os genomas em eucariotos. Por exemplo, a *Escherichia coli* possui um genoma com apenas 4.639 Kbp, ou seja, dois quintos do tamanho do genoma da levedura.

Organismo	**Tamanho do genoma (Mbp)**
Procariotos	
Mycoplasma genitalium	0,58
Escherichia coli	4,64
Bacillus megaterium	30,00
Eucariotos	
Fungos	
Saccharomyces cerevisiae (levedura)	12,10
Aspergillus nidulans	25,40
Protozoários	
Tetrahymena pyriformis	190,00
Invertebrados	
Caenorhabditis elegans (verme)	100,00
Drosophila melanogaster (mosca de fruta)	140,00
Bombyx mori (bicho da seda)	490,00
Strongylocentrotus purpuratus (ouriço do mar)	845,00
Locusta migratoria (gafanhoto)	5.000,00
Vertebrados	
Fugu rubripes (peixe)	400,00
Homo sapiens (homem)	3.000,00
Mus musculus (rato)	3.300,00
Plantas	
Arabidopsis thaliana	100,00
Oryza sativa (arroz)	565,00
Pisum sativum (ervilha)	4.800,00
Zea mays (milho)	5.000,00
Triticum aestivum (trigo)	17.000,00
Fritillaria assyriaca	120.000,00

Tabela 2.4: Lista de organismos com o respectivo tamanho de seu genoma [8].

A disposição física difere bastante também. A grande maioria dos genomas (em procariotos) está contido em uma única molécula de DNA e esta é circular ao invés de linear. Além disso, os cromossomos de eucariotos geralmente aparecem em duas (para organismos diplóides) ou várias cópias (para poliplóides). Os humanos, por exemplo, são organismos diplóides. Nosso genoma é composto por um total de 46 cromossomos, com cada um dos pais contribuindo com metade dos cromossomos. Cada um dos pais contribui com uma cópia dos comossomos 1 a 22 e um dos cromossomos sexuais (X ou Y). A combinação XY dá origem a um macho e a combinação XX dá origem a uma fêmea. A cana-de-açúcar é um exemplo de organismo poliplóide.

Além de seu único cromossomo, em procariotos são também encontrados genes adicionais em moléculas circulares independentes e menores chamadas plasmídeos. Em tais porções de genoma são codificadas propriedades tais como resistência a antibióticos.

O genoma em procariotos, em regra geral, possui também menor número de genes. Por exemplo, *Escherichia coli* possui 4.397 genes.

2.4 Os mecanismos da genética molecular

Cada célula – com exceção das células germinativas – de um organismo possui uma cópia de todo o genoma, que chamaremos simplesmente por DNA. No DNA há trechos que codificam as informações necessárias para construir uma dada proteína, tais trechos são conhecidos por genes.

Nas células existem mecanismos capazes de reconhecer onde começam e terminam os genes. Com o auxílio de uma enzima chamada transcriptase, é feita uma cópia do gene em uma molécula de RNA que recebe o nome de RNA mensageiro ou mRNA. Durante tal processo, chamado de transcrição, são substituídas as bases T por U.

Posteriormente esta molécula de mRNA irá migrar até uma organela celular chamada ribossomo, que é basicamente composta de proteínas e um tipo de RNA chamado RNA ribossomal ou rRNA. Esta organela tem como função a síntese protéica. Usando como entrada uma molécula de mRNA e moléculas de um tipo de RNA chamado de RNA transportador ou tRNA, o ribossomo produz uma proteína.

Agrupando as bases nitrogenadas três a três, formando o que chamamos de códons, os ribossomos promoverão a ligação destes códons a moléculas de tRNA que, por sua vez, estão ligadas a um aminoácido, cada uma. Na Tabela 2.5 temos o código genético padrão de conversão de códons em aminoácidos. Nem todos os organismos utilizam exatamente este código, mas o procedimento é semelhante para todos os organismos. Desta forma a cadeia de aminoácidos vai sendo formada. Ao final do processo, as moléculas de mRNA e os tRNAs são liberados para degradação e futura utilização. Temos também a proteína montada. Este processo é chamado de tradução.

Primeira posição	Segunda posição				Terceira posição
	G	A	C	U	
G	Gly	Glu	Ala	Val	G
	Gly	Glu	Ala	Val	A
	Gly	Asp	Ala	Val	C
	Gly	Asp	Ala	Val	U
A	Arg	Lys	Thr	Met	G
	Arg	Lys	Thr	Ile	A
	Ser	Asn	Thr	Ile	C
	Ser	Asn	Thr	Ile	U
C	Arg	Gln	Pro	Leu	G
	Arg	Gln	Pro	Leu	A
	Arg	His	Pro	Leu	C
	Arg	His	Pro	Leu	U
U	Trp	STOP	Ser	Leu	G
	STOP	STOP	Ser	Leu	A
	Cys	Tyr	Ser	Phe	C
	Cys	Tyr	Ser	Phe	U

Tabela 2.5: O código genético padrão de mapeamento de códons em aminoácidos [26].

Observe que há, na Tabela 2.5, três ocorrências do chamado `STOP` códon. É através deste códon que os mecanismos celulares reconhecem o final de um gene. O gene sempre inicia com uma metionina (AUG).

Há uma enzima que realiza a operação inversa da transcriptase gerando o que chamamos de cDNA, um DNA composto apenas de exons. Chamamos tal enzima de transcriptase reversa. O DNA composto de exons e íntrons é chamado de DNA genômico. Na Figura 2.8 apresentamos o fluxo da informação genética na célula.

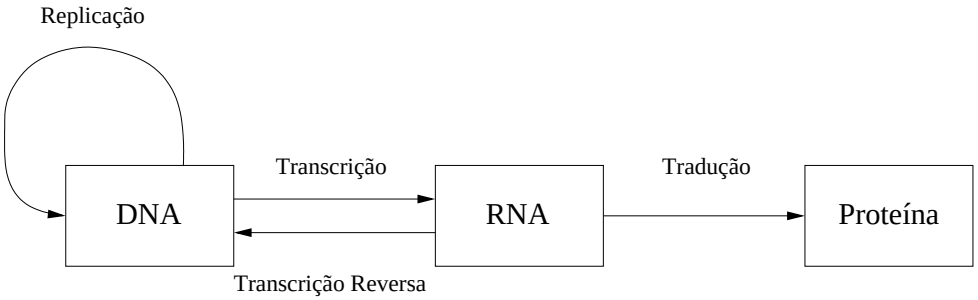


Figura 2.8: Fluxo da informação genética em uma célula: dogma central da biologia molecular [26].

2.5 Evolução de genomas

Genomas são entidades dinâmicas que evoluem através do efeito acumulado de eventos de mutação ou recombinação. As mutações são responsáveis por alterações em pequena escala, já as recombinações são responsáveis por alterações em maior escala, que dão origem ao que chamamos de rearranjo [8].

Uma mutação é uma alteração de uma pequena região na seqüência de nucleotídeos de um genoma. Muitas mutações realizam apenas a alteração de apenas um nucleotídeo. Outras promovem a inserção ou remoção de um ou alguns (poucos) nucleotídeos. As mutações são resultado de erros no processo de replicação do DNA ou do efeito de fatores mutagênicos (químicos ou físicos).

Recombinações promovem a reestruturação de parte do genoma. Tais eventos podem trocar segmentos de cromossomos homólogos durante o processo de meiose ou alterar a posição de uma porção do genoma.

Nos organismos diplóides e poliplóides podem existir dois tipos de células, que chamomos de somáticas e germinativas. As células somáticas possuem o genoma completo. As células germinativas possuem apenas metade do genoma e têm a função de reprodução.

Mutações e recombinações podem ter efeitos importantes na célula em que ocorrem, que podem causar um defeito tão grave que a leve à morte ou podem promover alterações em suas capacidades bioquímicas. Alterações que não resultem na morte da célula têm o potencial de contribuir com a evolução do genoma. Mas para que isto aconteça, estas alterações devem ser herdadas quando o organismo se reproduzir. Em organismos unicelulares, as alterações no genoma são sempre herdadas. Já nos organismos multicelulares, apenas os eventos que ocorrem nas células germinativas têm relevância para a evolução genômica. Observe que alterações no genoma de células somáticas podem afetar a saúde do organismo, por exemplo provocando um câncer.

Mutações nem sempre provocam efeitos perceptíveis, por exemplo: se a alteração ocorrer em regiões intergênicas ou se o códon modificado codifica o mesmo aminoácido como apresentamos a seguir:

AUA	GGA	ACG	CGC	CCU	UAC
I	G	T	R	P	Y

AUA	GGA	$AC\mathbf{C}$	CGC	$CC\mathbf{G}$	UAC
I	G	T	R	P	Y

Outras mutações que normalmente não apresentam efeitos perceptíveis são inserções ou deleções de três (ou múltiplo de três) nucleotídeos, que alteram um ou poucos aminoácidos. É importante observar que mesmo inserindo ou removendo apenas um aminoácido na seqüência a proteína pode perder propriedades.

AUA	GGA	ACG	CGC	CCU	UAC
I	G	T	R	P	Y

AUA	GGA	ACG	CGC	CCU	**UCG**	UAC
I	G	T	R	P	**S**	Y

Outras inserções ou remoções podem ter como efeito a alteração completa de porção da seqüência de aminoácidos.

AUA	GGA	ACG	CGA	CCU	UAC
I	G	T	R	P	Y

AUA	GGA	AC**U**	GCG	ACC	UUA	C
I	G	T	**A**	**T**	**L**	

Eventos de recombinação podem promover a troca de longas porções semelhantes em cromossomos homólogos, que recebe o nome de *crossing-over*. Podem também transferir segmentos de DNA de uma posição para outra.

2.6 Grafos

Um **grafo** G é uma tripla consistindo de um conjunto de **vértices** $V(G)$, um conjunto de **arestas** $E(G)$ e uma relação que associa cada aresta a dois vértices (não necessariamente distintos), chamados **extremos** [32]. Podemos desenhar um grafo representando vértices por pontos ou círculos e as arestas são representadas por linhas que ligam seus extremos. Na Figura 2.9 apresentamos um grafo com vértices $\{a, b, c\}$ e arestas $\{e1, e2, e3, e4, e5\}$.

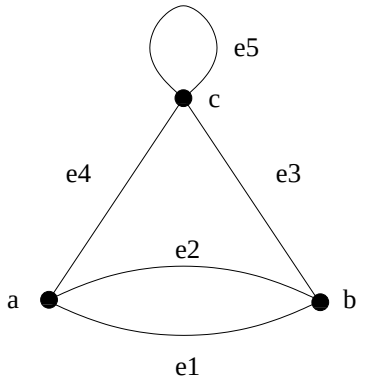


Figura 2.9: Um exemplo de grafo.

Uma aresta cujos extremos são iguais recebe o nome de **laço**. Duas ou mais arestas com o mesmo par de extremos recebem o nome de **arestas múltiplas**. A aresta $e5$ é

um laço e as arestas $e1$ e $e2$ definem arestas múltiplas no grafo da Figura 2.9. Um **grafo simples** é um grafo que não possui laços ou arestas múltiplas. Tais grafos (simples) podem ser especificados apenas pelo conjunto de vértices e conjunto de arestas, onde o último dos conjuntos é definido como um conjunto de pares não orientados de vértices, escrevendo $e = uv$ (ou $e = vu$) para uma aresta e com extremos u e v. Na Figura 2.10 apresentamos dois grafos simples. Doravante, quando falarmos em grafo, estaremos falando de grafo simples, a menos que seja especificado o contrário.

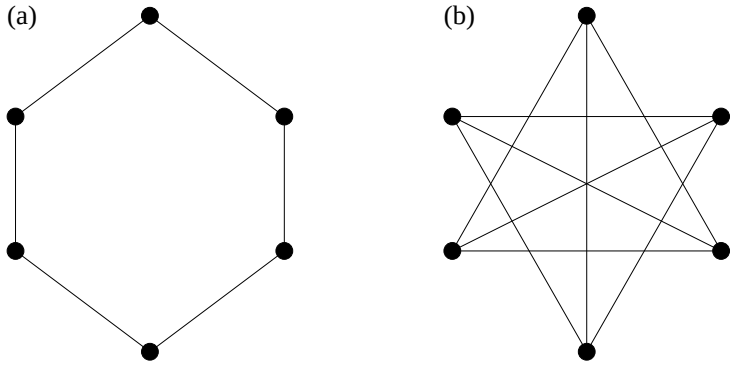


Figura 2.10: Exemplo de grafos complementares.

Sejam u e v os extremos de uma aresta. Dizemos que u e v são **adjacentes** ou **vizinhos**. O **complemento** $\overline{G}$ de um grafo G é o grafo com conjunto de vértices $V(G)$ e $uv \in E(\overline{G})$ se e somente se $uv \notin E(G)$. Na Figura 2.10 (a) apresentamos um exemplo de grafo simples e na Figura 2.10 (b) apresentamos seu complemento.

O grafo cujos vértices podem ser ordenados de tal forma que dois vértices são adjacentes se e somente se são consecutivos na lista é chamado **caminho**. Na Figura 2.11 apresentamos um exemplo de caminho.

Figura 2.11: Exemplo de caminho com quatro vértices.

Um **ciclo** é um grafo com um mesmo número de vértices e arestas cujos vértices podem ser dispostos em um círculo de tal forma que dois vértices são adjacentes se e somente se aparecem consecutivamente ao longo do círculo. Na Figura 2.10 (a) apresentamos um exemplo de ciclo.

Um **subgrafo** de um grafo G é um grafo H tal que $V(H) \subseteq V(G)$, $E(H) \subseteq E(G)$ e a associação dos extremos para as arestas em H é a mesma que em G. Na Figura 2.12 apresentamos um exemplo de subgrafo para o grafo apresentado na Figura 2.10 (b). Um **caminho** é uma seqüência de vértices tal que para cada um destes vértices há uma aresta para o próximo vértice na seqüência. Um grafo G é **conexo** se cada par de vértices

em G pertence a um caminho e **desconexo**, caso contrário. Os grafos apresentados nas Figuras 2.10 e 2.11 são grafos conexos. Já o grafo apresentado na Figura 2.12 é desconexo.

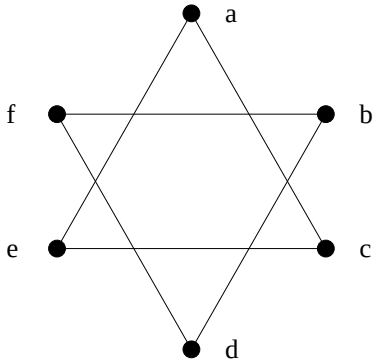


Figura 2.12: Exemplo de subgrafo do grafo apresentado na Figura 2.10 (b).

O **grau** do vértice v em um grafo G, denotado por $d(v)$, é o número de arestas incidentes a v. A **vizinhança** de v, denotada por $N(v)$, é o conjunto dos vértices adjacentes a v. No grafo da Figura 2.12 temos que todos os vértices possuem grau igual a dois, assim como $N(a) = \{c, e\}$.

Os **componentes** de um grafo G são seus subgrafos conexos maximais. Um componente (ou grafo) é **trivial** se não possui arestas. Um vértice que possui grau 0 é **isolado**. Um vértice isolado forma um componente trivial. No grafo da Figura 2.12 temos dois componentes. O primeiro é definido pelos vértices $\{a, c, e\}$ e o segundo por $\{b, d, f\}$.

Note a diferença entre maximal e máximo. No contexto deste trabalho, **máximo** significa "de tamanho máximo" e **maximal** significa "não há maior com tal propriedade". Todo caminho máximo é um caminho maximal, mas caminhos maximais não são necessariamente máximos.

Um **subgrafo induzido** é um subgrafo obtido pela remoção de um conjunto de vértices. Podemos falar também em subgrafo de G induzido por T, que é denotado por $G[T]$, significando $G - \overline{T}$ onde $\overline{T} = V(G) - T$. Na Figura 2.13 apresentamos o subgrafo, do grafo apresentado na Figura 2.12, induzido por $T = \{a, b, c, e\}$. Observe que o vértice b define agora um componente trivial.

A **ordem** de um grafo G, denotado por $n(G)$, é o número de vértices em G. O **tamanho** de um grafo G, denotado por $e(G)$, é o número de arestas em G. Note que, para G o grafo apresentado na Figura 2.13, temos $n(G) = 4$ e $e(G) = 3$.

Um **grafo de cordas sobrepostas** é um grafo onde vértices representam cordas de um círculo e arestas são criadas sempre que duas cordas se cruzam. Na Figura 2.14 apresentamos um exemplo de grafo de cordas sobrepostas.

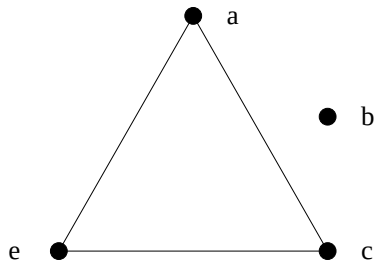


Figura 2.13: Subgrafo, do grafo apresentado na Figura 2.12, induzido por $T = \{a, b, c, e\}$.

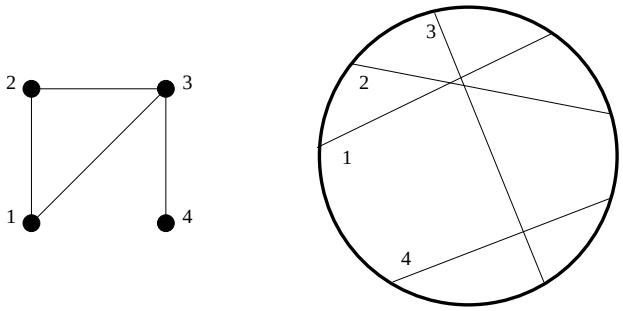


Figura 2.14: Exemplo de grafo de cordas sobrepostas

Capítulo 3

Comparação de Genomas

Neste capítulo iremos apresentar os conceitos e definições que vêem sendo empregados pela grande maioria dos pesquisadores em rearranjo de genomas, assim como uma visão geral do problema.

Uma forma de obter informações biologicamente importantes é comparar genomas, buscando por genomas (ou porções de genomas) já estudados que possuem semelhanças com o genoma em estudo ou mesmo procurando por padrões nos genomas já estudados ou não.

Há diversas formas de comparar genomas, uma delas é verificar as semelhanças nas seqüências de bases que compõem as moléculas de DNA dos genomas [27]. A seguir é apresentado um exemplo onde é exposto o resultado de uma busca por semelhanças entre as seqüências TACGTTGCCAATT e TACGGCCCATT. Em tais buscas procura-se emparelhar o maior número de bases e é possível inserir espaços (-) para tentar maximizar estes emparelhamentos.

```
TACGTTGCCAATT
TACG--GCCCATT
```

Na década de 1980 foi realizado um estudo onde os genomas mitocondriais do repolho e do nabo foram comparados e percebeu-se que as seqüências de genes destas espécies eram muito semelhantes, porém a ordem em que estes genes aparecem na molécula de DNA era extremamente diferente. Este e outros estudos mostraram que rearranjo de genomas é uma forma comum de evolução do DNA de cloroplastos, mitocôndrias, vírus e bactérias [3].

Tais descobertas levaram a uma outra forma de comparar genomas, onde as seqüências conservadas de DNA são rotuladas como **blocos** e passa-se a olhar o genoma como uma lista ordenada de blocos. Na Figura 3.1 apresentamos um exemplo de genoma. Para efeito de simplificação, chamaremos os blocos de **genes** no restante do texto. Em rearranjo de

genomas, dados dois genomas, π (origem) e σ (destino), e o conjunto das **operações** ou **eventos de rearranjo** (reversão, transposição, etc) permitidos, deseja-se descobrir o número mínimo de operações, que transformam π em σ.

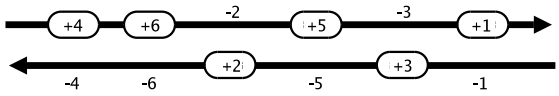


Figura 3.1: Abstração de um genoma como uma seqüência de genes.

Em rearranjo de genomas os genes são rotulados $1, 2, \ldots, n$ e sua ordem na molécula é representada por uma permutação $\pi = [\pi_1, \pi_2, \ldots, \pi_n]$. Seja o genoma

$$\pi = [\pi_1, \pi_2, \ldots, \pi_{i-1}, \pi_i, \pi_{i+1}, \ldots, \pi_{j-1}, \pi_j, \pi_{j+1}, \ldots, \pi_n]$$

A **reversão** $\rho(i, j)$ aplicada sobre o genoma π, resulta em

$$\rho\pi = [\pi_1, \pi_2, \ldots, \pi_{i-1}, \pi_j, \pi_{j-1}, \ldots, \pi_{i+1}, \pi_i, \pi_{j+1}, \ldots, \pi_n].$$

Se estiver trabalhando com genomas orientados, onde a cada gene é associado um sinal $+$ ou $-$, a reversão, além de reverter a ordem dos genes, altera o sinal de cada um dos genes. Os sinais simbolizam a orientação do bloco na molécula, ou seja, se $+i$ é um gene, $-i$ é o seu complemento reverso. Veja a Figura 3.1.

Digamos que estamos trabalhando com rearranjo de genomas e nos restringimos a operações de reversão. Sejam $\pi = [+1, -3, -2, -5, -4, +6]$ e $\sigma = [+1, +2, +3, +4, +5, +6]$ dois genomas. A Figura 3.2 apresenta um conjunto mínimo de operações que transforma π em σ.

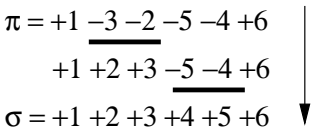


Figura 3.2: Seqüência mínima de operações que transforma π em σ.

Dadas as permutações π e σ, uma solução para o problema da **distância de reversão** é uma série de reversões $\rho_1, \rho_2, \ldots, \rho_t$, tal que $\rho_t \ldots \rho_2\rho_1\pi = \sigma$ e t é mínimo. A distância de reversão entre π e σ, denotada por $d(\pi, \sigma)$, é dada por este valor mínimo de t. O problema da **ordenação de π por reversões** é encontrar a distância de reversão entre π e a identidade $[+1, +2, \ldots, +n]$. Neste caso, como σ é a identidade, $d(\pi, \sigma)$ será denotado simplesmente por $d(\pi)$.

O primeiro algoritmo polinomial para ordenação por reversões orientadas foi apresentado por Hannenhalli e Pevzner [11]. Este algoritmo executa em tempo $O(n^4)$. Posteriormente, foram desenvolvidos algoritmos com desempenho assintoticamente melhor. Recentemente, Tannier e Sagot [29] apresentaram um algoritmo que executa em tempo $O(n\sqrt{n \log n})$. Existe também um algoritmo que executa em tempo linear, que foi apresentado por Bader, Moret e Yan [2]. Porém este só é capaz de determinar a distância, não sendo capaz de determinar a seqüência de reversões. O problema da distância de reversão sem sinais é NP-difícil [9].

Na formulação geral do problema são dados dois genomas (ou partes deles), π (origem) e σ (destino), na forma de lista ordenada de genes e um conjunto de operações (transposições, reversões, etc) permitidas. A partir desta entrada, deseja-se encontrar o número mínimo de operações que transformam o genoma π no genoma σ. Geralmente o número de operações necessários para transformar π em σ é o mesmo que para transformar σ em π, pois a inversa de uma operação é uma outra operação do mesmo tipo [10].

Ao comparar dois genomas, supõe-se que ambos possuam o mesmo conjunto de genes e não há genes repetidos em qualquer dos genomas. Restringe-se a eventos conservativos, ou seja, eventos que não alteram o conjunto de genes do genoma. Sendo assim eventos como duplicações e exclusões são descartados.

No restante deste trabalho nos restringiremos a operações de reversão.

Nas Seções 3.1 e 3.2 apresentamos as duas principais ferramentas no tratamento da distância de reversão – o problema que iremos tratar. Tais ferramentas são o "Diagrama de pontos de quebra" e o "Grafo de entrelaçamento de arcos". O conceito de obstáculo e uma série de conceitos e definições relacionadas apresentamos na Seção 3.3.

3.1 Diagrama de pontos de quebra

Vamos primeiramente considerar o caso dos genomas não orientados e, em seguida, mostraremos como as definições podem ser utilizadas em genomas orientados.

Antes da definição do diagrama precisamos apresentar uma série de definições e notações. Iremos denotar a relação $|i - j| = 1$ por $i \sim j$. Definimos a **extensão** de uma permutação da seguinte forma, se $\pi = [\pi_1\ \pi_2\ \ldots\ \pi_n]$ é uma permutação, adicionamos $\pi_0 = 0$ e $\pi_{n+1} = n + 1$. Dizemos que um par de elementos (π_i, π_{i+1}), $0 \leq i \leq n$, de π é uma **adjacência** se $\pi_i \sim \pi_{i+1}$, e um **ponto de quebra** se $\pi_i \nsim \pi_{i+1}$ [11]. O número de pontos de quebra é denotado por $b(\pi)$.

A identidade é uma permutação onde não há pontos de quebra, sendo assim, a tarefa de ordenar por reversões pode ser vista como uma eliminação dos pontos de quebra. Uma reversão pode eliminar no máximo dois pontos de quebra. Diante disto podemos afirmar que:

$$d(\pi) \geq \frac{b(\pi)}{2}$$

O **diagrama de pontos de quebra** de uma permutação π é um grafo $B(\pi)$ colorido nas arestas com $n+2$ vértices $\{\pi_0, \pi_1, \ldots, \pi_n, \pi_{n+1}\}$ [11]. Os vértices π_i e π_j são ligados por uma **aresta preta** caso (π_i, π_j) seja um ponto de quebra em π ($\pi_i \nsim \pi_j$ e $i \sim j$) e são ligados por uma **aresta cinza**, ou **arco**, caso (i, j) seja um ponto de quebra em π^{-1} (inversa de π), que pode ser escrito também como $\pi_i \sim \pi_j$ e $i \nsim j$. A Figura 3.3 apresenta um exemplo de diagrama de pontos de quebra.

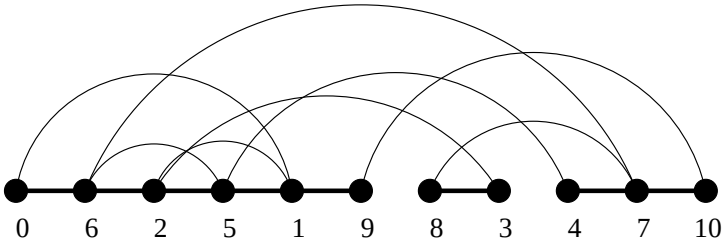


Figura 3.3: Diagrama de pontos de quebra de uma permutação sem sinais.

Sejam G um grafo colorido nas arestas e C um ciclo em G. O ciclo C é **alternante** se as cores de cada duas arestas consecutivas são distintas. Quando falarmos em ciclo, no restante do texto, entenda-se ciclo alternante. Dado um ciclo C, seu tamanho $l(C)$ é dado pelo número de arestas pretas.

Considere uma **decomposição em ciclos** de $B(\pi)$ em um número máximo $c(\pi)$ de ciclos alternantes disjuntos nas arestas. Para $\pi = [0, 6, 2, 5, 1, 9, 8, 3, 4, 7, 10]$, na Figura 3.3, temos $c(\pi) = 3$, já que $B(\pi)$ pode ser decomposto nos seguintes ciclos $(0, 1, 5, 6, 0)$, $(6, 7, 10, 9, 1, 2, 6)$ e $(2, 3, 8, 7, 4, 5, 2)$. Bafna e Pevzner [4] mostraram que uma reversão pode alterar o parâmetro $b(\pi) - c(\pi)$ por no máximo 1, e assim podemos dizer que:

$$d(\pi) \geq b(\pi) - c(\pi)$$

Bafna e Pevzner [4] perceberam que a definição de grafo de pontos de quebra poderia ser facilmente estendida para o caso de genomas orientados, ou seja, quando usam-se permutações com sinais. Para isso eles definiram uma transformação que leva uma permutação com sinais π de ordem n, ou seja, com n elementos para uma permutação sem sinais π' de ordem $2n$. Em tal transformação, um elemento positivo $+x$ em π é substituído por $2x-1, 2x$ em π' e um elemento negativo $-x$ em π é substituído por $2x, 2x-1$ em π'. O diagrama de pontos de quebra pode ser construído da mesma forma que anteriormente, só que utiliza-se a permutação π'. A Figura 3.4 apresenta um exemplo de diagrama de pontos de quebra de uma permutação com sinais. O diagrama de pontos de quebra para

permutações com sinais possui apenas vértices com grau 2 ou 0 e uma única decomposição em ciclos. Tais fatores o tornam mais fácil de ser manipulado.

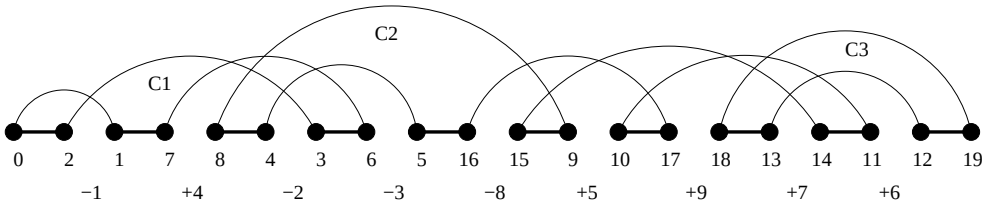


Figura 3.4: Diagrama de pontos de quebra de uma permutação com sinais.

Bergeron [5] sugere uma forma mais simples de montar o diagrama de pontos de quebra para permutações com sinais. Os vértices são obtidos pelo mesmo procedimento, o que muda é na definição das arestas. Junta-se com arestas pretas elementos consecutivos π_i e π_{i+1}, par sim, par não, começando do 0. Junta-se com arestas cinzas inteiros consecutivos, par sim, par não, começando com $(0, 1)$. É importante notar que os diagramas definidos por Hannenhalli e Pevzner [11] e por Bergeron [5] possuem uma pequena diferença. No primeiro não há ciclos de tamanho um, já no segundo, podem existir tais ciclos. Podemos simular o efeito de uma reversão $\rho(i, j)$ em π pela reversão $\rho(2i - 1, 2j)$ em π' [4].

Dada uma permutação π e uma reversão ρ, denota-se por Δb o incremento no número de pontos de quebra em π quando aplicada ρ, ou seja, $\Delta b = b(\rho\pi) - b(\pi)$, e por Δc o incremento no número de ciclos em π quando aplicada ρ, ou seja, $\Delta c = c(\rho\pi) - c(\pi)$. Tem-se que $\Delta(b - c) = \Delta b - \Delta c \geq -1$ [4]. Um reversão é **própria** caso $\Delta(b - c) = -1$.

Sejam os pontos de quebra (π_{i-1}, π_i) e (π_j, π_{j+1}). Diz-se diz que a reversão $\rho(i, j)$ **atua** nas arestas pretas (π_{i-1}, π_i) e (π_j, π_{j+1}). Uma reversão $\rho(i, j)$ **atua em um ciclo** C de $B(\pi)$ caso as arestas pretas (π_{i-1}, π_i) e (π_j, π_{j+1}) pertençam a C. Um arco g é **orientado** caso uma reversão atuando nas duas arestas pretas incidentes a g seja própria, e é **não orientado** caso tal reversão não seja própria [11]. Numa simplificação, Bergeron [5] define **suporte** de um arco como o intervalo de elementos de π' entre, e incluindo, os extremos. Ela define que um arco é orientado caso contenha um número ímpar de elementos no suporte, e é não orientado no caso em que é par. No exemplo da Figura 3.4 temos que apenas $(0, 1)$, $(6, 7)$, $(14, 15)$ e $(16, 17)$ são arcos orientados.

Um ciclo em $B(\pi)$ é **orientado** caso possua pelo menos um arco orientado, caso contrário é **não orientado**. No diagrama da Figura 3.4 os ciclos $C1$ e $C2$ são orientados. A partir da definição de ciclo orientado pode-se inferir que não existe reversão própria atuando em ciclos não orientados e que uma permutação só possuirá uma reversão própria caso tenha um ciclo orientado.

3.2 Grafo de entrelaçamento de arcos

Sejam g_1 e g_2 dois arcos, g_1 e g_2 **entrelaçam** se a interseção dos suportes não for vazia e cada suporte não está contido no outro.

O **grafo de entrelaçamento de arcos** é o grafo cujos vértices são os arcos no diagrama de pontos de quebra e cria-se aresta sempre que dois arcos entrelaçam [13]. O grafo de entrelaçamento de arcos referente ao diagrama apresentado na Figura 3.4 é apresentado na Figura 3.5. Um vértice relativo a um arco orientado é chamado **vértice orientado** (vértices pretos na figura), caso contrário é chamado **vértice não orientado** (vértices brancos na figura).

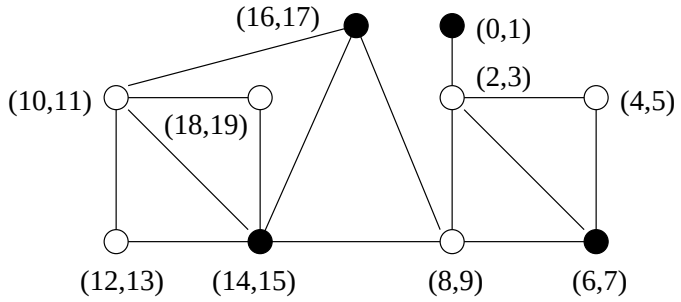


Figura 3.5: Grafo de entrelaçamento de arcos

A orientação é estendida para os componentes conexos do grafo de entrelaçamento de arcos, que chamaremos de **componentes**. Um componente é **orientado** caso possua pelo menos um vértice orientado, e é **não orientado** caso contrário. Note que um vértice possui grau ímpar se, e somente se, for orientado.

Ao aplicar-se uma reversão induzida por um vértice (arco) orientado v, o efeito no grafo de entrelaçamento será o complemento do subgrafo de v e os vértices adjacentes a v. Além disso, cada vértice adjacente a v terá sua orientação alterada [13]. A Figura 3.6 apresenta o grafo de entrelaçamento de arcos ao ser aplicada a reversão induzida pelo vértice orientado $(16, 17)$. A Figura 3.7 apresenta o resultado, da aplicação da reversão, sobre o diagrama de pontos de quebra.

Uma reversão é **segura** se, quando aplicada, não cria novos componentes não orientados, exceto para vértices isolados. Hannenhalli e Pevzner [11] mostram que qualquer seqüência de reversões orientadas e seguras é ótima, ou seja, diminui a distância de uma unidade.

A dificuldade em ordenar componentes orientados recai sobre o problema de encontrar reversões seguras. Bergeron [5] resolve este problema através de uma pontuação. Ela define a pontuação $p(v)$ de uma reversão orientada correspondente a um vértice orientado v conforme exposto na Equação 3.1. Na equação, T denota o número total de vértices

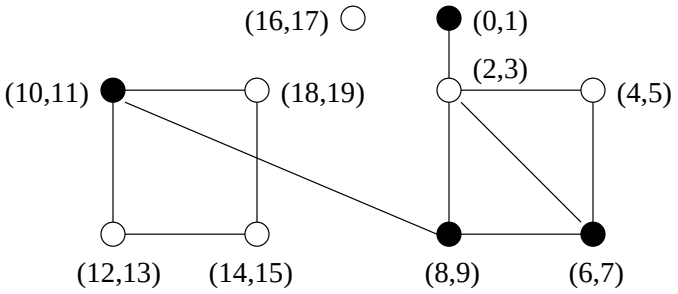


Figura 3.6: Grafo de entrelaçamento de arcos depois de aplicada a reversão.

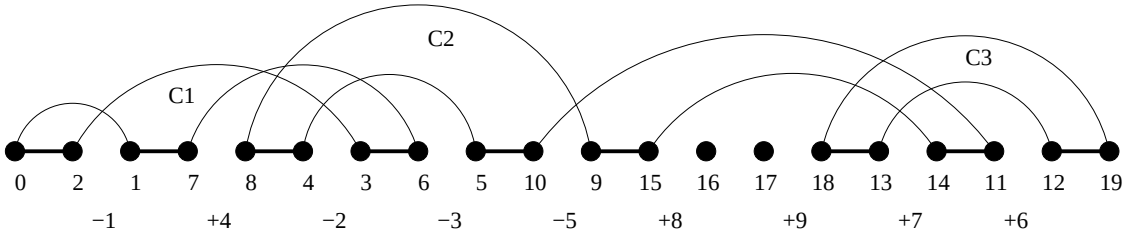


Figura 3.7: Diagrama de pontos quebra após a aplicação da reversão induzida pelo arco $(16, 17)$.

orientados no grafo de entrelaçamento, U denota o número de vértices não orientados adjacentes a v e O denota o número de vértices orientados adjacentes a v. Bergeron mostra que uma reversão orientada de pontuação máxima é segura.

$$p(v) = T + U - O - 1 \tag{3.1}$$

3.3 Obstáculos

Seja $\pi_{i_1}, \pi_{i_2}, \ldots, \pi_{i_k}$ a maior subseqüência da permutação estendida $0, \pi_1, \ldots, \pi_n, n+1$ tal que cada π_{i_j}, com $1 \leq j \leq k$, é um elemento incidente a um arco pertencente a um dos componentes não orientados do grafo de entrelaçamento. Ordene $\pi_{i_1}, \pi_{i_2}, \ldots, \pi_{i_k}$ em um ciclo CR tal que π_{i_j} segue $\pi_{i_{j-1}}$ para $2 \leq j \leq k$ e π_{i_1} segue π_{i_k}. Seja M um componente não orientado no grafo de entrelaçamento. Seja $E(M) \subset \{\pi_{i_1}, \pi_{i_2}, \ldots, \pi_{i_k}\}$ o conjunto dos terminais dos arcos em M. Dizemos que M é um **obstáculo** se os elementos de $E(M)$ ocorrem consecutivamente em CR [13]. Em nosso exemplo da Figura 3.4 não há componentes não orientados e, assim, não há obstáculos.

Uma reversão qualquer ρ que atua em um ciclo de um obstáculo K, elimina o obstáculo K, porém, eventualmente cria outros obstáculos [11]. O procedimento de aplicar tal reversão é uma operação que recebe o nome de **corte de obstáculo**. Um obstáculo que,

quando cortado, decrementa o número de obstáculos é um **obstáculo simples**. Quando um obstáculo não é simples ele é um **super obstáculo** [5]. Um corte de um obstáculo simples é uma reversão segura [11].

Sejam π uma permutação e ρ uma reversão que atua em arestas pretas de dois obstáculos distintos, digamos L e M, em π. A reversão ρ elimina os obstáculos L e M em $\rho\pi$ [11]. Tal procedimento define uma operação chamada **junção de obstáculos**. A operação de junção de obstáculos pode ser utilizada na busca por reversões seguras na ausência de componentes orientados e obstáculos simples.

Hannenhalli e Pevzner [11] mostram que, se $h(\pi) = 2$, existe uma reversão segura que realiza a junção dos dois obstáculos em π. Mostram também que, se $h(\pi) = 1$, existe uma reversão segura que corta o único obstáculo em π. Ainda mostram que, se $h(\pi) > 3$, existe uma reversão segura que realiza a junção de dois obstáculos em π.

Uma permutação π é uma **fortaleza** se possuir um número ímpar de obstáculos e todos estes são super obstáculos.

Em linhas gerais, Hannenhalli e Pevzner [11] mostram que:

- Caso π contenha apenas componentes orientados, $d(\pi) = b(\pi) - c(\pi)$.
- Caso π contenha obstáculos e não seja uma fortaleza, será necessária a aplicação de uma reversão a mais para cada obstáculo e assim $d(\pi) = b(\pi) - c(\pi) + h(\pi)$.
- Caso π seja uma fortaleza, será necessária uma reversão para "destruir" a fortaleza e $h(\pi)$ reversões para eliminar os obstáculos, assim $d(\pi) = b(\pi) - c(\pi) + h(\pi) + 1$.

Capítulo 4

Grupos de Permutação

Neste capítulo iremos apresentar os conceitos e definições relativas a permutações. Permutações são funções, com determinadas restrições, que são a base para a construção do formalismo algébrico para solução de problemas em rearranjo de genomas, inicialmente definido por Meidanis e Dias [17], que neste trabalho é estendido objetivando a apresentação de uma solução formal (algébrica) para o problema da distância de reversão.

Para maiores informações sobre funções e permutações veja as obras sobre álgebra na lista de referências [12, 16, 19].

Na Seção 4.1 são apresentados os conceitos e definições relativas a função. Os conceitos e definições clássicas de permutação, assim como novas operações, que serão úteis no formalismo algébrico, são apresentadas na Seção 4.2.

4.1 Função

Sejam X e Y dois conjuntos. Uma **função** f de X em Y, denotada por $f : X \mapsto Y$, relaciona cada elemento $x \in X$ a um único elemento $y = f(x)$ tal que $y \in Y$. Em outras palavras, f é uma relação binária entre os dois conjuntos tal que:

1. f é unívoca: se $y = f(x)$ e $z = f(x)$, então $y = z$.

2. f é total: para todo $x \in X$, existe um $y \in Y$ tal que $y = f(x)$.

Sejam os conjuntos $X = \{1, 2, 3\}$ e $Y = \{a, b, c, d, e\}$. A Figura 4.1 apresenta a função $f : X \mapsto Y$ que mapeia 1 em a, 2 em c e 3 em d. Tal função pode ser explicitamente definida pela seguinte expressão:

$$f(x) = \begin{cases} a, & se \quad x = 1 \\ c, & se \quad x = 2 \\ d, & se \quad x = 3. \end{cases}$$

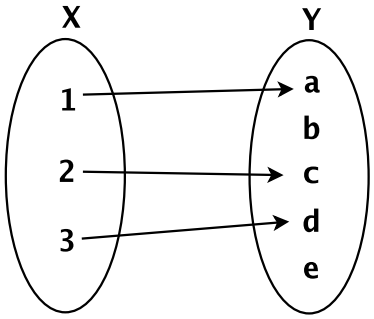


Figura 4.1: Função injetora $f : X \mapsto Y$.

Ao conceito de função estão associados três conjuntos especiais chamados domínio, contradomínio e imagem. O **domínio** (D) contém todos os elementos para os quais a função deve ser definida. O **contradomínio** (CD) contém os elementos que podem ser relacionados a elementos do domínio. Já o conjunto **imagem** (I) refere-se aos valores que efetivamente $f(x)$ assume e assim $I \subset CD$. Na função apresentada na Figura 4.1 temos os seguintes conjuntos $D = \{1, 2, 3\}$, $CD = \{a, b, c, d, e\}$ e $I = \{a, c, d\}$.

Uma função é chamada **sobrejetora** se todos os elementos do contradomínio estão associados a algum elemento do domínio ou, em outras palavras, $I = CD$. Uma função é chamada **injetora** se cada elemento do contradomínio está associado a no máximo um elemento do domínio. Sejam X e Y conjuntos, $x_1 \in X$, $x_2 \in X$ e $f : X \mapsto Y$ uma função. Se f é um função injetora e $x_1 \neq x_2$ então $f(x_1) \neq f(x_2)$. Uma função é **bijetora** se é injetora e sobrejetora ao mesmo tempo. Denotamos por $S(E)$ o conjunto de todas as funções bijetoras sobre o conjunto E. A Figura 4.1 ilustra um exemplo de função injetora, uma função sobrejetora é apresentada na Figura 4.2 e a Figura 4.3 nos mostra uma função bijetora.

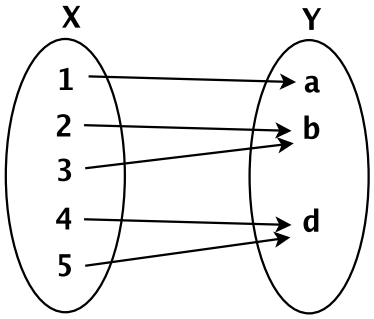


Figura 4.2: Função sobrejetora $f : X \mapsto Y$.

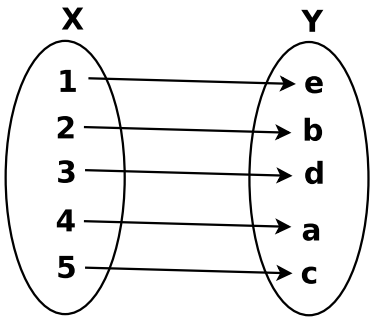


Figura 4.3: Função bijetora $f : X \mapsto Y$.

4.2 Permutação

Uma **permutação** α sobre um conjunto base E é uma função bijetora $\alpha : E \mapsto E$. Seja o conjunto base $E = \{x_u, x_v, \ldots, x_w\}$ e uma permutação α. Representamos a permutação α da seguinte maneira:

$$\begin{pmatrix} x_u & x_v & \ldots & x_w \\ \alpha(x_u) & \alpha(x_v) & \ldots & \alpha(x_w) \end{pmatrix}.$$

Seja α uma permutação. Utilizaremos $E(\alpha)$ para denotar o conjunto base de α.

Para que as permutações sejam qualificadas como grupo, além do conjunto base E, é necessária a definição de uma operação binária de $S(E) \times S(E)$ em $S(E)$. Tal operação deve satisfazer os seguintes axiomas: associatividade, existência de elemento identidade e existência de elemento inverso para todos os elementos.

A operação binária definida em permutações é chamada **composição** ou **produto**. Para calcular um produto entre as permutações α e β, denotado por $\alpha\beta$, aplica-se sucessivamente as funções da direita para esquerda, ou seja, para $x \in E$ tem-se que no produto ele é mapeado em $\alpha(\beta(x))$. O produto $\alpha\alpha$ pode ser denotado também por α^2, assim como $\alpha\alpha\alpha = \alpha^3$ e assim sucessivamente.

Seja α uma permutação. Diz-se que x é **fixo** em α se $\alpha(x) = x$. A **identidade**, denotada por 1, é a permutação onde todo $x \in E$ é fixo. Para qualquer permutação α, temos $\alpha 1 = \alpha = 1\alpha$. A permutação identidade é dada por

$$\begin{pmatrix} x_u & x_v & \ldots & x_w \\ x_u & x_v & \ldots & x_w \end{pmatrix}.$$

Dada uma permutação α, seu **inverso** α^{-1} é dado por

$$\begin{pmatrix} \alpha(x_u) & \alpha(x_v) & \ldots & \alpha(x_w) \\ x_u & x_v & \ldots & x_w \end{pmatrix}.$$

Para qualquer permutação α, temos $\alpha\alpha^{-1} = 1 = \alpha^{-1}\alpha$.

Lema 1. *Seja α uma permutação. Se $\alpha^2 = 1$ então $\alpha = \alpha^{-1}$.*

Demonstração. Sabe-se que $\alpha\alpha^{-1} = 1$. Como $\alpha^2 = 1$ então:

$$\alpha\alpha^{-1} = 1$$
$$\alpha^2\alpha^{-1} = \alpha$$
$$\alpha^{-1} = \alpha$$

□

Sejam o conjunto base E, a permutação α e o elemento $x \in E$. A aplicação sucessiva de α: x, $\alpha(x)$, $\alpha(\alpha(x))$, $\alpha(\alpha(\alpha(x)))$, ..., define um conjunto finito ao qual chamaremos de **órbita** de x. Uma permutação dá origem a uma ou mais órbitas e o número de órbitas de uma permutação α é denotado por $orb(\alpha)$. Uma permutação que possui exatamente uma órbita de cardinalidade maior ou igual a 2 recebe o nome de **ciclo**. Se o tamanho desta órbita for r, dizemos que é um **r-ciclo**.

Uma outra forma de se representar uma permutação é através de ciclos como apresentado a seguir

$$\alpha = (b\ d\ a\ e\ c)(f\ g) = (d\ a\ e\ c\ b)(g\ f) = (a\ e\ c\ b\ d)(f\ g)$$

significando que b é mapeado em d, d em a, a em e, e em c, c em b, f em g e g em f. Para qualquer elemento $x \in E$ e $x \notin \{a, b, c, d, e, f, g\}$ tem-se $\alpha(x) = x$. Tais ciclos correspondem a órbitas da permutação. Nesta representação, ciclos de tamanho 1 não são escritos, sendo assim, $E = \{a, b, c, d, e, f, g\}$ é o conjunto base mínimo para α. Caso $E = \{a, b, c, d, e, f, g\}$ para α, tem-se $orb(\alpha) = 2$ e órbitas $\{a, b, c, d, e\}$ e $\{f, g\}$.

Note que as permutações possuem um caráter circular e uma mesma permutação pode ser escrita de diferentes maneiras, assim como exibido no exemplo anterior. Para obter o inverso, nesta representação, basta inverter a ordem dos ciclos e a ordem dos elementos dentro de cada ciclo. No exemplo anterior temos que $\alpha^{-1} = (g\ f)(c\ e\ a\ d\ b)$.

Para exemplificar o produto, considere a seguinte situação. Dadas as permutações $\alpha = (b\ e\ c\ a)$ e $\beta = (e\ c\ b\ a\ d)$, o produto $\alpha\beta$ é:

$$\alpha\beta = (b\ e\ c\ a)(e\ c\ b\ a\ d) = (a\ d\ c\ e)(b) = (a\ d\ c\ e)$$

Observe que para o elemento a tem-se que $\beta(a) = d$ e $\alpha(d) = d$, então no produto $\alpha\beta$ temos que a é levado em d, o que indicaremos por $a \mapsto d$. Para d, temos $\beta(d) = e$ e $\alpha(e) = c$, então $d \mapsto c$. Para c, temos $\beta(c) = b$ e $\alpha(b) = e$, então $c \mapsto e$. Para e, temos $\beta(e) = c$ e $\alpha(c) = a$, então $e \mapsto a$ e o ciclo é fechado. Para b, temos $\beta(b) = a$ e $\alpha(a) = b$, então $b \mapsto b$. Lembramos que ciclos de tamanho 1 não são escritos.

Dada uma permutação α, seu **suporte**, denotado por $Supp(\alpha)$, é o conjunto dos elementos não fixos na permutação, conforme apresentado na Equação 4.1.

$$Supp(\alpha) = \{x \in E \mid \alpha(x) \neq x\} \tag{4.1}$$

Se α é escrita como $\alpha_1\alpha_2 \ldots \alpha_k$, onde α_i e α_j, com $1 \leq i \leq k$ e $1 \leq j \leq k$, são ciclos e $Supp(\alpha_i) \cap Supp(\alpha_j) = \emptyset$ para $i \neq j$, diz-se que a permutação está em sua **decomposição em ciclos disjuntos**. Toda permutação possui uma única decomposição em ciclos disjuntos, a menos da ordem dos ciclos.

4.2.1 Ciclos curtos

Seja $\alpha = (x\ y)$ um 2-ciclo. Ao realizar o produto de α por uma permutação β qualquer tem-se que se x e y estão em um mesmo ciclo de β, então este ciclo será quebrado em dois na permutação $\alpha\beta$. Caso x e y estejam em ciclos distintos de β, então os dois ciclos serão fundidos em $\alpha\beta$. Comportamento semelhante é apresentado para $\beta\alpha$. Tais situações podem ser evidenciadas nos produtos que seguem para $\alpha = (a\ b)$, $\beta = (a\ b\ c\ d)$ e $\gamma = (a\ c)(b\ d)$.

$$\alpha\beta = (a)(b\ c\ d)$$
$$\beta\alpha = (a\ c\ d)(b)$$
$$\alpha\gamma = (a\ c\ b\ d)$$
$$\gamma\alpha = (a\ d\ b\ c)$$

4.2.2 Novas operações

Como ferramental para o formalismo para rearranjo de genomas, Meidanis e Dias [17] ainda definem outras operações e conceitos tais como conjugação, norma e divisibilidade, os quais são descritos a seguir.

Conjugação

A **conjugação** de β por α, denotada por $\alpha \bullet \beta$, é o produto $\alpha\beta\alpha^{-1}$. A conjugação é uma operação que renomeia os elementos do suporte de β. Dada uma permutação $\beta = (b_1\ b_2\ \ldots\ b_n)$, a conjugação $\alpha \bullet \beta$ resulta em $(\alpha(b_1)\ \alpha(b_2)\ \ldots\ \alpha(b_n))$. Se β é um produto de ciclos, α afeta cada ciclo desta mesma forma. Se $\beta = (b_1\ b_2\ b_3)(b_4\ b_5)$ então $\alpha \bullet \beta = (\alpha(b_1)\ \alpha(b_2)\ \alpha(b_3))(\alpha(b_4)\ \alpha(b_5))$.

Norma e Divisibilidade

Permutações podem ser escritas como produto de 2-ciclos. Por exemplo:

$$\alpha = (a\ b\ c\ d)(e\ f\ g) = (a\ b)(b\ c)(c\ d)(e\ f)(f\ g)$$

A **norma** de uma permutação α, denotada por $\|\alpha\|$, é o menor número k, tal que é possível escrever α como um produto de k 2-ciclos. No exemplo anterior é apresentada uma das formas de se obter o menor produto de k 2-ciclos para α. Uma outra forma de se obter a norma é apresentada na Equação 4.2.

$$\|\alpha\| = |E| - orb(\alpha) \tag{4.2}$$

Lema 2. *Para quaisquer que sejam α e β permutações, tem-se:*

1. $\|\alpha\| = 0 \Leftrightarrow \alpha = 1$
2. $\|\alpha^{-1}\| = \|\alpha\|$
3. $\|\beta \bullet \alpha\| = \|\alpha\|$
4. $\|\alpha\beta\| \leq \|\alpha\| + \|\beta\|$
5. $\|\alpha\beta\| = \|\beta\alpha\|$
6. $orb(\alpha\beta) = orb(\beta\alpha)$

Demonstração. Seguem as demonstrações abaixo:

1. $\Rightarrow$ o produto de zero 2-ciclos é a identidade.

   $\Leftarrow \|1\| = 0$.

2. Se $\alpha = \alpha_1 \ldots \alpha_k$, com cada α_i um 2-ciclo e $1 \leq i \leq k$, então $\alpha^{-1} = \alpha_k \ldots \alpha_1$. Sendo assim, para cada decomposição de $\|\alpha\|$ em 2-ciclos nós temos uma decomposição de $\|\alpha^{-1}\|$ com o mesmo número de ciclos. Isto prova que $\|\alpha^{-1}\| \leq \|\alpha\|$. Por simetria concluímos que $\|\alpha\| \leq \|\alpha^{-1}\|$ e então $\|\alpha\| = \|\alpha^{-1}\|$.

3. Seja $\alpha = \alpha_1 \ldots \alpha_k$, para α_i com $1 \leq i \leq k$ um 2-ciclo. Seque que:

   $$\beta \bullet \alpha = (\beta \bullet \alpha_1)(\beta \bullet \alpha_2) \ldots (\beta \bullet \alpha_k)$$

   com cada $\beta \bullet \alpha_i$ um 2-ciclo. Isto permite-nos concluir que $\|\beta \bullet \alpha\| \leq \|\alpha\|$. Como isto é válido para todo α e β, temos $\|\beta^{-1} \bullet (\beta \bullet \alpha)\| \leq \|\beta \bullet \alpha\|$ ou, equivalentemente, $\|\alpha\| \leq \|\beta \bullet \alpha\|$.

4. Seja

$$\|\alpha\| = k \Leftrightarrow \alpha = \alpha_1 \dots \alpha_k$$

$$\|\beta\| = l \Leftrightarrow \beta = \beta_1 \dots \beta_l$$

para α_i, com $1 \leq i \leq k$, assim como β_j, com $1 \leq j \leq l$, 2-ciclos. Segue que:

$$\alpha\beta = \alpha_1 \dots \alpha_k \beta_1 \dots \beta_l \Rightarrow \|\alpha\beta\| \leq k + l$$

e então $\|\alpha\beta\| \leq \|\alpha\| + \|\beta\|$.

5. De acordo com o item 3 podemos dizer que:

$$\begin{aligned} \|\alpha\beta\| &= \|\beta \bullet (\alpha\beta)\| \\ &= \|\beta\alpha\beta\beta^{-1}\| \\ &= \|\beta\alpha\| \end{aligned}$$

6. Pela Equação 4.2 podemos afirmar que $\|\alpha\beta\| = |E(\alpha\beta)| - orb(\alpha\beta)$ e que $\|\beta\alpha\| = |E(\beta\alpha)| - orb(\beta\alpha)$. Observe que $|E(\alpha\beta)| = |E(\beta\alpha)|$. Pelo Item 5 temos que $\|\alpha\beta\| = \|\beta\alpha\|$ e assim temos:

$$\begin{aligned} \|\alpha\beta\| &= \|\beta\alpha\| \\ |E(\alpha\beta)| - orb(\alpha\beta) &= |E(\beta\alpha)| - orb(\beta\alpha) \\ orb(\alpha\beta) &= orb(\beta\alpha) \end{aligned}$$

□

Sejam as permutações α e β. Diz-se que α **divide** β, denotado por $\alpha \mid \beta$, quando $\|\beta\alpha^{-1}\| = \|\beta\| - \|\alpha\|$. O conceito de divisibilidade em permutações tem uma finalidade semelhante a encontrada no contexto dos números inteiros. Quando este conceito é empregado deseja-se descobrir se algo compõe o todo. Por exemplo, se $\alpha = (a\ c\ e)$ e $\beta = (a\ b\ c\ d\ e)$ tem-se que $\alpha \mid \beta$. Note que a, c e e aparecem em um mesmo ciclo de β e que aparecem na mesma ordem.

Seja α uma permutação. Formalmente, um **ciclo** em α é uma permutação β tal que $orb(\beta) = 1$, $\beta \mid \alpha$ e não existe um $\gamma \neq \beta$ tal que $orb(\gamma) = 1$, $\beta \mid \gamma$ e $\gamma \mid \alpha$.

Lema 3. *Para quaisquer que sejam α, β, γ permutações, tem-se:*

1. $1 \mid \alpha$ *para todo* α
2. $\alpha \mid \alpha$ *para todo* α
3. $\alpha \mid \beta$ *e* $\beta \mid \gamma \Rightarrow \alpha \mid \gamma$

4. $\alpha \mid \beta \Leftrightarrow \alpha^{-1} \mid \beta^{-1}$

5. $\alpha \mid \beta \Leftrightarrow \gamma \bullet \alpha \mid \gamma \bullet \beta$ *para todo* α, β *e* γ

6. $\alpha \mid \beta$ *e* $\beta \mid \alpha \Rightarrow \alpha = \beta$

Demonstração. Seguem as demonstrações abaixo:

1. Para afirmar que $1 \mid \alpha$ precisamos ter $\|\alpha 1^{-1}\| = \|\alpha\| - \|1\|$. Observe que $\|\alpha 1^{-1}\| = \|\alpha 1\| = \|\alpha\|$ e que $\|\alpha\| - \|1\| = \|\alpha\| - 0 = \|\alpha\|$.

2. Para que $\alpha \mid \alpha$ precisamos ter $\|\alpha\alpha^{-1}\| = \|\alpha\| - \|\alpha\|$. Observe que $\|\alpha\alpha^{-1}\| = \|1\| = 0$ e que $\|\alpha\| - \|\alpha\| = 0$

3. Como $\alpha \mid \beta$ e $\beta \mid \gamma$ podemos afirmar que $\|\beta\alpha^{-1}\| = \|\beta\| - \|\alpha\|$ e $\|\gamma\beta^{-1}\| = \|\gamma\| - \|\beta\|$. Com alguma álgebra podemos afirmar que $\|\beta\alpha^{-1}\| + \|\gamma\beta^{-1}\| = \|\gamma\| - \|\alpha\|$, assim como, de acordo com as propriedades da norma, podemos afirmar que $\|\gamma\beta^{-1}\beta\alpha^{-1}\| \leq \|\gamma\beta^{-1}\| + \|\beta\alpha^{-1}\|$ e conseqüentemente $\|\gamma\alpha^{-1}\| \leq \|\gamma\| - \|\alpha\|$. Pelas propriedades da norma é possível também afirmar que

$$\begin{gathered}\|\gamma\alpha^{-1}\alpha\| \leq \|\gamma\alpha^{-1}\| + \|\alpha\| \\ \|\gamma\| \leq \|\gamma\alpha^{-1}\| + \|\alpha\| \\ \|\gamma\| - \|\alpha\| \leq \|\gamma\alpha^{-1}\|\end{gathered}$$

   Finalmente, como $\|\gamma\alpha^{-1}\| \leq \|\gamma\| - \|\alpha\|$ e $\|\gamma\alpha^{-1}\| \geq \|\gamma\| - \|\alpha\|$, concluímos que $\|\gamma\alpha^{-1}\| = \|\gamma\| - \|\alpha\|$ e conseqüentemente $\alpha \mid \gamma$.

4. $\Rightarrow$ Se $\alpha \mid \beta$ então

$$\begin{aligned}\|\beta\alpha^{-1}\| &= \|\beta\| - \|\alpha\| \\ \|(\beta\alpha^{-1})^{-1}\| &= \|\beta^{-1}\| - \|\alpha^{-1}\| \\ \|\alpha\beta^{-1}\| &= \|\beta^{-1}\| - \|\alpha^{-1}\| \\ \|\beta^{-1}\alpha\| &= \|\beta^{-1}\| - \|\alpha^{-1}\|\end{aligned}$$

   e, assim, $\alpha^{-1} \mid \beta^{-1}$.

   $\Leftarrow$ Se $\alpha^{-1} \mid \beta^{-1}$ então

$$\begin{aligned}\|\beta^{-1}\alpha\| &= \|\beta^{-1}\| - \|\alpha^{-1}\| \\ \|(\beta^{-1}\alpha)^{-1}\| &= \|\beta\| - \|\alpha\| \\ \|\alpha^{-1}\beta\| &= \|\beta\| - \|\alpha\| \\ \|\beta\alpha^{-1}\| &= \|\beta\| - \|\alpha\|\end{aligned}$$

   e, assim, $\alpha \mid \beta$.

5. $\Rightarrow$ Se $\alpha \mid \beta$ então

$$\begin{aligned} \|\beta\alpha^{-1}\| &= \|\beta\| - \|\alpha\| \\ \|\gamma \bullet (\beta\alpha^{-1})\| &= \|\gamma \bullet \beta\| - \|\gamma \bullet \alpha\| \\ \|(\gamma \bullet \beta)(\gamma \bullet \alpha^{-1})\| &= \|\gamma \bullet \beta\| - \|\gamma \bullet \alpha\| \\ \|(\gamma \bullet \beta)(\gamma \bullet \alpha)^{-1}\| &= \|\gamma \bullet \beta\| - \|\gamma \bullet \alpha\| \end{aligned}$$

   e, assim, $\gamma \bullet \alpha \mid \gamma \bullet \beta$.

   $\Leftarrow$ Se $\gamma \bullet \alpha \mid \gamma \bullet \beta$ então

$$\begin{aligned} \|(\gamma \bullet \beta)(\gamma \bullet \alpha)^{-1}\| &= \|\gamma \bullet \beta\| - \|\gamma \bullet \alpha\| \\ \|\gamma\beta\gamma^{-1}\gamma\alpha^{-1}\gamma^{-1}\| &= \|\beta\| - \|\alpha\| \\ \|\gamma \bullet (\beta\alpha^{-1})\| &= \|\beta\| - \|\alpha\| \\ \|\beta\alpha^{-1}\| &= \|\beta\| - \|\alpha\| \end{aligned}$$

   e, assim, $\alpha \mid \beta$.

6. Como $\alpha \mid \beta$ e $\beta \mid \alpha$ podemos afirmar que $\|\beta\alpha^{-1}\| = \|\beta\| - \|\alpha\|$ e $\|\alpha\beta^{-1}\| = \|\alpha\| - \|\beta\|$. Observe que $\|\beta\alpha^{-1}\| = \|(\beta\alpha^{-1})^{-1}\| = \|\alpha\beta^{-1}\|$ e desta forma $\|\beta\| - \|\alpha\| = \|\alpha\| - \|\beta\|$ que resulta em $\|\alpha\| = \|\beta\|$. Note que, sendo $\|\alpha\| = \|\beta\|$ e $\|\beta\alpha^{-1}\| = \|\beta\| - \|\alpha\|$, temos $\|\beta\alpha^{-1}\| = 0$, ou seja, $\beta\alpha^{-1} = 1$. Finalmente, a última observação nos leva a concluir que $\alpha = \beta$.

□

Lema 4. *Sejam α e β permutações. A expressão $\alpha \mid \beta$ é verdadeira se e somente se $(\beta \bullet \alpha) \mid \beta$.*

Demonstração. Temos que $\alpha \mid \beta \Leftrightarrow \gamma \bullet \alpha \mid \gamma \bullet \beta$, para quaisquer α, β e γ. Desta forma, se $\gamma = \beta$ então $\alpha \mid \beta \Leftrightarrow \beta \bullet \alpha \mid \beta$, pois $\beta \bullet \beta = \beta$. □

Máximo divisor comum

Definimos ainda o máximo divisor comum entre duas permutações.

Sejam α e β permutações. O **máximo divisor comum** de α e β, denotado por $\alpha \wedge \beta$, é uma permutação de maior norma que divide α e β. Por exemplo, caso $\alpha = (a\ b\ c\ d\ e)$ e $\beta = (a\ c\ f)$ então $\alpha \wedge \beta = (a\ c)$.

Observe que pode existir mais que um máximo divisor comum para um dado par de permutações α e β. Por exemplo, se $\alpha = (a\ b\ c)$ e $\beta = (a\ c\ b)$ então $\alpha \wedge \beta$ pode assumir $(a\ b)$, $(a\ c)$ ou $(b\ c)$.

Capítulo 5

Um formalismo algébrico para rearranjo de genomas

Neste capítulo iremos apresentar o formalismo algébrico definido por Meidanis e Dias [17] para solução de problemas na área de rearranjo de genomas. Utilizaremos tal formalismo em nossa solução para o problema da distância de reversão que apresentamos no Capítulo 6.

Na Seção 5.1 é introduzido o formalismo com a definição do conjunto-base, gene e a permutação Γ, que mapeia cada elemento em seu oposto. O conceito de genoma e correlatos na ótica do formalismo é apresentado na Seção 5.2. Na Seção 5.3 é apresentado o conceito de operação em genoma, assim como a definição de reversão. A definição de distância de reversão e uma série de definições relacionadas são apresentadas na Seção 5.4. Finalmente, na Seção 5.5 apresentamos como realizar o mapeamento de um genoma linear para um circular e vice-versa. Uma vez que o formalismo é aplicável diretamente apenas a genomas circulares, tal mapeamento permite que genomas lineares sejam também tratados pelo formalismo.

5.1 Genes e complementaridade

Doravante denotaremos a aplicação de permutações em elementos sem a utilização de parênteses, escrevendo simplesmente αa em lugar de $\alpha(a)$. Como elementos são sempre representados por letras latinas e permutações por letras gregas, não deverá haver confusão.

Introduzimos o formalismo com a definição de um conjunto, o qual chamamos de **conjunto-base**, $E_n = \{+e_1, -e_1, +e_2, -e_2, \ldots, +e_n, -e_n\}$ onde n é o **número de genes**. Cada elemento $+e_i$, onde $1 \leq i \leq n$, representa um **gene**, e seu oposto $-e_i$ simboliza seu complemento reverso, ou seja, a seqüência que se encontra na mesma localização porém

na fita oposta da molécula de DNA. Define-se Γ como uma permutação que mapeia cada elemento em seu oposto. A permutação Γ pode ser escrita como:

$$\Gamma = (+e_1\ -e_1)(+e_2\ -e_2)\dots(+e_n\ -e_n).$$

A permutação Γ possui as seguintes propriedades:

1. $\Gamma a \neq a$ para todo $a \in E_n$
2. $\Gamma^2 = 1$
3. $\Gamma^{-1} = \Gamma$

5.2 Genomas circulares

Dizemos que uma permutação α, com conjunto-base contido no conjunto-base de Γ, é **admissível** caso a única permutação que divida α e Γ ao mesmo tempo seja a identidade. Uma permutação π é um **genoma** caso seja admissível e satisfaça a igualdade $\pi^{-1} = \Gamma \bullet \pi$. Cada ciclo de um genoma recebe o nome de **fita**.

Seja $\Gamma = (+a\ -a)(+b\ -b)(+c\ -c)(+d\ -d)(+e\ -e)$. A permutação $\alpha = (+a\ -c\ +e)$ é admissível sob Γ, $\pi = (+c\ -a\ -d\ +b\ -e)(+e\ -b\ +d\ +a\ -c)$ é um genoma sob Γ e $\pi_1 = (+c\ -a\ -d\ +b\ -e)$ e $\pi_2 = (+e\ -b\ +d\ +a\ -c)$ são fitas de π.

Seja π um genoma e π_1 uma fita de π. A permutação $\Gamma \bullet \pi_1^{-1}$ define a **fita complementar** de π_1. Um par de fitas complementares de um genoma recebe o nome de **cromossomo**. Um genoma é composto por cromossomos possuindo conjuntos-bases disjuntos.

Lema 5. *A igualdade $\Gamma\pi^{-1} = \pi\Gamma$ é válida para qualquer que seja o genoma π definido sob Γ.*

Demonstração. Pela definição de genoma temos que $\pi^{-1} = \Gamma \bullet \pi$. Segue que:

$$\begin{aligned} \Gamma\pi^{-1} &= \Gamma(\Gamma \bullet \pi) \\ &= \Gamma\Gamma\pi\Gamma^{-1} \\ &= \pi\Gamma \end{aligned}$$

□

Diz-se que dois genomas são **compatíveis** caso seus conjuntos-bases sejam iguais e sejam definidos sob uma mesma Γ.

Lema 6. *Seja π um genoma definido sob Γ e α uma permutação tal que $\alpha^2 = 1$. A expressão $\alpha \mid \pi$ é verdadeira se e somente se $(\Gamma \bullet \alpha) \mid \pi$.*

Demonstração. $\Rightarrow$ Se $\alpha \mid \pi$ então $\|\pi\alpha^{-1}\| = \|\pi\| - \|\alpha\|$ ou ainda, como $\alpha^2 = 1$, $\|\pi\alpha\| = \|\pi\| - \|\alpha\|$. Por outro lado $\|\pi\| - \|\Gamma \bullet \alpha\| = \|\pi\| - \|\alpha\|$. Agora observe que:

$$\begin{aligned}
\|\pi(\Gamma \bullet \alpha)^{-1})\| &= \|\pi(\Gamma\alpha\Gamma^{-1})^{-1}\| \\
&= \|\pi\Gamma\alpha^{-1}\Gamma^{-1}\| \\
&= \|\Gamma\pi^{-1}\alpha^{-1}\Gamma^{-1}\| \\
&= \|\Gamma \bullet (\pi^{-1}\alpha^{-1})\| \\
&= \|(\alpha\pi)^{-1}\| \\
&= \|\alpha\pi\| \\
&= \|\pi\alpha\|
\end{aligned}$$

Segue que $\|\pi(\Gamma \bullet \alpha)^{-1}\| = \|\pi\| - \|\alpha\| = \|\pi\| - \|\Gamma \bullet \alpha\|$ e assim $(\Gamma \bullet \alpha) \mid \pi$.
$\Leftarrow$ Se $(\Gamma \bullet \alpha) \mid \pi$ e $\alpha^2 = 1$ então:

$$\begin{gathered}
\|\pi(\Gamma \bullet \alpha)^{-1}\| = \|\pi\| - \|\Gamma \bullet \alpha\| \\
\|\pi(\Gamma\alpha\Gamma^{-1})^{-1}\| = \|\pi\| - \|\alpha\| \\
\|\pi\Gamma\alpha^{-1}\Gamma^{-1}\| = \|\pi\| - \|\alpha\| \\
\|\Gamma\pi^{-1}\alpha^{-1}\Gamma^{-1}\| = \|\pi\| - \|\alpha\| \\
\|\Gamma \bullet (\pi^{-1}\alpha^{-1})\| = \|\pi\| - \|\alpha\| \\
\|(\alpha\pi)^{-1}\| = \|\pi\| - \|\alpha\| \\
\|\alpha\pi\| = \|\pi\| - \|\alpha\| \\
\|\pi\alpha\| = \|\pi\| - \|\alpha\| \\
\|\pi\alpha^{-1}\| = \|\pi\| - \|\alpha\|
\end{gathered}$$

Finalmente $\alpha \mid \pi$.

□

5.3 Operações

Uma **operação** é uma permutação θ que, quando aplicada a um genoma π, gera o genoma $\pi\theta$.

Uma **reversão** de u a v é uma operação que inverte a ordem e a orientação dos elementos delimitados pelo caminho de u a v (excluindo u e incluindo v), realizando uma operação equivalente na fita complementar para manter a validade do genoma. A fórmula $\rho(u, v, \pi) = (u\ \Gamma\pi v)(v\ \Gamma\pi u)$ define a operação de reversão de u a v no genoma π quando u e v pertecem a uma mesma órbita de π. Observe que $\rho(u, v, \pi) = \rho(v, u, \pi) = \rho(\Gamma\pi u, \Gamma\pi v, \pi) = \rho(\Gamma\pi v, \Gamma\pi u, \pi)$.

Seja π é um genoma e ρ é uma reversão em π. Podemos afirmar que:

- $\|\pi\rho\| = \|\pi\|$
- Se $\rho = \rho(u, v, \pi)$ é uma reversão em π então $(u\ v) \mid \pi$ e $(u\ v) \nmid \pi\rho$, pois v muda de fita em relação a u

5.4 Distância de Reversão

O **problema da distância de reversão** consiste em encontrar o número mínimo de reversões que transformam um genoma em outro. Este número é a distância de reversão $d_r(\pi, \sigma)$ entre os genomas π e σ. Em outras palavras, nós queremos encontrar a seqüência de reversões $\rho_1\ \rho_2\ \ldots\ \rho_k$, tal que:

$$\sigma = \pi\rho_1\rho_2 \ldots \rho_{k-1}\rho_k$$

onde cada ρ_i, para $1 \leq i \leq k$, é uma reversão, ρ_{j+1}, para $1 \leq j < k$, é uma reversão no genoma $\pi\rho_1\rho_2 \ldots \rho_{j-1}\rho_j$ e k é mínimo.

O conceito de genoma está intrinsecamente ligado a permutação Γ e para que seja possível determinar a distância de reversão entre dois genomas é necessário que estes sejam compatíveis. A partir deste momento definiremos um **problema** pela tripla (π, σ, Γ), onde π é o genoma de origem, σ é o genoma de destino e ambos estão definidos sob Γ. Um problema é **monocromossomal** se π e σ são compostos por apenas um cromossomo e é **multicromossomal** caso contrário.

As permutações π, σ e Γ de um dado problema possuem conjuntos-bases iguais. Sendo assim, podemos nos referir a um conjunto-base de um problema, o qual será o conjunto-base de qualquer de suas permutações. Utilizaremos $E(P)$ para denotar o conjunto-base de um problema P e $n(P)$ para denotar o número de genes. Observe que $n(P) = |E(P)|/2$.

Podemos, alternativamente, utilizar $d_r(P)$ para nos referir a distância de reversão de um dado problema P. Note que, como desejamos transformar π em σ, as reversões sempre serão aplicadas a π. Sendo assim, podemos falar em aplicar uma reversão a um problema quando, na verdade, quisermos falar em aplicar a reversão ao genoma π do problema. Utilizaremos $P\rho$ para denotar a aplicação de uma reversão ρ a um problema P.

Lema 7. *Seja $P = (\pi, \sigma, \Gamma)$ um problema. Só é possível determinar a distância de reversão $d_r(P)$ caso para cada cromossomo em π exista um cromossomo em σ com o mesmo conjunto-base.*

Demonstração. Note que por definição uma reversão, quando aplicada, não altera o conjunto-base dos cromossomos de um genoma.

Digamos que pelo menos um dos cromossomos de π não possui um cromossomo em σ que possui o mesmo conjunto-base. Como a única operação de que dispomos é a reversão

e esta não possui a propriedade de alterar o conjunto-base dos cromossomos então nunca conseguiremos fazer $\pi = \sigma$. Tal fato nos leva a concluir que para que seja possível determinar a distância de reversão de P é necessário que para cada cromossomo em π exista um cromossomo correspondente (mesmo conjunto-base) em σ. □

Seja $P = (\pi, \sigma, \Gamma)$ um problema. Dizemos que P é **reversível** caso para cada cromossomo em π exista um cromossomo correspondente (mesmo conjunto-base) em σ e vice-versa. A partir deste momento todas as vezes que falarmos em problemas estaremos nos referindo a problemas reversíveis.

Teorema 1 (Particionamento de problema)**.** *Seja $P = (\pi, \sigma, \Gamma)$ um problema (reversível), tal que $\pi = \pi_1\pi_2 \dots \pi_n$, $\sigma = \sigma_1\sigma_2 \dots \sigma_n$, $\Gamma = \Gamma_1\Gamma_2 \dots \Gamma_n$ e Γ_1, Γ_2, ..., Γ_n possuam conjuntos-bases disjuntos. Se, para cada $1 \leq i \leq n$, π_i e σ_i são genomas compatíveis sob a permutação Γ_i então o problema P pode ser particionado nos problemas $P_1 = (\pi_1, \sigma_1, \Gamma_1)$, $P_2 = (\pi_2, \sigma_2, \Gamma_2)$, ..., $P_n = (\pi_n, \sigma_n, \Gamma_n)$ de forma que $d_r(P) = d_r(P_1) + d_r(P_2) + \ldots + d_r(P_n)$.*

Demonstração. Observe que reversões não alteram o conjunto-base dos cromossomos de um genoma. Cada cromossomo possui um conjunto-base disjunto do restante e cada cromossomo em π possui um cromossomo correspondente em σ. Sendo assim, podemos dizer que para cada cromossomo de π existe um conjunto de reversões que, quando aplicadas em uma determinada seqüência, torna este cromossomo igual ao seu cromossomo correspondente em σ.

As permutações π_1, π_2, ..., π_n são genomas disjuntos cujo produto $\pi_1\pi_2 \dots \pi_n$ resulta no genoma π, ou em outra palavras, π_1, π_2, ..., π_n contêm, cada um, um subconjunto dos cromossomos de π. As permutações σ_1, σ_2, ..., σ_n são genomas disjuntos cujo produto $\sigma_1\sigma_2 \dots \sigma_n$ resulta no genoma σ, ou em outra palavras, σ_1, σ_2, ..., σ_n contêm, cada um, um subconjunto dos cromossomos de σ.

Seja π_i compatível com σ_i, para $1 \leq i \leq n$. Como já sabemos que π_i e σ_i são compostos por subconjuntos de cromossomos de π e σ respectivamente, então o custo $d_r(\pi_i, \sigma_i)$ para ordenar π_i em relação a σ_i (fazer com que $\pi_i = \sigma_i$) é igual ao custo de tornar os cromossomos de π que também pertencem a π_i iguais aos seus cromossomos correspondentes em σ. □

O caráter circular das permutações implica em uma modelagem direta, através do formalismo, para genomas circulares. Porém, na natureza, são encontrados genomas circulares e lineares. Na Seção 5.5 serão comentadas as adaptações necessárias para o caso de genomas lineares.

5.5 Genomas Lineares

Quando o genoma é circular o formalismo algébrico aqui definido pode ser aplicado diretamente. Porém, quando o genoma em questão é linear, é necessário fazer um mapeamento para o circular e quando obtidos os resultados realizar um mapeamento de volta ao linear.

O mapeamento linear $\mapsto$ circular consiste na adição de um pseudo-gene $+p_i$, para cada i entre 1 e k, onde k é o número de cromossomos do genoma, para unir as extremidades de cada par de fitas do genoma linear. Para realizar o mapeamento circular $\mapsto$ linear, basta remover os pseudo-genes. A Figura 5.1 ilustra o mapeamento linear $\mapsto$ circular. Maiores informações podem ser encontradas nas referências [18, 30].

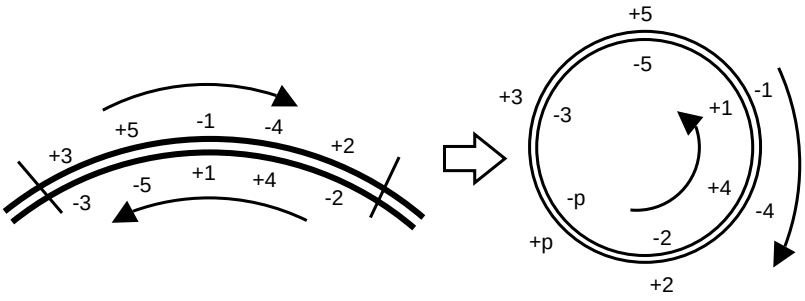


Figura 5.1: Mapeamento linear $\mapsto$ circular.

O genoma da Figura 5.1 é modelado em nosso formalismo algébrico pela seguinte permutação

$$(+3\ +5\ -1\ -4\ +2\ +p)(-p\ -2\ +4\ +1\ -5\ -3).$$

Capítulo 6

Uma solução algébrica para o problema da Distância de Reversão

Neste capítulo apresentamos uma solução formal para o problema da distância de reversão. Em tal solução empregamos um formalismo algébrico definido por Meidanis e Dias [17], que descrevemos no Capítulo 5.

Na Seção 6.1 apresentamos o quociente de um problema, que, como veremos, é importante na definição da distância de reversão. Duas das ferramentas que utilizaremos na solução do problema serão definidas nas Seções 6.2 e 6.3. Na primeira apresentamos o diagrama de pontos de quebra e na segunda apresentamos o grafo de entrelaçamento de arcos. Os conceitos de componentes, obstáculos e fortalezas são definidos na Seção 6.4. Finalmente, na Seção 6.5 apresentamos o algoritmo para solução da distância de reversão assim como uma demonstração formal de sua corretude.

6.1 Quociente

Seja $P = (\pi, \sigma, \Gamma)$ um problema. O **quociente** de P, definido pelo produto $\pi^{-1}\sigma$, fornece uma permutação onde para cada ciclo α, da sua decomposição em ciclos disjuntos, temos um ciclo companheiro de α, dado por $(\Gamma\pi) \bullet \alpha^{-1}$, em $\pi^{-1}\sigma$. Em um quociente cada par de ciclos companheiros pode ser visto como uma unidade e a partir deste momento este binômio será chamado de **par de ciclos**. O **número de pares de ciclos** (companheiros) em $\pi^{-1}\sigma$ é denotado por $c(P)$. Por exemplo, seja o problema $P_1 = (\pi_1, \sigma_1, \Gamma_1)$, onde

$$\begin{aligned}\pi_1 = (&+0\ +7\ +1\ -5\ -3\ -4\ -2\ +6\ +8\ +10\ +9\ +11\ +14\ +13\ +12)\\ &(-12\ -13\ -14\ -11\ -9\ -10\ -8\ -6\ +2\ +4\ +3\ +5\ -1\ -7\ -0),\\ \sigma_1 = (&+0\ +1\ +2\ +3\ +4\ +5\ +6\ +7\ +8\ +9\ +10\ +11\ +12\ +13\ +14)\\ &(-14\ -13\ -12\ -11\ -10\ -9\ -8\ -7\ -6\ -5\ -4\ -3\ -2\ -1\ -0)\ \text{e}\end{aligned}$$

$$\Gamma_1 = (+0\ -0)\ (+1\ -1)\ (+2\ -2)\ (+3\ -3)\ (+4\ -4)\ (+5\ -5)\ (+6\ -6)\ (+7\ -7)\ (+8\ -8)$$
$$(+9\ -9)\ (+10\ -10)\ (+11\ -11)\ (+12\ -12)\ (+13\ -13)\ (+14\ -14).$$

Segue que:

$$\pi_1^{-1}\sigma_1 = (-13\ -0)\ (-8\ -1\ -7)\ (-2\ +5)\ (+2\ +4\ +3)\ (+1\ -6)\ (-11\ -9\ -10)$$
$$(-12\ -14)\ (+14\ +12)\ (+11\ +13)\ (+8\ +10\ +9)\ (+0\ +7\ +6)\ (-5\ -3\ -4).$$

e, assim, temos $c(P_1) = 6$. O número de pares de ciclos (companheiros) de um problema $P = (\pi, \sigma, \Gamma)$ pode ser obtido através da Equação 6.1.

$$c(P) = \frac{orb(\pi^{-1}\sigma)}{2} = \frac{|E(P)| - \|\pi^{-1}\sigma\|}{2} \tag{6.1}$$

Note que, como $n(P) = |E(P)|/2$, a Equação 6.1 pode ser escrita também como:

$$\frac{\|\pi^{-1}\sigma\|}{2} = n(P) - c(P)$$

Seja α um par de ciclos. O **tamanho** do par de ciclos $\alpha = \alpha_1\alpha_2$, denotado por $l(\alpha)$, é $|Supp(\alpha)|/2$. Se $l(\alpha) = 0$ dizemos que α é um par de ciclos **trivial**.

Observe que nosso objetivo é fazer com que $\pi = \sigma$ e assim, pelas propriedades de permutações, podemos dizer que nosso objetivo é obter um problema onde o quociente $\pi^{-1}\sigma$ seja a identidade, ou ainda, onde há apenas pares de ciclos triviais.

O lema apresentado a seguir mostra o que pode ocorrer com o número de órbitas do quociente do problema quando aplicamos uma reversão.

Lema 8. *Seja $P = (\pi, \sigma, \Gamma)$ um problema. Se ρ é uma reversão em π, então $orb(\rho\pi^{-1}\sigma) = orb(\pi^{-1}\sigma) + x$, onde $x \in \{-2, 0, +2\}$.*

Demonstração. Seja $\rho = \rho(u, v, \pi) = (u\ \Gamma\pi v)(v\ \Gamma\pi u)$. Faremos uma análise em três casos da aplicação da reversão ρ sob o quociente $\pi^{-1}\sigma$.

1. **u e v estão no mesmo ciclo de $\pi^{-1}\sigma$:** Seja α_1 o ciclo ao qual u e v pertencem, e α_2 o ciclo companheiro de α_1. Como $\{u, v\} \subset Supp(\alpha_1)$, então $\{\Gamma\pi u, \Gamma\pi v\} \subset Supp(\alpha_2)$. No produto $(v\ \Gamma\pi u)\pi^{-1}\sigma$ os ciclos α_1 e α_2 são unidos num único ciclo γ, e no produto $\rho\pi^{-1}\sigma$, o ciclo γ é novamente separado em dois ciclos. Desta forma temos que $orb(\rho\pi^{-1}\sigma) = orb(\pi^{-1}\sigma)$.

2. **u e v estão em ciclos distintos porém companheiros de $\pi^{-1}\sigma$:** Sejam α_1 e α_2 ciclos companheiros em $\pi^{-1}\sigma$. Se $u \in Supp(\alpha_1)$ e $v \in Supp(\alpha_2)$, então podemos afirmar que $\Gamma\pi v \in Supp(\alpha_1)$ e $\Gamma\pi u \in Supp(\alpha_2)$. A reversão $\rho(u, v, \pi)$ é composta

por duas órbitas $A = \{u, \Gamma\pi v\}$ e $B = \{v, \Gamma\pi u\}$. De acordo com o que observamos acima, podemos dizer que $A \subset Supp(\alpha_1)$ e $B \subset Supp(\alpha_2)$. Desta forma, no produto $\rho\pi^{-1}\sigma$ teremos que α_1 e α_2 serão quebrados em dois ciclos cada, e assim $orb(\rho\pi^{-1}\sigma) = orb(\pi^{-1}\sigma) + 2$.

3. u **e** v **estão em ciclos distintos e não companheiros de** $\pi^{-1}\sigma$**:** Sejam α_1 e α_2 ciclos não companheiros em $\pi^{-1}\sigma$. Se $u \in Supp(\alpha_1)$ e $v \in Supp(\alpha_2)$, então podemos afirmar que $\Gamma\pi u \in (\Gamma\pi) \bullet \alpha_1^{-1} = \alpha_3$ e $\Gamma\pi v \in (\Gamma\pi) \bullet \alpha_2^{-1} = \alpha_4$. Note que α_3 é o ciclo companheiro de α_1 e α_4 é o companheiro de α_2. Sabemos que $\rho(u, v, \pi) = (u\ \Gamma\pi v)(v\ \Gamma\pi u)$ é um produto de dois 2-ciclos disjuntos e que cada elemento do suporte de ρ está em uma órbita de $\pi^{-1}\sigma$. Segue que no produto $\rho\pi^{-1}\sigma$ as órbitas que contêm u e $\Gamma\pi v$ em $\pi^{-1}\sigma$ são unidas em uma única órbita de $\rho\pi^{-1}\sigma$, assim como as órbitas que contêm v e $\Gamma\pi u$ em $\pi^{-1}\sigma$ são unidas em uma única órbita de $\rho\pi^{-1}\sigma$. Desta forma temos que $orb(\rho\pi^{-1}\sigma) = orb(\pi^{-1}\sigma) - 2$.

□

Seja $P = (\pi, \sigma, \Gamma)$ um problema e ρ uma reversão em π. O **incremento no número pares de ciclos** $c(P\rho) - c(P)$ é denotado por Δc. O Lema 8 nos leva a classificar reversões em três categorias. Dizemos que a reversão ρ é uma **junção** caso $\Delta c = -1$, uma **recombinação** caso $\Delta c = 0$ ou uma **quebra** caso $\Delta c = 1$. Uma reversão ρ **atua** em um par de ciclos α, quando $Supp(\rho) \subset Supp(\alpha)$, ou seja, quando for uma recombinação ou uma quebra. No exemplo anterior temos que $\rho(+1, -2, \pi_1)$ é uma quebra, $\rho(-8, -1, \pi_1)$ é uma recombinação e $\rho(+11, +14, \pi_1)$ é uma junção.

Deriva imediatamente do Lema 8 o seguinte corolário:

Corolário 1. *Seja $P = (\pi, \sigma, \Gamma)$ um problema. Se ρ é uma reversão em π então podemos dizer que $\Delta c \in \{-1, 0, +1\}$.*

Conforme observamos, nosso objetivo é fazer com que $\pi^{-1}\sigma$ seja composta apenas por ciclos triviais, podemos, assim, dizer que nosso objetivo é "quebrar" os pares de ciclos não triviais até que só restem pares de ciclos triviais. Neste contexto, as quebras surgem como uma boa ferramenta. Um par de ciclos γ é **bom** caso exista uma quebra de ciclos ρ que atua em γ e é **ruim**, caso contrário. No exemplo P_1, temos apenas um par de ciclos bom, dado por $(-2\ +5)(+1\ -6)$.

Observações no Lema 8 nos permite afirmar que serão necessárias $\|\pi^{-1}\sigma\|/2$ reversões, no mínimo, para resolver o problema $P = (\pi, \sigma, \Gamma)$ conforme demonstramos no Lema 9.

Lema 9. *Se $P = (\pi, \sigma, \Gamma)$ é um problema então $d_r(P) \geq \|\pi^{-1}\sigma\|/2$.*

Demonstração. Objetivamos partir de P e encontrar o número mínimo de reversões que, aplicadas seqüencialmente, transformem P em $P_f = (\pi_f, \sigma, \Gamma)$ tal que $\pi_f^{-1}\sigma = 1$. O Lema 8 nos mostra que, na melhor das hipóteses, uma reversão ρ em P incrementa $orb(\pi^{-1}\sigma)$ de duas unidades. Sabemos que $\|\alpha\| = |E| - orb(\alpha)$, para α uma permutação qualquer, onde E é o conjunto-base de α. Segue que:

$$\begin{array}{rcl} orb(\rho\pi^{-1}\sigma) & \leq & orb(\pi^{-1}\sigma) + 2 \\ |E| - \|\rho\pi^{-1}\sigma\| & \leq & |E| - \|\pi^{-1}\sigma\| + 2 \\ - \|\rho\pi^{-1}\sigma\| & \leq & - \|\pi^{-1}\sigma\| + 2 \\ \|\rho\pi^{-1}\sigma\| & \geq & \|\pi^{-1}\sigma\| - 2 \end{array}$$

para uma reversão qualquer ρ em P. Então podemos dizer que, na melhor das hipóteses, uma reversão reduz a norma do quociente de um problema em duas unidades. Tal observação nos permite afirmar que $d_r \geq \|\pi^{-1}\sigma\|/2$. □

A seguir apresentamos um problema e o resolvemos passo-a-passo. É importante salientar que um problema não possui necessariamente apenas uma solução. No momento não entraremos em detalhes sobre como é feita a escolha das reversões.

Seja o problema $P_2 = (\pi_2, \sigma_2, \Gamma_2)$ com

$$P_2 = \left\{ \begin{array}{rcl} \pi_2 & = & (+0\ -1\ -3\ +2)(-2\ +3\ +1\ -0) \\ \sigma_2 & = & (+0\ +1\ +2\ +3)(-3\ -2\ -1\ -0) \\ \Gamma_2 & = & (+0\ -0)(+1\ -1)(+2\ -2)(+3\ -3) \\ \pi_2^{-1}\sigma_2 & = & (+0\ +3\ +2\ -2)(-0\ -1\ +1\ -3) \end{array} \right.$$

Há quebras em P_2, tal como $\rho_1 = \rho(+0, -1, \pi_2)$, que se aplicada resulta em:

$$P_2\rho_1 = \left\{ \begin{array}{rcl} \pi_2\rho_1 & = & (+0\ +1\ -3\ +2)(-2\ +3\ -1\ -0) \\ \sigma_2 & = & (+0\ +1\ +2\ +3)(-3\ -2\ -1\ -0) \\ \Gamma_2 & = & (+0\ -0)(+1\ -1)(+2\ -2)(+3\ -3) \\ (\pi_2\rho_1)^{-1}\sigma_2 & = & (+1\ -3\ -0)(+2\ -2\ +3)(-1)(+0) \end{array} \right.$$

Em $P_2\rho_1$ também há quebras, tal como $\rho_2 = \rho(+2, -3, \pi_2\rho_1)$, que quando aplicada resulta em:

$$P_2\rho_1\rho_2 = \left\{ \begin{array}{rcl} \pi_2\rho_1\rho_2 & = & (+0\ +1\ -3\ -2)(+2\ +3\ -1\ -0) \\ \sigma_2 & = & (+0\ +1\ +2\ +3)(-3\ -2\ -1\ -0) \\ \Gamma_2 & = & (+0\ -0)(+1\ -1)(+2\ -2)(+3\ -3) \\ (\pi_2\rho_1\rho_2)^{-1}\sigma_2 & = & (+1\ -0)(+3\ -2)(+0)(-1)(-3)(+2) \end{array} \right.$$

Finalmente, em $P_2\rho_1\rho_2$ há apenas a quebra $\rho_3 = \rho(+3, -0, \pi_2\rho_1\rho_2)$, que quando aplicada resulta em:

$$P_2\rho_1\rho_2\rho_3 = \begin{cases} \pi_2\rho_1\rho_2\rho_3 &= (+0\ +1\ +2\ +3)(-3\ -2\ -1\ -0) \\ \sigma_2 &= (+0\ +1\ +2\ +3)(-3\ -2\ -1\ -0) \\ \Gamma_2 &= (+0\ -0)(+1\ -1)(+2\ -2)(+3\ -3) \\ (\pi_2\rho_1\rho_2\rho_3)^{-1}\sigma_2 &= (+0)(+1)(+2)(+3)(-3)(-2)(-1)(-0) \end{cases}$$

Observe que $(\pi_2\rho_1\rho_2\rho_3)^{-1}\sigma_2 = 1$ e $\pi_2\rho_1\rho_2\rho_3 = \sigma_2$. Observe também que $\left\|\pi_2^{-1}\sigma_2\right\| = 6$ e $d_r(P_2) \geq 3$, assim concluímos que o problema foi resolvido. Note também que nem todo problema possui quebras e, mesmo nos que possuem, é necessário escolher as quebras com cuidado. Por exemplo, se tivéssemos aplicado $\rho_4 = \rho(+1, -2, \pi_2)$ a P_2 teríamos:

$$P_2\rho_4 \begin{cases} \pi_2\rho_4 &= (+0\ +3\ +1\ +2)(-2\ -1\ -3\ -0) \\ \sigma_2 &= (+0\ +1\ +2\ +3)(-3\ -2\ -1\ -0) \\ \Gamma_2 &= (+0\ -0)(+1\ -1)(+2\ -2)(+3\ -3) \\ (\pi_2\rho_4)^{-1}\sigma_2 &= (+0\ +3\ +2)(-0\ -1\ -3)(+1)(-2) \end{cases}$$

onde não há quebras.

No decorrer deste capítulo mostraremos como fazer a seleção correta das quebras, assim como as formas de encontrar reversões, que levem a solução, na ausência de quebras.

6.2 Diagrama de pontos de quebra

O diagrama de pontos de quebra é uma ferramenta que permite uma visualização gráfica do problema. Ele permite verificar o estágio atual do problema (realidade) e comparar com o que desejamos alcançar. Alguns autores, tais como Setubal e Meidanis [26] chamam tal diagrama de realidade-desejo. Antes da definição do diagrama são necessárias algumas definições, que apresentamos a seguir.

Seja π um genoma e π_1 e π_2 as fitas de π. As permutações $\Gamma\pi_1$ e $\Gamma\pi_2$ definem os **anéis** de π. Note que π_1 e π_2 são fitas complementares e $\Gamma\pi_1 = (\Gamma\pi_2)^{-1}$. Um anel $\theta(\pi)$ de um genoma π sempre apresenta $orb(\theta(\pi)) = 1$ e $Supp(\theta(\pi)) = E(\pi)$.

Seja $P = (\pi, \sigma, \Gamma)$ um problema. Uma **aresta preta** é um 2-ciclo que divide $\Gamma\pi$. Um **arco** ou **aresta cinza** é um 2-ciclo que divide $\Gamma\sigma$. Um elemento $b \in E(P)$ é um **ponto de quebra** se $\pi b \neq \sigma b$. Um ponto de quebra indica uma posição de π que não está em concordância com σ, ou seja, é um ponto que deve sofrer alteração para igualar π a σ. Os pontos de quebra sempre ocorrem aos pares, de forma que se b é um ponto de quebra então $\Gamma\pi b$ também é um ponto de quebra, só que na fita complementar a fita de b. Os pares de pontos de quebra definem o 2-ciclo $(b\ \Gamma\pi b)$.

Note que toda aresta preta α tal que $\alpha \nmid \Gamma\sigma$ é um par de pontos de quebra. Tal observação parte do fato de que tanto arestas pretas quanto pares de pontos de quebra

assumem a forma $(b\ \Gamma\pi b)$ e que se $\pi b \neq \sigma b$ então $\Gamma\pi b \neq \Gamma\sigma b$ e, finalmente, chegamos a conclusão de que $(b\ \Gamma\pi b) \nmid \Gamma\sigma$.

Seja $P = (\pi, \sigma, \Gamma)$ um problema. O **diagrama de pontos de quebra** de P, denotado por $B(P)$, é construído da seguinte forma:

- Seja $\theta(\pi)$ um dos anéis de π. Distribua os elementos de $Supp(\theta(\pi))$ em um ciclo mantendo a ordem definida por $\theta(\pi)$. Pelas propriedades do diagrama os dois anéis terão o mesmo efeito.
- Para cada aresta preta de P, una os elementos do seu suporte seguindo o contorno do ciclo.
- Para cada arco de P, una os elementos do seu suporte.

Para exemplificar, iremos montar o diagrama de pontos de quebra do problema P_2. Inicialmente calculamos anel, arestas pretas e arestas cinzas.

$$\theta(\pi_2) = (+0\ +1\ -1\ +3\ -3\ -2\ +2\ -0)$$
$$\Gamma_2\pi_2 = (+0\ +1)(-1\ +3)(-3\ -2)(+2\ -0)$$
$$\Gamma_2\sigma_2 = (+0\ -1)(+1\ -2)(+2\ -3)(+3\ -0)$$

Em seguida construímos o diagrama, que é apresentado na Figura 6.1.

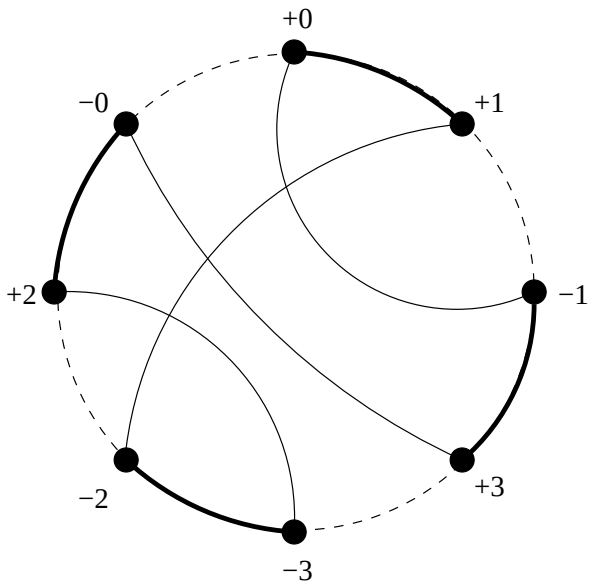


Figura 6.1: Diagrama de pontos de quebra do problema P_2.

Note que quando o problema está resolvido temos $\Gamma\pi = \Gamma\sigma$ e, assim, a cada aresta preta temos uma aresta cinza pareada. Neste contexto, também podemos ver o problema como compatibilizar arestas pretas e arestas cinzas (arcos) e assim passaremos a utilizar os arcos como guia na definição das reversões.

No restante desta seção iremos nos concentrar nos arcos. Inicialmente apresentaremos o conceito de orientação de um arco e mostraremos a reversão que um arco induz, sempre objetivando compatibilizar arcos e arestas pretas. Posteriormente iremos apresentar os conceitos de entrelaçamento e intercalação de arcos. Tais conceitos focam no relacionamento entre os arcos mostrando, por exemplo, arcos que serão afetados com a aplicação de uma reversão. Iniciaremos ainda uma análise dos efeitos das reversões.

6.2.1 Arcos: orientação e reversões

Seja $P = (\pi, \sigma, \Gamma)$ um problema e α um arco em P. Dizemos que α é **orientado** caso $\alpha \mid \pi$ e é **não orientado** caso contrário. Conforme veremos a seguir no Lema 10, os arcos orientados induzem quebras e, como já havíamos comentado anteriormente, as quebras são reversões do tipo mais eficiente no que se refere a reduzir a norma do quociente.

Seja $P = (\pi, \sigma, \Gamma)$ um problema, α um arco em P e $v \in \mathit{Supp}(\alpha)$. O arco α induz uma reversão, que denotamos por $\rho(\alpha, \pi)$, definida como segue:

- $\rho(\alpha, \pi) = \rho(v, \Gamma\sigma v, \pi) = ((\Gamma\pi) \wedge (\Gamma\pi\alpha\Gamma\pi\alpha\Gamma\pi))(\Gamma\pi\alpha\Gamma\pi\alpha\Gamma\pi) = (\Gamma\pi\alpha\Gamma\pi\alpha\Gamma\pi)((\Gamma\pi) \wedge (\Gamma\pi\alpha\Gamma\pi\alpha\Gamma\pi))$, para α orientado.
- $\rho(\alpha, \pi) = \rho(\Gamma\pi v, \Gamma\sigma v, \pi) = \alpha\Gamma\pi\alpha\Gamma\pi = \Gamma\pi\alpha\Gamma\pi\alpha$, para α não orientado.

Observe que, quando α é orientado, $\alpha \mid \pi$ e assim v e $\Gamma\sigma v$ estão na mesma fita de π e definem uma reversão. Já no caso em que α é não orientado, v e $\Gamma\sigma v$ estão em fitas complementares de π. Desta forma, podemos afirmar que $\Gamma\pi v$ e $\Gamma\sigma v$ estão na mesma fita e definem uma reversão.

Quanto às fórmulas, observe que no caso em que α é não orientado já temos um dos 2-ciclos que definem a reversão e daí basta fazer o produto com o conjugado de α por $\Gamma\pi$ que resulta em $\alpha(\Gamma\pi)\bullet\alpha = \alpha\Gamma\pi\alpha(\Gamma\pi)^{-1} = \alpha\Gamma\pi\alpha\Gamma\pi$. No caso em que α é orientado temos um 2-ciclo que não é um dos que definem a reversão. Ao fazermos o produto $\alpha\Gamma\pi\alpha\Gamma\pi$ obtemos dois 2-ciclos com todos os elementos que definem a reversão, mas ainda não estão no posicionamento correto. Quando fazemos o produto $\Gamma\pi\alpha\Gamma\pi\alpha\Gamma\pi$ conseguimos o posicionamento correto dos elementos, mas agora temos outros 2-ciclos que nada têm a ver com a reversão. O produto com o máximo divisor comum elimina tais ciclos indesejados.

Lema 10. *Seja $P = (\pi, \sigma, \Gamma)$ um problema. Se α é um arco orientado em P então a reversão $\rho(\alpha, \pi)$ induzida por α em π é uma quebra.*

Demonstração. Seja $v \in Supp(\alpha)$. Sendo α orientado então α induz a reversão $\rho(\alpha,\pi) = (v\ \pi^{-1}\sigma v)(\Gamma\sigma v\ \Gamma\pi v)$. Seja $u = \pi^{-1}\sigma v$. Segue que:

$$\begin{aligned}\rho(\alpha,\pi) &= (v\ \pi^{-1}\sigma v)(\Gamma\sigma v\ \Gamma\pi v)\\ &= (\sigma^{-1}\pi u\ u)(\Gamma\pi u\ \Gamma\pi\sigma^{-1}\pi u)\\ &= (\sigma^{-1}\pi u\ u)(\Gamma\pi u\ \pi^{-1}\Gamma\sigma^{-1}\pi u)\\ &= (\sigma^{-1}\pi u\ u)(\Gamma\pi u\ \pi^{-1}\sigma\Gamma\pi u)\end{aligned}$$

Observe que u é o elemento mapeado por $\sigma^{-1}\pi u$ em $\pi^{-1}\sigma$. Observe também que $\Gamma\pi u$ é um elemento do ciclo companheiro ao ciclo que contém u e $\pi^{-1}\sigma\Gamma\pi u$ é o elemento mapeado por $\Gamma\pi u$ em $\pi^{-1}\sigma$. Tais observações permite-nos afirmar que $\rho(\alpha,\pi) \mid \pi^{-1}\sigma$, $\Delta c = 1$ e, finalmente, chegamos a conclusão de que $\rho(\alpha,\pi)$ é uma quebra. □

Os Lemas 11, 12 e 13 a seguir mostram um pouco do efeito das reversões e seus resultados serão utilizados nas análises que faremos na seção seguinte.

Lema 11. *Seja $P = (\pi,\sigma,\Gamma)$ um problema e α um arco em P. Se $\rho(\alpha,\pi)$ é a reversão induzida por α no genoma π então $\alpha \nmid \pi\rho(\alpha,\pi)$.*

Demonstração. Faremos uma demonstração em dois casos conforme segue:

1. **O arco α é orientado:** seja $\alpha = (v\ \Gamma\sigma v)$. Como $\alpha \mid \pi$ então $\rho(\alpha,\pi) = \rho(v,\Gamma\sigma v,\pi)$. Segue das propriedades de reversões que $\alpha \nmid \pi\rho(\alpha,\pi)$.

2. **O arco α é não orientado:** como $\alpha \nmid \pi$ então $\rho(\alpha,\pi) = \Gamma\pi\alpha\Gamma\pi\alpha$. Segue que:

$$\begin{aligned}\|\pi\rho(\alpha,\pi)\alpha^{-1}\| &= \|\pi\Gamma\pi\alpha\Gamma\pi\alpha\alpha^{-1}\|\\ &= \|\Gamma\pi^{-1}\pi\alpha\Gamma\pi\|\\ &= \|\Gamma\alpha\Gamma^{-1}\pi\|\\ &= \|(\Gamma\bullet\alpha)\pi\|\end{aligned}$$

   Sabe-se que α é um 2-ciclo e $\alpha \nmid \pi$. Sabe-se também que, pelo fato de π ser um genoma sob Γ, se β é um ciclo em π então $\Gamma\bullet\beta^{-1}$ também é um ciclo em π. Desta forma podemos afirmar que $(\Gamma\bullet\alpha) \nmid \pi$. Pela última observação podemos afirmar que $\|\pi\rho(\alpha,\pi)\alpha^{-1}\| = \|\pi\| + 1$. Por outro lado $\|\pi\rho(\alpha,\pi)\| = \|\pi\|$ e então $\|\pi\rho(\alpha,\pi)\| - \|\alpha\| = \|\pi\| - 1$. Assim concluímos que $\alpha \nmid \pi\rho(\alpha,\pi)$.

□

Lema 12. *Seja $P = (\pi,\sigma,\Gamma)$ um problema, α um arco em P e $\rho(\alpha,\pi)$ a reversão induzida por α em π. Seja $\rho(\alpha,\pi\rho(\alpha,\pi))$ a reversão induzida por α em $\pi\rho(\alpha,\pi)$. Se $\alpha \nmid \pi$ então $\pi\rho(\alpha,\pi)\rho(\alpha,\pi\rho(\alpha,\pi)) = \pi$.*

Demonstração. Como $\alpha \nmid \pi$ então $\rho(\alpha,\pi) = \alpha\Gamma\pi\alpha\Gamma\pi = \Gamma\pi\alpha\Gamma\pi\alpha$. Pelo Lema 11 podemos afirmar que $\alpha \nmid \pi\rho(\alpha,\pi)$ e assim:

$$\begin{aligned}\rho(\alpha,\pi\rho(\alpha,\pi)) &= \alpha\Gamma(\pi\Gamma\pi\alpha\Gamma\pi\alpha)\alpha\Gamma(\pi\Gamma\pi\alpha\Gamma\pi\alpha)\\ &= \alpha(\Gamma\pi)(\Gamma\pi)\alpha\Gamma\pi\alpha\alpha(\Gamma\pi)(\Gamma\pi)\alpha\Gamma\pi\alpha\\ &= \alpha\alpha\Gamma\pi\alpha\Gamma\pi\alpha\\ &= \Gamma\pi\alpha\Gamma\pi\alpha\end{aligned}$$

E desta forma:

$$\begin{aligned}\pi\rho(\alpha,\pi)\rho(\alpha,\pi\rho(\alpha,\pi)) &= \pi(\alpha\Gamma\pi\alpha\Gamma\pi)(\Gamma\pi\alpha\Gamma\pi\alpha)\\ &= \pi\alpha\Gamma\pi\alpha(\Gamma\pi)(\Gamma\pi)\alpha\Gamma\pi\alpha\\ &= \pi\alpha\Gamma\pi\alpha\alpha\Gamma\pi\alpha\\ &= \pi\alpha(\Gamma\pi)(\Gamma\pi)\alpha\\ &= \pi\alpha\alpha\\ &= \pi\end{aligned}$$

□

Lema 13. *Seja $P = (\pi, \sigma, \Gamma)$ um problema, α um arco em P e $\rho(\alpha,\pi)$ a reversão induzida por α em π. Seja $\rho(\alpha,\pi\rho(\alpha,\pi))$ a reversão induzida por α em $\pi\rho(\alpha,\pi)$. Se $\alpha \mid \pi$ então $\pi\rho(\alpha,\pi)\rho(\alpha,\pi\rho(\alpha,\pi)) = \pi\rho(\alpha,\pi)$.*

Demonstração. Como $\alpha \mid \pi$ então:

$$\rho(\alpha,\pi) = ((\Gamma\pi) \wedge (\Gamma\pi\alpha\Gamma\pi\alpha\Gamma\pi))(\Gamma\pi\alpha\Gamma\pi\alpha\Gamma\pi) = (\Gamma\pi\alpha\Gamma\pi\alpha\Gamma\pi)((\Gamma\pi) \wedge (\Gamma\pi\alpha\Gamma\pi\alpha\Gamma\pi))$$

Pelo Lema 11 podemos afirmar que $\alpha \nmid \pi\rho(\alpha,\pi)$ e assim:

$$\begin{aligned}\rho(\alpha,\pi\rho(\alpha,\pi)) &= \Gamma\pi((\Gamma\pi) \wedge (\Gamma\pi\alpha\Gamma\pi\alpha\Gamma\pi))(\Gamma\pi\alpha\Gamma\pi\alpha\Gamma\pi)\alpha\Gamma\pi\\ &\quad (\Gamma\pi\alpha\Gamma\pi\alpha\Gamma\pi)((\Gamma\pi) \wedge (\Gamma\pi\alpha\Gamma\pi\alpha\Gamma\pi))\alpha\\ &= \Gamma\pi((\Gamma\pi) \wedge (\Gamma\pi\alpha\Gamma\pi\alpha\Gamma\pi))\Gamma\pi\alpha\Gamma\pi\alpha\Gamma\pi\alpha(\Gamma\pi)(\Gamma\pi)\alpha\Gamma\pi\alpha\Gamma\pi\\ &\quad ((\Gamma\pi) \wedge (\Gamma\pi\alpha\Gamma\pi\alpha\Gamma\pi))\alpha\\ &= \Gamma\pi((\Gamma\pi) \wedge (\Gamma\pi\alpha\Gamma\pi\alpha\Gamma\pi))\Gamma\pi\alpha\Gamma\pi\alpha\Gamma\pi\alpha\alpha\Gamma\pi\alpha\Gamma\pi\\ &\quad ((\Gamma\pi) \wedge (\Gamma\pi\alpha\Gamma\pi\alpha\Gamma\pi))\alpha\\ &= \Gamma\pi((\Gamma\pi) \wedge (\Gamma\pi\alpha\Gamma\pi\alpha\Gamma\pi))\Gamma\pi\alpha\Gamma\pi\alpha(\Gamma\pi)(\Gamma\pi)\alpha\Gamma\pi\\ &\quad ((\Gamma\pi) \wedge (\Gamma\pi\alpha\Gamma\pi\alpha\Gamma\pi))\alpha\\ &= \Gamma\pi((\Gamma\pi) \wedge (\Gamma\pi\alpha\Gamma\pi\alpha\Gamma\pi))\Gamma\pi\alpha\Gamma\pi\alpha\alpha\Gamma\pi\\ &\quad ((\Gamma\pi) \wedge (\Gamma\pi\alpha\Gamma\pi\alpha\Gamma\pi))\alpha\\ &= \Gamma\pi((\Gamma\pi) \wedge (\Gamma\pi\alpha\Gamma\pi\alpha\Gamma\pi))\Gamma\pi\alpha(\Gamma\pi)(\Gamma\pi)\\ &\quad ((\Gamma\pi) \wedge (\Gamma\pi\alpha\Gamma\pi\alpha\Gamma\pi))\alpha\\ &= \Gamma\pi((\Gamma\pi) \wedge (\Gamma\pi\alpha\Gamma\pi\alpha\Gamma\pi))\Gamma\pi\alpha((\Gamma\pi) \wedge (\Gamma\pi\alpha\Gamma\pi\alpha\Gamma\pi))\alpha\end{aligned}$$

Seja $\alpha = (u\ v)$. Pelo enunciado, $\alpha \mid \pi$, sendo assim u, v, $\Gamma\pi u$, $\Gamma\pi v$ são todos distintos. Observe que o produto $\Gamma\pi((\Gamma\pi) \wedge (\Gamma\pi\alpha\Gamma\pi\alpha\Gamma\pi))$ resulta em $(u\ \Gamma\pi u)(v\ \Gamma\pi v)$. O produto $(u\ \Gamma\pi u)(v\ \Gamma\pi v)\Gamma\pi$, por sua vez, resulta em $((\Gamma\pi) \wedge (\Gamma\pi\alpha\Gamma\pi\alpha\Gamma\pi))$ e nos resta $((\Gamma\pi) \wedge (\Gamma\pi\alpha\Gamma\pi\alpha\Gamma\pi))\alpha((\Gamma\pi) \wedge (\Gamma\pi\alpha\Gamma\pi\alpha\Gamma\pi))\alpha$. Agora observe que $((\Gamma\pi) \wedge (\Gamma\pi\alpha\Gamma\pi\alpha\Gamma\pi))$ e α são disjuntos e $((\Gamma\pi) \wedge (\Gamma\pi\alpha\Gamma\pi\alpha\Gamma\pi))$ é um produto de 2-ciclos. Com isso temos $((\Gamma\pi) \wedge (\Gamma\pi\alpha\Gamma\pi\alpha\Gamma\pi))\alpha((\Gamma\pi) \wedge (\Gamma\pi\alpha\Gamma\pi\alpha\Gamma\pi))\alpha = 1$.

Concluímos assim que $\rho(\alpha, \pi\rho(\alpha, \pi)) = 1$ e, finalmente, $\pi\rho(\alpha, \pi)\rho(\alpha, \pi\rho(\alpha, \pi)) = \pi\rho(\alpha, \pi)$.

□

6.2.2 Entrelaçamento de arcos

Sejam $P = (\pi, \sigma, \Gamma)$ um problema, ρ uma reversão em π, α e β arcos em P e $\rho(\alpha, \pi)$ a reversão induzida por α em π. Dizemos que ρ e α se **entrelaçam** caso α possua orientações distintas em π e $\pi\rho$. Dizemos que β **entrelaça** α caso β possua orientações distintas em π e $\pi\rho(\alpha, \pi)$.

Seja α uma permutação e $A \subset E(\alpha)$. A **subtração** de A em α resulta na permutação $\alpha \setminus A$ tal que $\alpha \setminus A \mid \alpha$, $|E(\alpha \setminus A)| = |E(\alpha)| - |A|$ e se $x \in A$ então $x \notin E(\alpha \setminus A)$.

Seja $P = (\pi, \sigma, \Gamma)$ um problema, G um conjunto de genes de P e ΓG o conjunto dos complementos dos genes em G. A **subtração** de G em P resulta no problema $P \setminus G = (\pi \setminus (G \cup \Gamma G), \sigma \setminus (G \cup \Gamma G), \Gamma \setminus (G \cup \Gamma G))$.

Seja $P = (\pi, \sigma, \Gamma)$ um problema e α e β arcos distintos de P. Dizemos que α e β se **intercalam** quando existe um anel θ de π tal que

$$(x\ y\ \Gamma\sigma x\ \Gamma\sigma y) \mid \theta$$

para $x \in \mathit{Supp}(\alpha)$ e $y \in \mathit{Supp}(\beta)$.

Os Lemas 14, 15 e 16 a seguir analisam o efeito da subtração sobre os arcos com suas orientações e relações de intercalação.

Lema 14. *Seja $P = (\pi, \sigma, \Gamma)$ um problema e G um conjunto de genes em P. Se $\alpha = (x\ \Gamma\sigma x)$ é um arco de P e $\mathit{Supp}(\alpha) \cup \{\sigma x\} \subset E(P \setminus G)$ então α é arco de $P \setminus G$.*

Demonstração. De acordo com o enunciado podemos afirmar que $\{x, \Gamma x, \sigma x, \Gamma\sigma x\} \subset E(P \setminus G)$. Observe que $\sigma \setminus (G \cup \Gamma G)x = \sigma x$, assim, $\Gamma \setminus (G \cup \Gamma G)\sigma \setminus (G \cup \Gamma G)x = \Gamma\sigma x$ e, desta forma, α é arco também em $P \setminus G$. □

Lema 15. *Seja $P = (\pi, \sigma, \Gamma)$ um problema e G um conjunto de genes em P. Se $\alpha = (x\ \Gamma\sigma x)$ é um arco de P e $\mathit{Supp}(\alpha) \cup \{\sigma x\} \subset E(P \setminus G)$ então α é arco em $P \setminus G$ e sua orientação é a mesma em P e $P \setminus G$.*

Demonstração. O Lema 14 nos garante que α é arco em $P \setminus G$.

Pelo enunciado temos que $Supp(\alpha) \subset E(P \setminus G)$ e, de acordo com a definição de subtração, podemos afirmar que $\pi \setminus (G \cup \Gamma G) \mid \pi$ logo, se $\alpha \mid \pi$ então $\alpha \mid \pi \setminus (G \cup \Gamma G)$ e se $\alpha \nmid \pi$ então $\alpha \nmid \pi \setminus (G \cup \Gamma G)$. □

Lema 16. *Seja $P = (\pi, \sigma, \Gamma)$ um problema e G um conjunto de genes em P. Se $\alpha = (x\ \Gamma\sigma x)$ e $\beta = (y\ \Gamma\sigma y)$ são arcos de P e $Supp(\alpha) \cup Supp(\beta) \cup \{\sigma x, \sigma y\} \subset E(P \setminus G)$ então α e β são arcos em $P \setminus G$, α e β se intercalam em $P \setminus G$ caso se intercalem em P e α e β não se intercalam em $P \setminus G$ caso não se intercalem em P.*

Demonstração. O Lema 14 nos garante que α e β são arcos em $P \setminus G$.

De acordo com a definição de subtração, $\pi \setminus (G \cup \Gamma G) \mid \pi$. Desta forma os elementos x, $\Gamma\sigma x$, y e $\Gamma\sigma y$ estarão, num dos anéis de $P \setminus G$, na mesma ordem que estavam num dos anéis de P e, assim, α e β se intercalam em $P \setminus G$ se, e somente se, se intercalam em P. □

Nos Lemas 17 e 18 assim como na Conjectura 1 e Corolário 2 faremos uma análise da relação entre os conceitos de entrelaçamento e intercalação. Ao final concluímos que os conceitos são eqüivalentes.

Lema 17. *Seja $P = (\pi, \sigma, \Gamma)$ um problema e α e β arcos distintos deste problema. Se os arcos α e β se intercalam então β entrelaça α.*

Demonstração. Inicialmente observemos a estrutura do anel. Note que, no anel, cada elemento fica acompanhado de seu complemento, os elementos de uma fita e outra alternam, os elementos da fita tomada como base, digamos π_1, ficam na mesma ordem de π ou, algebricamente, também podemos dizer que $\pi_1 | \Gamma\pi_1$. Os da outra fita, digamos π_2, ficam na ordem inversa ou, algebricamente, também podemos dizer que $\pi_2^{-1} \mid \Gamma\pi_1$.

Faremos agora uma análise de casos variando α e β de orientação. São quatro casos no total. Para o estudo iremos utilizar o seguinte problema como referência

$$P = \begin{cases} \pi &= (\Gamma a\ \Gamma x\ \Gamma y\ \Gamma f\ \ \Gamma i)(i\ f\ \ y\ \ x\ \ a) \\ \sigma &= \textit{qualquer} \\ \Gamma &= (a\ \Gamma a)(x\ \Gamma x)(y\ \Gamma y)(f\ \Gamma f)(i\ \Gamma i) \end{cases}$$

onde temos

$$\theta(\pi) = (a\ \Gamma a\ \ x\ \Gamma x\ \ y\ \Gamma y\ f\ \ \Gamma f\ \ i\ \Gamma i)$$

como um dos anéis. Temos também que $x \in Supp(\alpha)$, $y \in Supp(\beta)$, $\Gamma\sigma x = f$ ou $\Gamma\sigma x = \Gamma f$ e $\Gamma\sigma y = i$ ou $\Gamma\sigma y = \Gamma i$. Note que os valores de $\Gamma\sigma x$ e $\Gamma\sigma y$ variam de acordo com a orientação. Temos ainda que $a = \pi x$.

Para verificar se β entrelaça α iremos observar se β altera de orientação quando aplicamos a reversão induzida por α. A seguir apresentamos a análise em casos.

- α orientado: neste caso temos $\alpha = (x\ f)$ e $\rho(\alpha, \pi) = \rho(x, f, \pi)$. Ao aplicar $\rho(\alpha, \pi)$ teremos

$$\pi\rho(\alpha, \pi) = (\Gamma a\ y\ x\ \Gamma f\ \Gamma i)(i\ f\ \Gamma x\ \Gamma y\ a)$$

  e

$$\theta(\pi\rho(\alpha, \pi)) = (a\ \Gamma a\ \Gamma y\ y\ \Gamma x\ x\ f\ \Gamma f\ i\ \Gamma i)$$

  - β orientado: neste caso temos $\beta = (y\ i)$. Observe que $(y\ i) \nmid \pi\rho(\alpha, \pi)$ e, assim, podemos afirmar que β entrelaça α.
  - β não orientado: neste caso temos $\beta = (y\ \Gamma i)$. Observe que $(y\ \Gamma i) \mid \pi\rho(\alpha, \pi)$ e, assim, podemos afirmar que β entrelaça α.

- α não orientado: neste caso temos $\alpha = (x\ \Gamma f)$ e $\rho(\alpha, \pi) = \rho(\Gamma a, \Gamma f, \pi)$. Ao aplicar $\rho(\alpha, \pi)$ teremos

$$\pi\rho(\alpha, \pi) = (\Gamma a\ f\ y\ x\ \Gamma i)(i\ \Gamma x\ \Gamma y\ \Gamma f\ a)$$

  e

$$\theta(\pi\rho(\alpha, \pi)) = (a\ \Gamma a\ \Gamma f\ f\ \Gamma y\ y\ \Gamma x\ x\ i\ \Gamma i)$$

  - β orientado: neste caso temos $\beta = (y\ i)$. Observe que $(y\ i) \nmid \pi\rho(\alpha, \pi)$ e, assim, podemos afirma que β entrelaça α.
  - β não orientado: neste caso temos $\beta = (y\ \Gamma i)$. Observe que $(y\ \Gamma i) \mid \pi\rho(\alpha, \pi)$ e, assim, podemos afirmar que β entrelaça α.

Ao concluir a análise destes casos chegamos a conclusão de que β entrelaça α em qualquer das possíveis combinações de orientação para α e β em P. Observe que a subtração do gene a e conseqüente substituição de $\pi x = a$ para $\pi x = i$ não alteraria os resultados.

Observe também o que os Lemas 15 e 16 nos permitem. Digamos que tivéssemos um problema com o seguinte π

$$\pi = (\Gamma x\ \Gamma A_1\ \Gamma y\ \Gamma A_2\ \Gamma f\ \Gamma A_3\ \Gamma i\ \Gamma A_4)(A_4\ i\ A_3\ f\ A_2\ y\ A_1\ x)$$

onde A_1, A_2, A_3 e A_4 são blocos contendo zero ou mais elementos. Utilizando a subtração podemos construir qualquer problema com pelo menos dois arcos e reduzí-lo ao problema base desta demonstração. Finalmente, concluímos que se α e β se intercalam então β entrelaça α. □

Conjectura 1. *Seja P um problema e α e β arcos de P. Se α entrelaça β então α e β se intercalam.*

Lema 18. *Seja P um problema e α e β arcos de P. Se α entrelaça β então β entrelaça α.*

Demonstração. De acordo com a Conjectura 1 podemos dizer que, como α entrelaça β, α e β intercalam. Agora pelo Lema 17 podemos concluir que β entrelaça α. □

O resultado demonstrado pelo Lema 18 nos leva agora a dizer que α e β se entrelaçam ao invés de dizer que α entrelaça β. Os Lemas 17 e 18 e a Conjectura 1 nos levam ao seguinte corolário.

Corolário 2. *Seja P um problema e α e β arcos distintos de P. Os arcos α e β se intercalam se, e somente se, se entrelaçam.*

Os Lemas 19 e 20 a seguir iniciam uma análise do efeito de uma reversão induzida por um arco no que se refere aos entrelaçamentos.

Lema 19. *Seja $P = (\pi, \sigma, \Gamma)$ um problema e α e β arcos em P. Se $\alpha \mid \pi$ e α e β se entrelaçam em π então α e β não se entrelaçam em $\pi\rho(\alpha, \pi)$.*

Demonstração. Como $\alpha \mid \pi$, pelo Lema 13 podemos afirmar que $\pi\rho(\alpha,\pi)\rho(\alpha,\pi\rho(\alpha,\pi)) = \pi\rho(\alpha,\pi)$.

Segue uma demonstração em dois casos.

- Quando $\beta \mid \pi$: como α e β se entrelaçam em π então podemos afirmar que $\beta \nmid \pi\rho(\alpha,\pi)$. Pela observação acima podemos dizer que $\beta \nmid \pi\rho(\alpha,\pi)\rho(\alpha,\pi\rho(\alpha,\pi))$ e assim α e β não se entrelaçam em $\pi\rho(\alpha,\pi)$.

- Quando $\beta \nmid \pi$: como α e β se entrelaçam em π então podemos afirmar que $\beta \mid \pi\rho(\alpha,\pi)$. Pela observação acima podemos dizer que $\beta \mid \pi\rho(\alpha,\pi)\rho(\alpha,\pi\rho(\alpha,\pi))$ e assim α e β não se entrelaçam em $\pi\rho(\alpha,\pi)$.

□

Lema 20. *Seja $P = (\pi, \sigma, \Gamma)$ um problema e α e β arcos em P. Se $\alpha \nmid \pi$ e α e β se entrelaçam em π então α e β se entrelaçam em $\pi\rho(\alpha, \pi)$.*

Demonstração. Como $\alpha \nmid \pi$, pelo Lema 12 podemos afirmar que $\pi\rho(\alpha,\pi)\rho(\alpha,\pi\rho(\alpha,\pi)) = \pi$. Segue uma demonstração em dois casos:

- $\beta \mid \pi$: como α e β se entrelaçam em π então podemos afirmar que $\beta \nmid \pi\rho(\alpha,\pi)$. Pela observação acima podemos dizer que $\beta \mid \pi\rho(\alpha,\pi)\rho(\alpha,\pi\rho(\alpha,\pi))$ e assim α e β se entrelaçam em $\pi\rho(\alpha,\pi)$.

- $\beta \nmid \pi$: como α e β se entrelaçam em π então podemos afirmar que $\beta \mid \pi\rho(\alpha,\pi)$. Pela observação acima podemos dizer que $\beta \nmid \pi\rho(\alpha,\pi)\rho(\alpha,\pi\rho(\alpha,\pi))$ e assim α e β se entrelaçam em $\pi\rho(\alpha,\pi)$.

□

6.3 Grafo de entrelaçamento de arcos $OV(P)$

Seja $P = (\pi, \sigma, \Gamma)$ um problema. Define-se o **grafo de entrelaçamento de arcos** de P, denotado por $OV(P)$, da seguinte forma:

- **Vértices**: são os arcos de P. Subdividimos os vértices em orientados e não orientados. Dizemos que um vértice é **orientado** (preto) caso o arco correspondente seja orientado e é **não orientado** (branco) caso contrário.

- **Arestas**: são criadas entre arcos entrelaçados.

Apresentamos exemplos de grafos de entrelaçamento de arcos na Figura 6.2. Tais grafos referem-se ao problema P_2.

Os **componentes conexos** do grafo de entrelaçamento são classificados como:

1. **Trivial**: contém apenas um vértice, necessariamente não orientado.

2. **Orientado**: contém mais que um vértice e, pelo menos, um de seus vértices é orientado.

3. **Não orientado**: contém mais que um vértice, mas nenhum de seus vértices é orientado.

O grafo da Figura 6.2 (d) é composto de quatro componentes conexos triviais. Já na Figura 6.2 (a) temos um único componente conexo e este é orientado. Um exemplo de componente conexo não orientado é apresentado na Figura 6.3.

Decorre diretamente da definição de entrelaçamento o seguinte corolário.

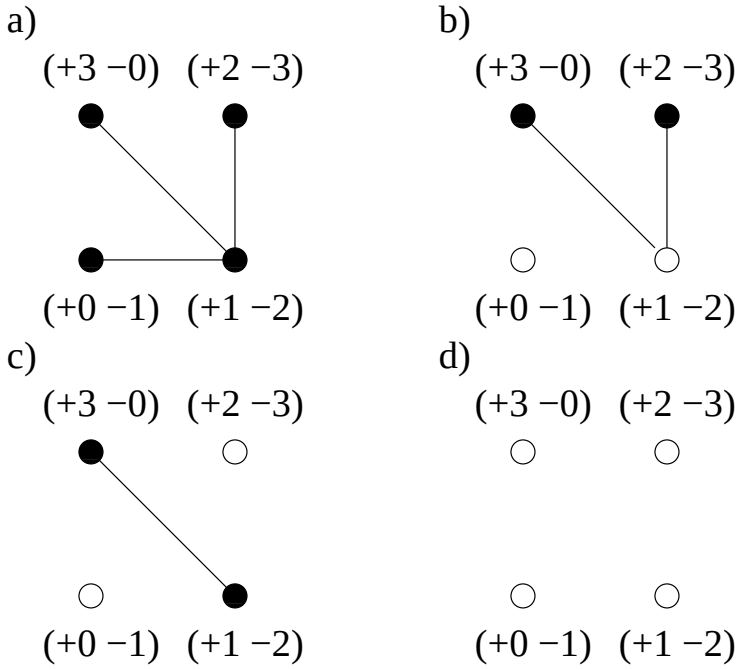


Figura 6.2: Exemplos de grafos de entrelaçamento de arcos. a) Apresenta o grafo de entrelaçamento de arcos do problema P_2. b) Apresenta o grafo de entrelaçamento de arcos do problema $P_2\rho_1$, onde $\rho_1 = \rho(+0, -1, \pi_2)$. c) Apresenta o grafo de entrelaçamento de arcos do problema $P_2\rho_1\rho_2$, onde $\rho_2 = \rho(+2, -3, \pi_2\rho_1)$. d) Apresenta o grafo de entrelaçamento de arcos de $P_2\rho_1\rho_2\rho_3$, onde $\rho_3 = \rho(+3, -0, \pi_2\rho_1\rho_2)$.

Corolário 3. *Seja $P = (\pi, \sigma, \Gamma)$ um problema, α e β vértices em $OV(P)$ e $\rho(\alpha, \pi)$ a reversão induzida por α em π. O arco β terá sua orientação invertida em $\pi\rho(\alpha, \pi)$ se, e somente se, β é adjacente a α em π.*

Seja $P = (\pi, \sigma, \Gamma)$ um problema. De acordo com o Colorolário 2 podemos ver o grafo de entrelaçamento de arcos de P como um grafo de cordas sobrepostas com um modelo semelhante ao diagrama de pontos de quebra de P, que podemos construir da seguinte forma:

- Cria-se um ciclo a partir de um dos anéis de π. Pode-se tomar qualquer um dos anéis de π, pois em qualquer deles as propriedades de interesse serão mantidas.
- Para cada vértice $\alpha = (x\ y)$ em OV(P) cria-se uma corda que toca o ciclo em x e y, ou, em outras palavras, descartaremos as arestas pretas e passaremos a olhar apenas para os arcos.

Na Figura 6.4 apresentamos o modelo associado ao grafo de cordas sobrepostas apresentado na Figura 6.2 (a).

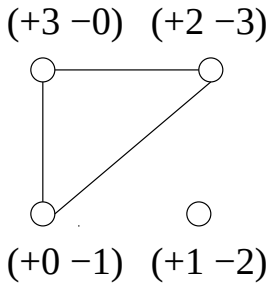


Figura 6.3: Grafo de entrelaçamento de arcos do problema $P_2\rho_4$, onde $\rho_4 = \rho(+1, -2, \pi_2)$.

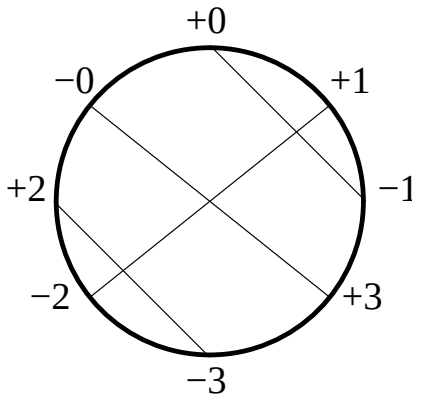


Figura 6.4: Modelo associado ao grafo de cordas sobrepostas apresentado na Figura 6.2 (a).

O Lema 21 assim como os Corolários 4, 5 e 6 prosseguem com nossa análise do efeito das reversões quanto a entrelaçamento. Agora também observamos as alterações nas orientações e temos também o grafo de entrelaçamento de arcos, que permite uma visualização mais simplificada dos efeitos. O Lema 21 nos trás um resultado novo bastante interessante, que permite a observação do que ocorre no grafo de entrelaçamento dada uma reversão qualquer. Nos trabalhos até então existentes quando se fazia análise de efeitos no grafo era sempre de uma reversão induzida por uma arco.

Lema 21. *Seja $P = (\pi, \sigma, \Gamma)$ um problema, ρ uma reversão em π e α e β arcos em P. Se α e β entrelaçam ρ em P então α e β se entrelaçam em $P\rho$ caso não se entrelacem em P ou não se entrelaçam em $P\rho$ caso se entrelacem em P.*

Demonstração. Seja $\theta(\pi)$ um dos anéis de π. Seja $\rho = (a, b, \pi)$. Para demonstrar este lema iremos utilizar a definição de grafo de cordas sobrepostas e uma análise em dois casos, que sumarizamos na Figura 6.5. No primeiro dos casos temos que os arcos α e β não entrelaçam em P, ou seja, as cordas (arcos) não se cruzam. Ao aplicarmos ρ, temos como efeito em $\theta(\pi)$, a inversão do trecho delimitado por a e b e assim passamos a uma

situação onde α e β agora se cruzam. No outro caso temos que os arcos α e β entrelaçam em P e então ocorre o oposto.

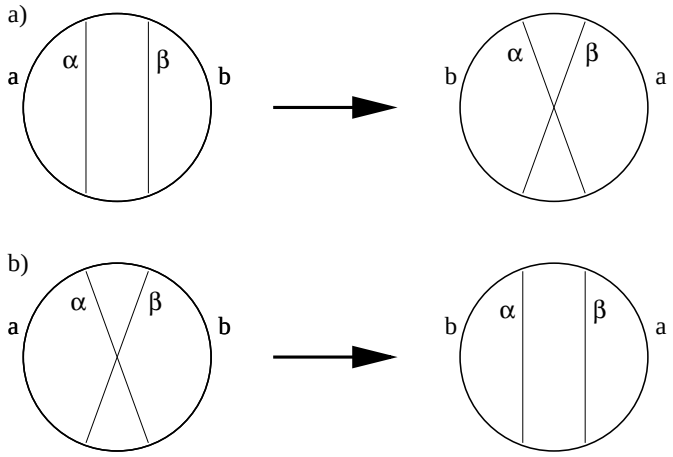


Figura 6.5: Efeito da reversão $\rho = (a, b, \pi)$ no modelo associado ao grafo de cordas sobrepostas.

□

Os Lemas 19, 20 e 21 e o Corolário 3 levam aos seguintes corolários.

Corolário 4. *Seja $P = (\pi, \sigma, \Gamma)$ um problema e α um vértice orientado em $OV(P)$. O efeito da aplicação da reversão induzida por α será o complemento do subgrafo de α e os vértices adjacentes a α assim como a inversão da orientação destes vértices.*

Corolário 5. *Seja $P = (\pi, \sigma, \Gamma)$ um problema e α um vértice não orientado em $OV(P)$. O efeito da aplicação da reversão induzida por α será o complemento do subgrafo dos vértices adjacentes a α assim como a inversão da orientação destes vértices.*

Corolário 6. *Seja $P = (\pi, \sigma, \Gamma)$ um problema e ρ uma reversão em π. O efeito de ρ no grafo de entrelaçamento de arcos de P será o complemento do subgrafo induzido pelos arcos que entrelaçam ρ assim como a inversão da orientação destes vértices.*

6.4 Componentes

Seja $P = (\pi, \sigma, \Gamma)$ um problema. Uma permutação ϕ é uma **transformação** quando ϕ^{-1} divide π e σ ao mesmo tempo, os produtos $\pi\phi$ e $\sigma\phi$ são genomas compatíveis sob $\Gamma\phi$ e $P\phi = (\pi\phi, \sigma\phi, \Gamma\phi)$ é um problema reversível. Ao **aplicar** uma transformação ϕ a um problema P obtém-se o problema $P\phi = (\pi\phi, \sigma\phi, \Gamma\phi)$.

Os Lemas 22 e 23 assim como o Corolário 7 nos trazem propriedades das transformações e decomposições.

Lema 22. *Seja $P = (\pi, \sigma, \Gamma)$ um problema. Se ϕ é uma transformação em P, então $orb(\pi\phi) = orb(\pi) + \|\phi\|$ e $orb(\sigma\phi) = orb(\sigma) + \|\phi\|$.*

Demonstração. Seja $E(\pi) = E(\pi\phi) = E(\phi\pi) = E$. Se ϕ é uma transformação em P então $\phi^{-1} \mid \pi$. Segue que:

$$\begin{aligned} \|\pi\phi\| &= \|\pi\| - \|\phi^{-1}\| \\ |E| - orb(\pi\phi) &= |E| - orb(\pi) - \|\phi^{-1}\| \\ -orb(\pi\phi) &= -orb(\pi) - \|\phi^{-1}\| \\ orb(\pi\phi) &= orb(\pi) + \|\phi^{-1}\| \\ orb(\pi\phi) &= orb(\pi) + \|\phi\| \end{aligned}$$

O mesmo ocorre com σ. □

Lema 23. *Seja $P = (\pi, \sigma, \Gamma)$ um problema. Se ϕ é uma transformação em P então $(\pi\phi) \mid \pi$ e $(\sigma\phi) \mid \sigma$.*

Demonstração. Se ϕ é uma transformação em P então $\phi^{-1} \mid \pi$. Segue que:

$$\begin{aligned} \|\pi\phi\| &= \|\pi\| - \|\phi^{-1}\| \\ \|\phi^{-1}\| &= \|\pi\| - \|\pi\phi\| \\ \|\pi \bullet \phi^{-1}\| &= \|\pi\| - \|\pi\phi\| \\ \|\pi(\pi\phi)^{-1}\| &= \|\pi\| - \|\pi\phi\| \end{aligned}$$

A última igualdade permite-nos afirmar que $(\pi\phi) \mid \pi$. O mesmo ocorre com σ. □

Seja $P = (\pi, \sigma, \Gamma)$ um problema monocromossomal e ϕ uma transformação em P. Define-se por **decomposição** o procedimento de aplicar ϕ a P seguido pelo particionamento do problema resultante em problemas monocromossomais. Desta forma também podemos falar em decomposição ϕ. Um problema monocromossomal é **indecomponível** caso não existam transformações no problema com norma maior que 0.

Decorre diretamente do Lema 23 o seguinte corolário.

Corolário 7. *Seja $P = (\pi, \sigma, \Gamma)$ um problema monocromossomal. Se o conjunto $C = \{P_1, P_2, \ldots, P_n\}$ é uma decomposição em componentes de P, onde $P_1 = (\pi_1, \sigma_1, \Gamma_1)$, $P_2 = (\pi_2, \sigma_2, \Gamma_2)$, $\ldots$, $P_n = (\pi_n, \sigma_n, \Gamma_n)$, então $\pi_i \mid \pi$, para $1 \leq i \leq n$.*

No problema $P_1 = (\pi_1, \sigma_1, \Gamma_1)$, anteriormente apresentado, há diversas decomposições, por exemplo: $\phi_1 = (+0\ +8)(-0\ -8)$, $\phi_2 = (+0\ +11)(-0\ -11)$, $\phi_3 = (+1\ +6)(-1\ -6)$, $\phi_4 = (-5\ -2)(+5\ +2)$, $\phi_5 = (+8\ +11)(-8\ -11)$ e $\phi_6 = (+11\ +8\ +0)(-0\ -8\ -11)$. Ao aplicarmos a decomposição ϕ_5 obtemos os seguintes sub-problemas:

$$P_3 = \begin{cases} \pi_3 = (+10\ +9\ +11)(-8\ -9\ -10) \\ \sigma_3 = (+9\ +10\ +11)(-8\ -10\ -9) \\ \Gamma_3 = (+9\ -9)(+10\ -10)(+11\ -8) \end{cases}$$

$$P_4 = \begin{cases} \pi_4 = (+0\ +7\ +1\ -5\ -3\ -4\ -2\ +6\ +8\ +14\ +13\ +12) \\ (-12\ -13\ -14\ -11\ -6\ +2\ +4\ +3\ +5\ -1\ -7\ -0) \\ \sigma_4 = (+0\ +1\ +2\ +3\ +4\ +5\ +6\ +7\ +8\ +12\ +13\ +14) \\ (-14\ -13\ -12\ -11\ -7\ -6\ -5\ -4\ -3\ -2\ -1\ -0) \\ \Gamma_4 = (+0\ -0)(+1\ -1)(+2\ -2)(+3\ -3)(+4\ -4)(+5\ -5)(+6\ -6)(+7\ -7) \\ (+8\ -11)(+12\ -12)(+13\ -13)(+14\ -14) \end{cases}$$

Observe que não há decomposições em P_3 com norma maior que zero. Porém, existem decomposições em P_4, tais como: $\phi_7 = (+0\ +8)(-11\ -0)$, $\phi_8 = (+1\ +6)(-1\ -6)$ e $\phi_9 = (-5\ -2)(+5\ +2)$. Ao aplicarmos a decomposição ϕ_7 obtemos os sub-problemas P_5 e P_6.

$$P_5 = \begin{cases} \pi_5 = (+0\ +14\ +13\ +12)(-12\ -13\ -14\ -11) \\ \sigma_5 = (+0\ +12\ +13\ +14)(-14\ -13\ -12\ -11) \\ \Gamma_5 = (+0\ -11)(+12\ -12)(+13\ -13)(+14\ -14) \end{cases}$$

$$P_6 = \begin{cases} \pi_6 = (+7\ +1\ -5\ -3\ -4\ -2\ +6\ +8)(-0\ -6\ +2\ +4\ +3\ +5\ -1\ -7) \\ \sigma_6 = (+7\ +8\ +1\ +2\ +3\ +4\ +5\ +6)(-6\ -5\ -4\ -3\ -2\ -1\ -0\ -7) \\ \Gamma_6 = (+1\ -1)(+2\ -2)(+3\ -3)(+4\ -4)(+5\ -5)(+6\ -6)(+7\ -7)(+8\ -0) \end{cases}$$

Note que P_5 é um sub-problema indecomponível e que há decomposições em P_6, tais como: $\phi_{10} = (+1\ +6)(-1\ -6)$ e $\phi_{11} = (-5\ -2)(+5\ +2)$. Ao aplicarmos a decomposição ϕ_{10} obtemos os sub-problemas P_7 e P_8.

$$P_7 = \begin{cases} \pi_7 = (+7\ +1\ +8)(-0\ -6\ -7) \\ \sigma_7 = (+7\ +8\ +1)(-6\ -0\ -7) \\ \Gamma_7 = (+1\ -6)(+7\ -7)(+8\ -0) \end{cases}$$

$$P_8 = \begin{cases} \pi_8 = (-5\ -3\ -4\ -2\ +6)(-1\ +2\ +4\ +3\ +5) \\ \sigma_8 = (+2\ +3\ +4\ +5\ +6)(-1\ -5\ -4\ -3\ -2) \\ \Gamma_8 = (+2\ -2)(+3\ -3)(+4\ -4)(+5\ -5)(+6\ -1) \end{cases}$$

Observe que P_7 é indecomponível e que em P_8 encontramos a decomposição $\phi_{12} = (-5\ -2)(+5\ +2)$, que quando aplicada gera os sub-problemas indecomponíveis P_9 e P_{10}.

$$P_9 = \begin{cases} \pi_9 = (-5\ +6)(-1\ +2) \\ \sigma_9 = (+2\ +6)(-1\ -5) \\ \Gamma_9 = (+2\ -5)(+6\ -1) \end{cases}$$

$$P_{10} = \begin{cases} \pi_{10} = (-3\ -4\ -2)(+5\ +4\ +3) \\ \sigma_{10} = (+3\ +4\ +5)(-2\ -4\ -3) \\ \Gamma_{10} = (+3\ -3)(+4\ -4)(+5\ -2) \end{cases}$$

Note que em P_1 há outras decomposições e dentre elas podemos citar

$$\phi_P = \phi_5\phi_7\phi_{10}\phi_{12} = (+0\ +11\ +8)(-8\ -11\ -0)(+1\ +6)(-6\ -1)(-5\ -2)(+5\ +2)$$

que quando aplicada resulta nos problemas P_3, P_5, P_7, P_9 e P_{10}.

Define-se **componente** como um sub-problema monocromossomal indecomponível. Seja $P = (\pi, \sigma, \Gamma)$ um problema. O problema P é **trivial** caso $\pi^{-1}\sigma = 1$. Denota-se por $k(P)$ o **número de componentes** no problema P. No exemplo P_1 há cinco componentes, são eles: P_3, P_5, P_7, P_9 e P_{10}, ou seja, $k(P) = 5$. Nenhum dos componentes é trivial.

Seja P um problema monocromossomal. Pode-se definir um diagrama, que chamaremos de **diagrama de decomposição em componentes**, como uma árvore com raiz. Cada nó representa um problema. A raiz da árvore é P. Os nós filhos de um dado nó N são sub-problemas obtidos pela aplicação de uma decomposição a N. As folhas da árvore são componentes e o conjunto das folhas define uma **decomposição em componentes** de P. Observe que um problema pode ter um ou mais diagramas de decomposição em componentes. A Figura 6.6 apresenta um diagrama de decomposição em componentes para o problema P_1. Note que se aplicarmos ϕ_P a P_1 teremos um outro diagrama.

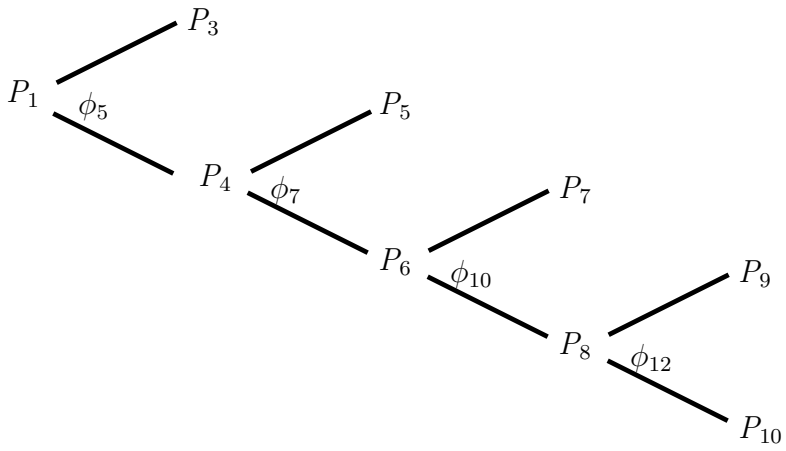


Figura 6.6: Ilustra um possível diagrama de decomposição em componentes de P_1 .

Uma transformação de maior norma em um problema recebe o nome de **transformação principal**. O particionamento do problema obtido pela aplicação de uma transformação principal nos fornece uma decomposição em componentes.

Um componente é **bom**, caso possua, pelo menos, uma quebra e é **ruim**, caso contrário. No exemplo há apenas um componente bom P_9. Dizemos que um problema é **quase bom** caso, ao ser decomposto em componentes, exista entre todos os componentes apenas um componente ruim e não trivial. Seja P um problema. Denota-se respectivamente por $k^+(P)$ e $k^-(P)$ o **número de componentes bons** e o **número de componentes ruins** em P. Note que $k(P) = k^+(P) + k^-(P)$.

Lema 24. *Seja P um componente não trivial. Se P é ruim então há, pelo menos, uma recombinação.*

Demonstração. Seja $P = (\pi, \sigma, \Gamma)$. Sejam A e B as órbitas de π. Como P é ruim então não há quebras, ou em outras palavras, não há dois elementos em A, digamos a_1 e a_2, tal que a_1 pertença a órbita do ciclo α de $\pi^{-1}\sigma$ e a_2 pertença a órbita de $(\Gamma\pi) \bullet \alpha^{-1}$, companheiro de α. Sendo P não trivial segue que existe pelo menos um par de ciclos γ tal que $l(\gamma) \geq 2$ e $\gamma_1 \subset A$, para γ_1 órbita de γ. Seja $\{g_1, g_2\} \subset \gamma_1$. A reversão $\rho(g_1, g_2, \pi)$ define uma recombinação. □

Os Lemas 25, 26 e 27 assim como os Corolários 8 e 9 fazem uma análise da correlação entre objetos de um problema e subproblema.

Lema 25. *Seja $P = (\pi, \sigma, \Gamma)$ um problema monocromossomal e ϕ uma decomposição em P. Se γ é um par de ciclos em $P\phi$ então $\phi \bullet \gamma$ é o par de ciclo correspondente em P.*

Demonstração. Observe que $\alpha \bullet \beta \bullet \omega = (\alpha\beta) \bullet \omega$ para quaisquer que sejam as permutações α, β e ω.

Se $P\phi$ resulta nos sub-problemas P_1, P_2, ..., P_n definidos por $(\pi_1, \sigma_1, \Gamma_1)$, $(\pi_2, \sigma_2, \Gamma_2)$, ..., $(\pi_n, \sigma_n, \Gamma_n)$, respectivamente, então temos $\pi\phi = \pi_1\pi_2 \ldots \pi_n$, $\sigma\phi = \sigma_1\sigma_2 \ldots \sigma_n$, $\Gamma\phi = \Gamma_1\Gamma_2 \ldots \Gamma_n$ e, finalmente, $(\pi\phi)^{-1}(\sigma\phi) = (\pi_1\pi_2 \ldots \pi_n)^{-1}(\sigma_1\sigma_2 \ldots \sigma_n)$. Segue que:

$$
\begin{gathered}
(\pi\phi)^{-1}(\sigma\phi) = (\pi_1\pi_2 \ldots \pi_n)^{-1}(\sigma_1\sigma_2 \ldots \sigma_n) \\
\phi^{-1}\pi^{-1}\sigma\phi = \pi_n^{-1} \ldots \pi_2^{-1}\pi_1^{-1}\sigma_1\sigma_2 \ldots \sigma_n \\
\phi^{-1} \bullet (\pi^{-1}\sigma) = (\pi_n^{-1}\sigma_n) \ldots (\pi_2^{-1}\sigma_2)(\pi_1^{-1}\sigma_1) \\
\pi^{-1}\sigma = \phi \bullet [(\pi_n^{-1}\sigma_n) \ldots (\pi_2^{-1}\sigma_2)(\pi_1^{-1}\sigma_1)] \\
\pi^{-1}\sigma = \phi \bullet (\pi_n^{-1}\sigma_n) \ldots \phi \bullet (\pi_2^{-1}\sigma_2)\phi \bullet (\pi_1^{-1}\sigma_1)
\end{gathered}
$$

O quociente $\pi_i^{-1}\sigma_i$, para $1 \leq i \leq n$, do sub-problema $P_i = (\pi_i, \sigma_i, \Gamma_i)$, é composto por um produto de pares de ciclos companheiros. Tais pares de ciclos possuem conjuntos-bases disjuntos e assim se $\gamma_1\gamma_2 \ldots \gamma_m = \pi_i^{-1}\sigma_i$, para γ_j, com $1 \leq j \leq m$, par de ciclo em P_i, então $\phi \bullet (\pi_i^{-1}\sigma_i) = (\phi \bullet \gamma_1)(\phi \bullet \gamma_2) \ldots (\phi \bullet \gamma_m)$. □

Lema 26. *Sejam os problemas P, P_1, P_2, ..., P_n, tal que P_1, P_2, ..., P_n são sub-problemas obtidos pela aplicação da decomposição ϕ a P. A reversão ρ em P_i, para $1 \leq i \leq n$, é a mesma que $\phi \bullet \rho$ em P.*

Demonstração. Sejam P, P_1, P_2, ..., P_n definidos por (π, σ, Γ), $(\pi_1, \sigma_1, \Gamma_1)$, $(\pi_2, \sigma_2, \Gamma_2)$, ..., $(\pi_n, \sigma_n, \Gamma_n)$, respectivamente. Sabemos que $\pi_1\pi_2 \dots \pi_n = \pi\phi$, $\sigma_1\sigma_2 \dots \sigma_n = \sigma\phi$ e $\Gamma_1\Gamma_2 \dots \Gamma_n = \Gamma\phi$. Segue:

$$\begin{gathered}
\rho_1 \dots \rho_{k-1}\rho_k = \pi_1^{-1}\sigma_1\pi_2^{-1}\sigma_2 \dots \pi_n^{-1}\sigma_n \\
\rho_1 \dots \rho_{k-1}\rho_k = \pi_1^{-1}\pi_2^{-1} \dots \pi_n^{-1}\sigma_1\sigma_2 \dots \sigma_n \\
\rho_1 \dots \rho_{k-1}\rho_k = (\pi_n \dots \pi_2\pi_1)^{-1}(\sigma_1\sigma_2 \dots \sigma_n) \\
\rho_1 \dots \rho_{k-1}\rho_k = (\pi\phi)^{-1}(\sigma\phi) \\
\rho_1 \dots \rho_{k-1}\rho_k = \phi^{-1} \bullet (\pi^{-1}\sigma) \\
\phi\rho_1 \dots \rho_{k-1}\rho_k\phi^{-1} = \pi^{-1}\sigma \\
\phi\rho_1\phi^{-1}\phi \dots \phi^{-1}\phi\rho_{k-1}\phi^{-1}\phi\rho_k\phi^{-1} = \pi^{-1}\sigma \\
(\phi \bullet \rho_1) \dots (\phi \bullet \rho_{k-1})(\phi \bullet \rho_k) = \pi^{-1}\sigma
\end{gathered}$$

□

Lema 27. *Seja $P = (\pi, \sigma, \Gamma)$ um problema. Se ϕ é uma decomposição em P então $\Gamma\sigma = \phi \bullet (\Gamma\phi\sigma\phi)$.*

Demonstração. Seja $P\phi = (\pi\phi, \sigma\phi, \Gamma\phi)$ decomposto nos sub-problemas $P_1 = (\pi_1, \sigma_1, \Gamma_1)$, $P_2 = (\pi_2, \sigma_2, \Gamma_2)$, ..., $P_n = (\pi_n, \sigma_n, \Gamma_n)$. Temos então que $\pi\phi = \pi_1\pi_2 \dots \pi_n$, $\sigma\phi = \sigma_1\sigma_2 \dots \sigma_n$ e $\Gamma\phi = \Gamma_1\Gamma_2 \dots \Gamma_n$. Note que ϕ é um genoma sob Γ. Sendo assim podemos dizer que $\Gamma\phi\Gamma = \phi^{-1}$ e que $\Gamma\phi = \phi^{-1}\Gamma$. Segue então:

$$\begin{gathered}
(\Gamma\phi)(\sigma\phi) = \phi^{-1}\Gamma\sigma\phi \\
\Gamma\phi\sigma\phi = \phi^{-1} \bullet (\Gamma\sigma) \\
\phi \bullet (\Gamma\phi\sigma\phi) = \Gamma\sigma
\end{gathered}$$

□

O Lema 27 nos leva aos Corolários 8 e 9, que apresentamos a seguir.

Corolário 8. *Seja $P = (\pi, \sigma, \Gamma)$ um problema. Se ϕ é uma decomposição em P então P e $P\phi$ possuem o mesmo número de arcos.*

Corolário 9. *Seja $P = (\pi, \sigma, \Gamma)$ um problema e ϕ uma decomposição em P. Se α é um arco em $P\phi$ então $\phi \bullet \alpha$ é um arco em P.*

6.4.1 Componentes e componentes conexos de $OV(P)$

Seja $P = (\pi, \sigma, \Gamma)$ um problema e ϕ uma decomposição que divide P nos sub-problemas $P_1, P_2, \ldots, P_n$. Podemos obter P se tivermos os sub-problemas e ϕ. Observe que $\pi\phi = \pi_1\pi_2 \ldots \pi_n$, $\sigma\phi = \sigma_1\sigma_2 \ldots \sigma_n$ e $\Gamma\phi = \Gamma_1\Gamma_2 \ldots \Gamma_n$. Assim, fazendo o produto com ϕ^{-1} pela direita em cada um dos produtos obteremos a tripla que define P. Denotamos este procedimento, oposto a decomposição por $P_1P_2 \ldots P_n\phi^{-1}$.

Lema 28. *Seja $P = (\pi, \sigma, \Gamma)$ um problema tal que $P_1 = (\pi_1, \sigma_1, \Gamma_1)$, $P_2 = (\pi_2, \sigma_2, \Gamma_2)$, ..., $P_n = (\pi_n, \sigma_n, \Gamma_n)$ são os problemas resultantes da aplicação da decomposição ϕ a P. Se α é um arco de P_1 então*

$$P\rho(\phi \bullet \alpha, \pi) = P_n \ldots P_2(P_1\rho(\alpha, \pi_1))\phi^{-1}$$

Demonstração. Observe que $P\phi = (\pi\phi, \sigma\phi, \Gamma\phi)$, onde $\pi\phi = \pi_1\pi_2 \ldots \pi_n$, $\sigma\phi = \sigma_1\sigma_2 \ldots \sigma_n$ e $\Gamma\phi = \Gamma_1\Gamma_2 \ldots \Gamma_n$. De acordo com o Lema 26 e Corolário 9 podemos dizer que $\rho(\alpha, \pi_1) = \phi^{-1} \bullet \rho(\phi \bullet \alpha, \pi)$. Segue que:

$$\begin{aligned} \pi_n \ldots \pi_2(\pi_1\rho(\alpha, \pi_1))\phi^{-1} &= \pi\phi\rho(\alpha, \pi_1)\phi^{-1} \\ &= \pi\phi\phi^{-1}\rho(\phi \bullet \alpha, \pi)\phi\phi^{-1} \\ &= \pi\rho(\phi \bullet \alpha, \pi) \end{aligned}$$

Como a reversão não afeta σ e Γ, podemos concluir que

$$P\rho(\phi \bullet \alpha, \pi) = P_n \ldots P_2(P_1\rho(\alpha, \pi_1))\phi^{-1}$$

□

Seja $P = (\pi, \sigma, \Gamma)$ um problema e $\alpha = (x\ y)$ um arco de P. Dizemos que x e y são os **extremos** do arco α. Seja C um componente conexo de $OV(P)$. Denotamos por $Ext(C)$ o **conjunto dos extremos** dos arcos de C ou, simplesmente, os extremos de C.

Dizemos que u e v são **consecutivos** na permutação γ se $v = \gamma(u)$ ou $u = \gamma(v)$. Uma seqüência de elementos consecutivos num dos anéis de P recebe o nome de **caminho**. Dizemos que um caminho S **passa** por C se $S \cap Ext(C) \neq \emptyset$.

Seja P um problema e C e D componentes conexos distintos de $OV(P)$. Dizemos que x é **extremo** de C em relação D quando x é extremo de C e existe um caminho que inicia em x, termina em D e não passa por C (desconsiderando x). Denotamos o conjunto dos extremos de C em relação a D por $Ext(C \to D)$.

Seja P um problema e C e D componentes conexos distintos de $OV(P)$. Dizemos que x é **elemento interno** de C em relação a D quando $x \in Ext(C)$ e $x \notin Ext(C \to D)$. Denotamos o conjunto dos elementos internos de C em relação a D por $Int(C \to D)$.

Lema 29. *Seja P um problema. Se C e D são componentes conexos distintos de $OV(P)$ então $|Ext(C \rightarrow D)| = 2$.*

Demonstração. Inicialmente iremos mostrar que $|Ext(C \rightarrow D)| \leq 2$ por absurdo. Seja $Ext(C \rightarrow D) = \{x_1, x_2, x_3\}$. Podemos afirmar que x_1, x_2 e x_3 dividem o anel em três regiões, são elas: x_1x_2, x_2x_3 e x_3x_1. Observe que não pode haver elementos de D em mais que uma destas regiões, senão C e D se alternariam no anel e haveria arco entrelaçando. Desta forma D está limitado a apenas uma das regiões. Sem perda de generalidade, digamos que D está na região delimitada por x_1 e x_2. Chegamos a conclusão de que qualquer caminho partindo de x_3 e terminando em D terá de passar por x_1 ou x_2. O que é um absurdo e o que nos mostra que $|Ext(C \rightarrow D)| \leq 2$.

Para provar que $|Ext(C \rightarrow D)| = 2$ precisamos ainda mostrar que $|Ext(C \rightarrow D)| \geq 2$.

Sabemos que $|Ext(C)| \geq 2$. Iremos mostrar que $|Ext(C \rightarrow D)| \geq 2$ a partir de uma demonstração construtiva. Escolha $d \in Ext(D)$ e uma direção a seguir no anel. Partindo de d percorra o anel na direção escolhida até encontrar o primeiro elemento de C, digamos c_1. De acordo com a definição c_1 é extremo de C em relação a D. De forma análoga, parta de d na direção oposta e pare ao encontrar o primeiro elemento de C, digamos c_2. Da mesma forma c_2 é extremo de C em relação a D e, como $|Ext(C)| \geq 2$, então $c_1 \neq c_2$. Assim concluímos que $|Ext(C \rightarrow D)| \geq 2$ e, finalmente $|Ext(C \rightarrow D)| = 2$. □

Seja P um problema e C e D componentes conexos de $OV(P)$. Se ϕ é uma decomposição de P tal que C e D estão em sub-problemas distintos de $P\phi$ então dizemos que ϕ **separa** C e D.

Conjectura 2. *Seja $P = (\pi, \sigma, \Gamma)$ um problema. Se C e D são componentes conexas distintas de $OV(P)$ então $\phi = (x\ \Gamma y)(y\ \Gamma x)$, para $\{x, y\} = Ext(C \rightarrow D)$, é uma decomposição de P.*

Conjectura 3. *Seja P um problema e C e D componentes conexas distintas de $OV(P)$. Se $\{x, y\} = Ext(C \rightarrow D)$ então $\phi = (x\ \Gamma y)(y\ \Gamma x)$ é uma decomposição e separa C e D.*

Seja $P = (\pi, \sigma, \Gamma)$ um problema tal que $\mathcal{A}$ representa uma decomposição em componentes de P e ϕ a decomposição principal de P que deu origem a $\mathcal{A}$. Denotamos por $Arcos(\mathcal{A})$ a função que, dado um conjunto problema, retorna, para cada problema, um conjunto de arcos.

Lema 30. *Seja $P = (\pi, \sigma, \Gamma)$ um problema tal que $\mathcal{A}$ representa uma decomposição em componentes de P e ϕ a decomposição principal de P que deu origem a $\mathcal{A}$. Se $\mathcal{C}$ é o conjunto dos componentes conexos de $OV(P)$ então $Arcos(\mathcal{A}) = \phi^{-1} \bullet \mathcal{C}$.*

Demonstração. Iremos demonstrar esta correspondência em dois passos. Primeiro iremos mostrar que $\phi^{-1} \bullet \mathcal{C} \subseteq Arcos(\mathcal{A})$ e depois mostraremos que $Arcos(\mathcal{A}) \subseteq \phi^{-1} \bullet \mathcal{C}$.

A primeira parte, $\phi^{-1} \bullet \mathcal{C} \subseteq Arcos(\mathcal{A})$, provamos por contradição. Seja C um componente conexo qualquer em $\mathcal{C}$ e c_1 e c_2 arcos distintos de C tal que $\phi^{-1} \bullet c_1$ é arco de A_1 e que $\phi^{-1} \bullet c_2$ é arco de A_2, onde A_1 e A_2 são componentes distintos de $\mathcal{A}$. Como c_1 e c_2 estão em um mesmo componente conexo de $OV(P)$ podemos afirmar que existe, pelo menos, um caminho que liga c_1 e c_2. Seja o caminho $c_1, l_1, l_2, \ldots, l_{n-1}, l_n, c_2$. Podemos dizer que c_1 e l_1, l_1 e l_2, l_{n-1} e l_n, l_n e c_2 se entrelaçam em P. Como $\phi^{-1} \bullet c_1$ é arco de A_1, de acordo com o Lema 28 e pela definição de entrelaçamento entre arcos, podemos afirmar que $\phi^{-1} \bullet l_1$ também é arco de A_1 e $\phi^{-1} \bullet c_1$ e $\phi^{-1} \bullet l_1$ se entrelaçam em A_1. Da mesma forma, terminamos por chegar a conclusão de que $\phi^{-1} \bullet c_1, \phi^{-1} \bullet l_1, \phi^{-1} \bullet l_2, \ldots, \phi^{-1} \bullet l_{n-1}, \phi^{-1} \bullet l_n, \phi^{-1} \bullet c_2$ são todos arcos de A_1, contrariando o que assumimos no início.

Agora mostraremos que $Arcos(\mathcal{A}) \subseteq \phi^{-1} \bullet \mathcal{C}$. Considere o menor subproblema de P que contém totalmente $C \in \mathcal{C}$. Chamemos este de A. Se, por absurdo, ele contiver algo a mais que C, digamos D, de acordo com a Conjectura 3 podemos decompô-lo de forma a separar D e C, contrariando a escolha de A. Logo, $Arcos(\mathcal{A}) = \phi^{-1} \bullet \mathcal{C}$. □

6.4.2 Obstáculos

Seja P um problema. Um **obstáculo** em P é um componente ruim e não trivial contido em um sub-problema quase bom que pode ser obtido com, no máximo, uma decomposição ϕ de norma igual a dois aplicada ao problema inicial P. Denota-se por $h(P)$, $\overline{h}(P)$ e $t(P)$, respectivamente, o número de obstáculos, número de componentes ruins que não são obstáculos nem triviais e componentes triviais no problema P. Note que $k^-(P) = h(P) + \overline{h}(P) + t(P)$.

Em nosso exemplo temos que P_3, P_5 e P_{10} são componentes ruins e obstáculos.

Lema 31. *Seja $P = (\pi, \sigma, \Gamma)$ um problema. Se $\alpha = (x\ y)$ é um arco em P então x e y pertencem a ciclos distintos porém companheiros de $\pi^{-1}\sigma$.*

Demonstração. Como α é arco então $\alpha \mid \Gamma\sigma$ ou ainda $(x\ \Gamma\sigma x) \mid \Gamma\sigma$. Substituindo x por $\Gamma\pi z$ obtemos $(\Gamma\pi z\ \sigma^{-1}\pi z) \mid \Gamma\sigma$.

Sabemos que cada ciclo γ em $\pi^{-1}\sigma$ possui um ciclo companheiro $(\Gamma\pi) \bullet \gamma^{-1}$. A partir daí podemos concluir que se z e $\sigma^{-1}\pi z$ pertencem ao ciclo β de $\pi^{-1}\sigma$ então $\Gamma\pi z$ pertence ao ciclo companheiro de β. □

Lema 32. *Seja $P = (\pi, \sigma, \Gamma)$ um problema. Se α é um arco em P e $\alpha \nmid \pi$ então $\rho(\alpha, \pi)$ é uma recombinação.*

Demonstração. Seja $x \in Supp(\alpha)$. Como $\alpha \nmid \pi$ então $\rho(\alpha, \pi) = \rho(\Gamma\pi x, \Gamma\sigma x, \pi)$. Pelo Lema 31 podemos afirmar que x e $\Gamma\sigma x$ pertencem a ciclos companheiros e que $(\Gamma\pi x\ \Gamma\sigma x) \mid \pi^{-1}\sigma$.

Da última observação chegamos a conclusão de que arcos não orientados originam recombinações. □

Lema 33 (Corte de Obstáculos)**.** *Seja P um obstáculo. Se ρ é a reversão induzida por um dos arcos de P então $h(P\rho) = 0$ e $k(P\rho) = 1$.*

Demonstração. Seja α um dos arcos em P e $\rho(\alpha, \pi)$ a reversão induzida por α em π. De acordo com o Lema 20 podemos afirmar que os vértices adjacentes a α em π permanecerão adjacentes a α em $\pi\rho$, ou seja, o componente se mantém unido depois da aplicação da reversão e, assim, $k(P\rho) = k(P) = 1$.

Basta agora mostrar que há, pelo menos, um arco orientado em $\pi\rho$. Tal fato é garantido por P ser um componente não trivial e pelo Corolário 3. □

Um obstáculo é **simples** caso, ao ser eliminado (cortado), o número de obstáculos em P é decrescido de uma unidade e é **super**, caso contrário. Os componentes P_3 e P_5 são obstáculos simples. Já o componente P_{10} é um super obstáculo. Denota-se respectivamente por $h_{si}(P)$ e $h_{su}(P)$ os números de obstáculos simples e o número de super obstáculos no problema P. Note que $h(P) = h_{si}(P) + h_{su}(P)$.

Seja P um problema monocromossomal. Caso P contenha um número ímpar de obstáculos e todos estes sejam super, diz-se que P é uma **fortaleza**. Define-se a função $f(P)$ que retorna 1 caso P (monocromossomal) seja uma fortaleza e 0, caso contrário. No caso de problemas multicromossomais a função f deve retornar a soma dos f's dos problemas monocromossomais.

Seja P um problema e H um componente ruim de P que não é obstáculo. Se H_{su} é um obstáculo em P que ao ser cortado por ρ faz com que H torne-se obstáculo em $P\rho$ dizemos que H_{su} **protege** H.

Seja $P = (\pi, \sigma, \Gamma)$ um problema. Sejam H_1 e H_2 obstáculos distintos em P. Dizemos que H_1 e H_2 são **vizinhos** se há um caminho partindo de H_1, chegando a H_2 e que não passa por outro obstáculo.

Seja $P = (\pi, \sigma, \Gamma)$ um problema tal que $k^+ = 0$. Sejam H_1 e H_2 obstáculos distintos e vizinhos. Se $a \in Int(H_1 \rightarrow H_2)$ e $b \in Int(H_2 \rightarrow H_1)$ tal que a consecutivo no anel a $x \in Ext(H_1 \rightarrow H_2)$ e b consecutivo no anel a $y \in Ext(H_2 \rightarrow H_1)$ e, além disso, $(a\ b) \mid \pi$ então dizemos que a reversão $\rho(a, b, \pi)$ é uma **junção de obstáculos**.

Lema 34. *Seja $P = (\pi, \sigma, \Gamma)$ um problema tal que $k^+ = 0$ e $h(P) = 2$. Sejam H_1 e H_2 os obstáculos em P e $a \in Int(H_1 \rightarrow H_2)$ e $b \in Int(H_2 \rightarrow H_1)$ tal que a consecutivo a $x \in Ext(H_1 \rightarrow H_2)$ e b consecutivo a $y \in Ext(H_2 \rightarrow H_1)$. Se $(a\ b) \mid \pi$ então $h(P\rho(a, b, \pi)) = 0$.*

Demonstração. Inicialmente iremos mostrar que ρ entrelaça, pelo menos, um arco, digamos α, de H_1 e com, pelo menos, um arco, digamos β, de H_2. Seja $|E(H)| = |Ext(H)|$. Pela definição de arcos temos que o número de arcos em cada obstáculo H é igual a metade da cardinalidade de seu conjunto base $E(H)$. Em outras palavras, cada elemento de um obstáculo (ou problema) é o extremo de um arco. Note que, pelo Lema 29, $|Ext(H_1 \rightarrow H_2)| = 2$ e $|Ext(H_2 \rightarrow H_1)| = 2$. Note também que $|Ext(H_1)| \geq 4$ e $|Ext(H_2)| \geq 4$, uma vez que H_1 e H_2 são obstáculos e, assim, não são triviais e cada arco contribui com dois extremos. Sejam $x \in Supp(\alpha)$ e $y \in Supp(\beta)$. Finalmente, observe que α e β intercalam ρ e, sendo assim, α e β entrelaçam $\rho = \rho(a, b, \pi)$.

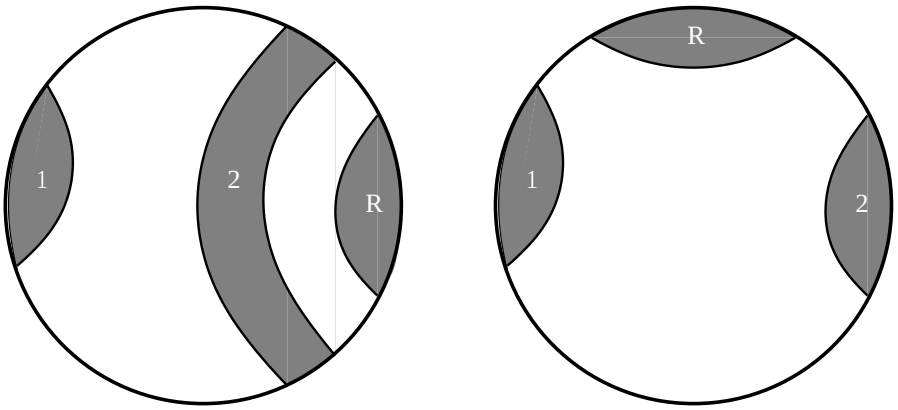
Agora mostraremos que os obstáculos serão unidos em uma única componente conexa de $OV(P\rho)$. De acordo com o Corolário 6 o efeito de ρ no grafo de entrelaçamento de arcos será o complemento do subgrafo induzido pelos arcos contidos em S, para S conjunto dos arcos que entrelaçam ρ. Uma vez que, pelo menos, um dos arcos em cada obstáculo pertence a S teremos que, quando fizermos o complemento do subgrafo induzido, serão criadas arestas partindo de todos os vértices de H_1 que pertecem a S para todos os vértices de H_2 que pertencem a S. As arestas entre vértices que pertencem a S e vértices que não pertencem a S são mantidas. Tais fatos garantem que os componentes irão se fundir por completo.

Quanto ao surgimento de, pelo menos, um arco orientado no componente conexo resultante da fusão, observe que temos cardinalidade de S maior ou igual a dois já que os arcos α de H_1 e β de H_2 entrelaçam ρ. Sendo assim, como todos os arcos em P são ruins, de acordo com o Corolário 6, teremos $|S|$ arcos bons em $P\rho$.

Para finalizar, precisamos agora mostrar que $h(P\rho) = 0$. Observe que se $\overline{h}(P) = 0$ então teremos $h(P\rho) = 0$. Temos que verificar os casos onde $\overline{h}(P) \geq 1$. Faremos uma análise de casos, quando $\overline{h}(P) = 1$, para verificar as possíveis posições do componente ruim em relação aos obstáculos. Em seguida mostraremos que $h(P\rho) = 0$. Para concluir, mostraremos que o mesmo ocorre quando $\overline{h}(P) > 1$.

São três casos possíveis para o posicionamento do componente ruim, que chamaremos de R, em relação aos obstáculos, que chamaremos de 1 e 2, no diagrama de pontos de quebra. Apresentamos o primeiro caso na Figura 6.7(a). Observe que o componente rotulado por 2 não é um obstáculo. Sendo assim, esta não é uma posição válida para R. Na Figura 6.7(b) apresentamos o segundo caso. Observe que o componente R, que deveria ser apenas um componente ruim, é um obstáculo. Sendo assim, esta também não é uma posição válida para R.

Finalmente, apresentamos na Figura 6.8 o último caso. Observe que os componentes 1 e 2 assumem o papel de obstáculos e R o de componente ruim. Com isso, concluímos que, se $h(P) = 2$ e $\overline{h}(P) = 1$, então a única possibilidade de posicionameto do componente ruim R, em relação aos obstáculos (1 e 2), é a apresentada na Figura 6.8.



(a) R protegendo 2. R é obstáculo. (b) R entre 1 e 2. R é obstáculo.

Figura 6.7: Estudo de caso para determinação dos possíveis posicionamentos de um componente ruim (R) em relação aos obstáculos (1 e 2) em um problema P com $h(P) = 2$.

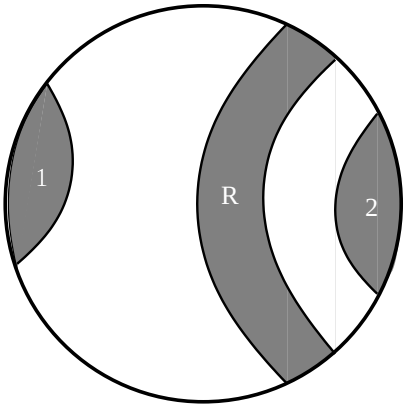


Figura 6.8: Estudo de caso para determinação dos possíveis posicionamentos de um componente ruim (R) em relação aos obstáculos (1 e 2) em um problema P com $h(P) = 2$.

Agora observe que, pelo menos um arco de R, digamos γ, intercala com $\rho = \rho(a, b, \pi)$, onde $a \in Ext(H_1)$ e $b \in Ext(H_2)$, e, assim, podemos afirmar que γ entrelaça ρ, o que, finalmente, garante que R estará no mesmo componente conexo de H_1 e H_2 em $P\rho$ e, assim, $h(P\rho) = 0$.

Para concluir a demonstração basta mostrar que $h(P\rho) = 0$ quando $\overline{h}(P) > 1$. De acordo com as observações anteriores podemos concluir que, se houver mais que um componente ruim, estes estarão em paralelo com R e também serão fundidos ao componente que conterá H_1 e H_2 em $P\rho$. □

6.5 Algoritmo algébrico

Seja $P = (\pi, \sigma, \Gamma)$ um problema monocromossomal e ϕ uma transformação em P. Dizemos que ϕ é uma **transformação equivalente** caso a variação do parâmetro $(\|\sigma\pi^{-1}\|/2 + h + f)$, denotado por $\delta(\|\sigma\pi^{-1}\|/2 + h + f)$, antes e depois de aplicar ϕ a P seja zero. Uma decomposição é **equivalente** caso sua transformação seja equivalente.

Seja $P = (\pi, \sigma, \Gamma)$ um problema e ρ uma reversão em P. Dizemos que ρ é **segura** se $\Delta(\|\sigma\pi^{-1}\|/2 + h) = -1$. Dizemos que ρ é **ordenante** caso $\Delta(\|\sigma\pi^{-1}\|/2 + h + f) = -1$.

Nosso **algoritmo** para solução do problema da distância de reversão é apresentado a seguir. A demonstração de sua corretude é apresentada em seguida.

```
    Ordena(P)
01. SE P é trivial
02.    retorna ∅
03. SENÃO SE P é particionável
04.    P_a,...,P_b ⇐ particiona(P)
05.    retorna Ordena(P_a) +...+ Ordena(P_b)
06. SENÃO SE P é não monocromossomal
07.    Interrompe a execução com um erro
08. SENÃO SE P possui quebra(s)
09.    ρ ⇐ seleciona uma quebra ordenante
10.    retorna ρ + Ordena(Pρ)
11. SENÃO SE h(P) = 1
12.    ρ ⇐ seleciona um corte de obstáculo
13.    retorna ρ + Ordena(Pρ)
14. SENÃO SE h(P) = 2
15.    ρ ⇐ seleciona uma junção de obstáculos
16.    retorna ρ + Ordena(Pρ)
17. SENÃO SE h(P) = 3 e P não é uma fortaleza
18.    ρ ⇐ seleciona um corte de um obstáculo simples
19.    retorna ρ + Ordena(Pρ)
20. SENÃO SE h(P) = 3 e P é uma fortaleza
21.    ρ ⇐ seleciona uma junção de obstáculos
22.    retorna ρ + Ordena(Pρ)
23. SENÃO /* h(P) ≥ 4 */
24.    φ ⇐ seleciona uma transformação equivalente
25.    l ⇐ Ordena(Pφ)
26.    retorna l substituindo cada reversão ρ por φ • ρ
```

O algoritmo recebe um problema como entrada e retorna uma lista ordenada de reversões. O símbolo $\emptyset$ indica uma lista vazia e o operador + é utilizado para concatenar duas listas.

6.5.1 Componentes Bons

Seja P um problema e α um vértice orientado no grafo de entrelaçamento de arcos de P. O **valor**[1] de α, denotado por $valor(\alpha)$, é dado por:

$$valor(\alpha) = T + U - O - 1$$

onde T é o número de vértices orientados no grafo, U é o número de vértices não orientados e adjacentes a α, e O é o número de vértices orientados e adjacentes a α.

Lema 35. *Uma quebra de valor máximo é segura.*

Demonstração. Suponha que α seja um vértice preto de valor máximo, e que a reversão induzida por α cria um novo componente não orientado C contendo mais que um vértice. Pelo menos um dos vértices em C deve ter sido adjacente a α, desde que apenas são afetadas pela reversão, arestas entre vértices adjacentes a α. Seja β um vértice adjacente a α, contido em C e considere os valores de α e β:

$$valor(\alpha) = T + U - O - 1$$
$$valor(\beta) = T + U' - O' - 1$$

Observe que todo vértice não orientado e adjacente a α deve ser adjacente a β. Pois caso exista um vértice não orientado adjacente a α e não adjacente a β tornaria-se orientado e adjacente a β, fazendo com que C fosse orientado e contrariando o que assumimos no início. Desta forma, temos que $U' \geq U$.

Note que todo vértice orientado adjacente a β deve ser adjacente a α. Pois caso exista um vértice orientado adjacente a β e não adjacente a α permaneceria orientado e adjacente a β, fazendo com que C fosse orientado e contrariando também o que assumimos no início. Desta forma, temos que $O \geq O'$.

Caso $O' = O$ e $U' = U$, os vértices α e β possuem o mesmo conjunto de vértices adjacentes e complementando o subgrafo de α e seus vértices adjacentes irá isolar α e β. E assim C tem apenas 1 vértice.

Caso $O > O'$ e $U' = U$, ou $U' > U$ e $O = O'$, ou ainda $O > O'$ e $U' > U$, teremos valor de β maior que valor de α, o que contraria o que foi assumido inicialmente e o que nos mostra que não há componente não orientado e não trivial novo. □

[1] Traduzimos por "valor" o termo, em inglês, *score*.

Lema 36. *Se $P = (\pi, \sigma, \Gamma)$ é um componente bom então há uma reversão ordenante.*

Demonstração. Sendo P um componente bom então podemos afirmar que há, pelo menos, uma quebra. Seja ρ uma quebra de pontuação máxima em P. O Lema 35 nos garante que ρ é segura e, então, temos que $\Delta(\|\sigma\pi^{-1}\|/2 + h) = -1$. Como o parâmetro f não é afetado por quebras seguras então chegamos a conclusão de que ρ é ordenante. □

6.5.2 Componentes Ruins

Lema 37. *Seja P um obstáculo e α um dos arcos em P. Se $\rho(\alpha, \pi)$ é a reversão induzida por α em π então $\rho(\alpha, \pi)$ é segura.*

Demonstração. Observe que α é não orientado. Sendo assim, de acordo com o Lema 32, temos $\|\rho\pi^{-1}\sigma\| = \|\pi^{-1}\sigma\|$.

Para que ρ seja segura precisamos que $h(P\rho) = h(P) - 1$. Tal fato é garantido pelo Lema 33. □

Lema 38. *Seja $P = (\pi, \sigma, \Gamma)$ um componente. Caso P seja um obstáculo então um corte de obstáculos qualquer em P define uma reversão ordenante.*

Demonstração. Seja α um dos arcos de P. Seja ρ a reversão induzida por α em π. De acordo com o Lema 37 temos que ρ é segura. Como P não possui fortaleza e $P\rho$ não possui componentes ruins então $\Delta f = 0$ e ρ é ordenante. □

Lema 39. *Seja $P = (\pi, \sigma, \Gamma)$ um problema tal que $h(P) = 2$. Se ρ é uma junção de obstáculos em P então ρ define uma reversão segura.*

Demonstração. De acordo com o Lema 34 temos que $\Delta h = -2$. Como ρ é uma junção então $\|\rho\pi^{-1}\sigma\| = \|\pi^{-1}\sigma\| + 2$. Segue que:

$$\frac{\|\rho\pi^{-1}\sigma\|}{2} + h(\pi\rho) = \frac{\|\pi^{-1}\sigma\| + 2}{2} + h(\pi) - 2 = \frac{\|\pi^{-1}\sigma\|}{2} + h(\pi) - 1$$

E, assim, concluímos que ρ é segura. □

Lema 40. *Seja P um problema tal que $h(P) = 2$. Se ρ é uma junção de obstáculos em P então ρ define uma reversão ordenante.*

Demonstração. De acordo com o Lema 39 podemos afirmar que ρ é segura. Como $h(P) = 2$ e $h(P\rho) = 0$ então não há fortalezas em qualquer momento. Assim $\Delta f = 0$ e ρ é ordenante. □

Lema 41. *Seja $P = (\pi, \sigma, \Gamma)$ um problema tal que $h(P) = 3$ e $f(P) = 0$. Se H é um obstáculo simples em P então o corte de H é ordenante.*

Demonstração. Observe que

$$d_r(P) = \frac{\|\pi^{-1}\sigma\|}{2} + h(P) + f(P) = \frac{\|\pi^{-1}\sigma\|}{2} + 3$$

Observe também que, como ρ é um corte, $\|\rho\pi^{-1}\sigma\| = \|\pi^{-1}\sigma\|$. De acordo com o Lema 33 podemos afirmar que $h(P\rho) = h(P) - 1$ e, por conseguinte, $f(P\rho) = f(P) = 0$. Assim

$$\begin{aligned} d_r(P\rho) &= \frac{\|\rho\pi^{-1}\sigma\|}{2} + h(P\rho) + f(P\rho) \\ &= \frac{\|\pi^{-1}\sigma\|}{2} + (h(P) - 1) + (f(P) + 0) \\ &= \frac{\|\pi^{-1}\sigma\|}{2} + h(P) + f(P) - 1 \\ &= d_r(P) - 1 \end{aligned}$$

o que nos leva a concluir que ρ é ordenante. □

Lema 42. *Seja $P = (\pi, \sigma, \Gamma)$ um problema tal que $h(P) = 3$ e $f(P) = 1$. Se ρ é uma junção de obstáculos então $h(P\rho) = 2$ e $f(P\rho) = 0$.*

Demonstração. Inicialmente iremos verificar o número mínimo de componentes ruins necessário em P para que P seja fortaleza, juntamente com as posições destes em relação aos obstáculos.

Para começar distribuímos os obstáculos no diagrama de pontos de quebra. A única configuração possível é apresentada na Figura 6.9.

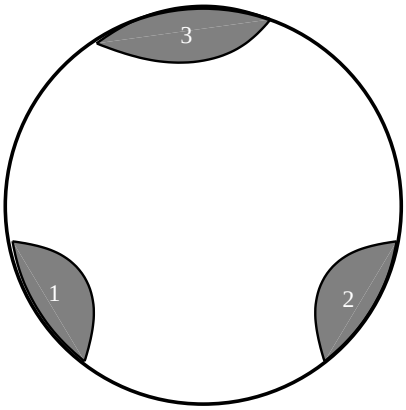


Figura 6.9: Estudo de caso para determinação dos possíveis posicionamentos de componentes ruins em relação aos obstáculos (1, 2 e 3) em um problema P com $h(P) = 3$.

Agora faremos uma análise com $\overline{h}(P) = 1$. Verificaremos quais as possíveis posições de R1, o único componente ruim, em relação aos obstáculos. Verificaremos também quantos super obstáculos existem em cada caso. Na Figura 6.10(a) apresentamos o primeiro caso.

Neste R1 protege 1. Observe que este não é um caso válido pois, assim, R1 é obstáculo e 1 é apenas componente ruim.

O segundo caso apresentamos na Figura 6.10(b). Neste, R1 é vizinho dos obstáculos 1 e 3. E, como podemos ver, R1 é obstáculo também, fazendo com que $h(P) = 4$, contrariando o enunciado que diz que $h(P) = 3$.

Na Figura 6.10(c) apresentamos o terceiro caso. Agora R1 encontra-se distribuido entre 1 e 2, entre 2 e 3 e entre 3 e 1. Observe que $h(P) = 3$ porém nenhum dos obstáculos é super. Como estamos em busca de uma fortaleza, esta situação não é de interesse.

O último dos casos apresentamos na Figura 6.10(d). Neste podemos ver que $h(P) = 3$ e que 1 é um super obstáculo. Finalmente, concluímos que, com apenas um componente ruim, conseguimos apenas um super obstáculo.

De acordo com as observações anteriores podemos ver que, com dois componentes ruins, podemos, na melhor das hipóteses, conseguir dois super obstáculos. Apresentamos tal caso na Figura 6.11(a). E, finalmente, na Figura 6.11(b) apresentamos o caso em que $h(P) = 3$, $\overline{h}(P) = 3$ e $f(P) = 1$.

Note que para cada componente ruim podemos ter, na verdade, um bloco de componentes ruins em paralelo com pelo menos um componente ruim. Assim concluímos que, se $h(P) = 3$ e $f(P) = 1$ então $\overline{h}(P) \geq 3$.

Uma vez que conhecemos a estrutura de uma fortaleza com três obstáculos agora precisamos observar o efeito de junção de obstáculos. Observe que se ρ é uma junção de 1 e 2 então R1 e R2 entrelaçam ρ. Se ρ é uma junção de 2 e 3 então R2 e R3 entrelaçam ρ e se ρ é uma junção de 3 e 1 então R3 e R1 entrelaçam ρ.

Desta forma, ao aplicar a junção ρ de dois dos obstáculos teremos que os dois obstáculos serão fundidos em um único componente juntamente com os blocos por eles protegidos em $P\rho$ e este grande componente será bom. Teremos ainda o obstáculo remanescente, que continuará obstáculo, e seu bloco protegido. Observe ainda que um dos componente ruins irá se tornar obstáculo.

Finalmente concluímos que $h(P\rho) = 2$ e conseqüentemente $f(P\rho) = 0$.

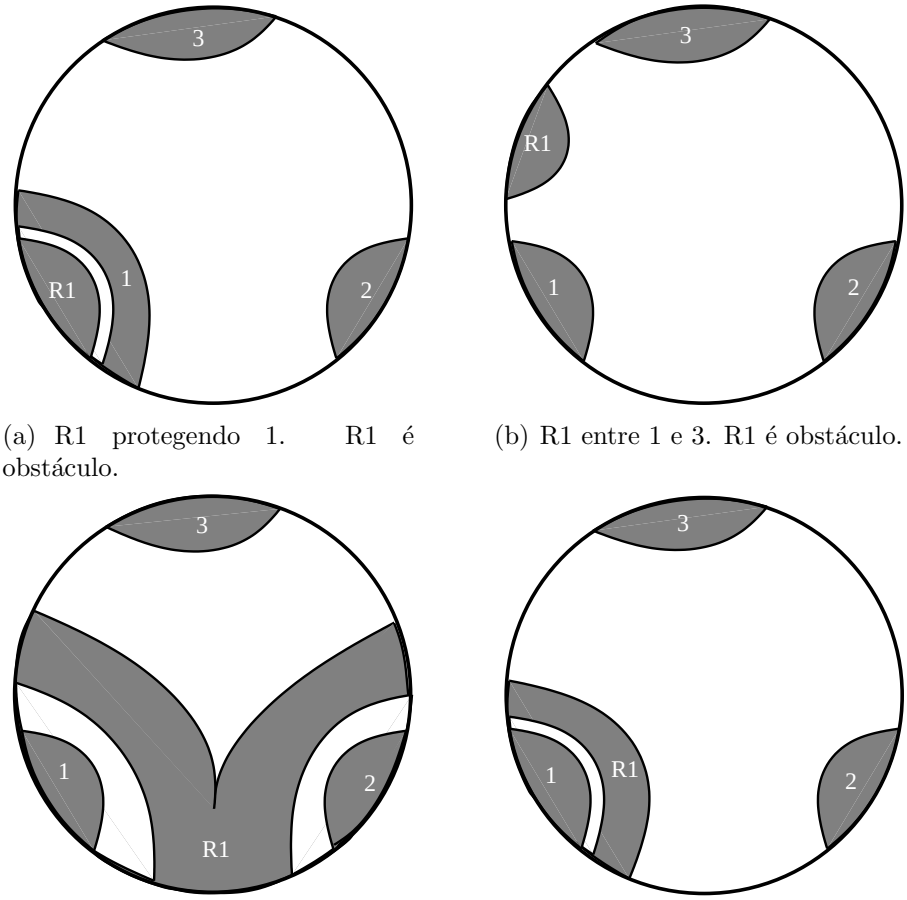
□

Lema 43. *Seja $P = (\pi, \sigma, \Gamma)$ um problema tal que $h(P) = 3$ e $f(P) = 1$. Se ρ é uma junção de obstáculos então ρ define uma reversão ordenante.*

Demonstração. Observe que

$$d_r(P) = \frac{\|\pi^{-1}\sigma\|}{2} + h(P) + f(P) = \frac{\|\pi^{-1}\sigma\|}{2} + 3 + 1 = \frac{\|\pi^{-1}\sigma\|}{2} + 4$$

De acordo com Lema 42 podemos afirmar que $h(P\rho) = 2$ e $f(P\rho) = 0$. Finalmente, como ρ é junção então $\|\rho\pi^{-1}\sigma\| = \|\pi^{-1}\sigma\| + 2$. O que nos leva a



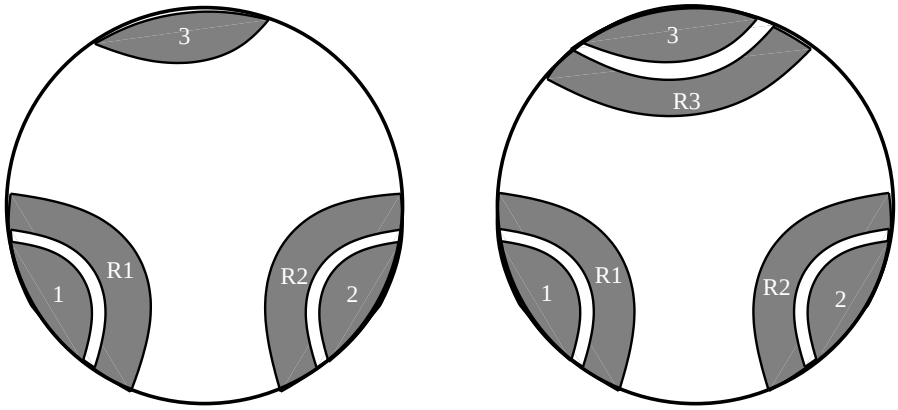
(a) R1 protegendo 1. R1 é obstáculo.

(b) R1 entre 1 e 3. R1 é obstáculo.

(c) R1 entre 1 e 2, entre 2 e 3 e entre 3 e 1. Todos os obstáculos são simples.

(d) R1 sendo protegido por 1. Apenas o obstáculo 1 é super.

Figura 6.10: Estudo de caso para determinação dos possíveis posicionamentos de um componente ruim (R1) em relação aos obstáculos (1, 2 e 3) em um problema P com $h(P) = 3$ e $\overline{h}(P) = 1$.



(a) P com $h(P) = 3$ e $\overline{h}(P) = 2$. Apenas os obstáculos 1 e 2 são super.

(b) P com $h(P) = 3$ e $\overline{h}(P) = 3$. Todos os obstáculos são super.

Figura 6.11: Estudo de caso para determinação dos possíveis posicionamentos de componentes ruins (R1, R2 e R3) em relação aos obstáculos (1, 2 e 3) em um problema P com $h(P) = 3$ e com dois ou três componentes ruins.

$$\begin{aligned} d_r(P\rho) &= \frac{\|\rho\pi^{-1}\sigma\|}{2} + h(P\rho) + f(P\rho) \\ &= \frac{\|\pi^{-1}\sigma\|+2}{2} + (h(P)-1) + (f(P)-1) \\ &= \frac{\|\pi^{-1}\sigma\|}{2} + h(P) + f(P) - 1 \\ &= d_r(P) - 1 \end{aligned}$$

e, assim, ρ é ordenante. □

Conjectura 4. *Seja P um problema monocromossomal. Se $h(P) \geq 4$, então há uma transformação equivalente em P.*

Capítulo 7

Conclusão

Ao final deste trabalho conseguimos avançar significativamente em direção a uma solução completamente algébrica para o problema de rearranjo por reversões com sinal.

Alteramos a definição de entrelaçamento entre arcos de forma a simplificar seu entendimento e assimilação, assim como, facilitar sua manipulação. Generalizamos o conceito de entrelaçamento definindo o entrelaçamento entre arcos e reversões. Definimos o grafo de entrelaçamento de arcos de um problema e apresentamos uma demonstração do efeito de uma reversão no grafo, inclusive com um resultado novo. Mostramos o efeito de uma reversão qualquer no grafo.

Definimos conceitos tais como transformação e decomposição, que permitem particionar um problema em subproblemas de solução mais simples. Nosso algoritmo utiliza decomposições para reduzir os problemas a uma das categorias em que é capaz de encontrar reversões ordenantes. Apresentamos o esboço de uma demonstração que correlaciona componentes oriundos de decomposições e os componentes conexos do grafo de entrelaçamento de arcos.

Fizemos diversas alterações no formalismo algébrico, que resultaram em definições compatíveis de conceitos e objetos em nossa teoria algébrica e na teoria clássica fundamentada no trabalho de Hannenhalli e Pevzner.

Por fim apresentamos um algoritmo que resolve o problema da distância de reversão e um esboço da demonstração de sua corretude.

Muitos são os trabalhos futuros. Dentre eles destacamos a conclusão da demonstração aqui esboçada e a implementação de um algoritmo eficiente. Este trabalho desde sua concepção teve como objetivo fazer uma tradução direta dos conceitos e solução já conhecida em rearranjo de genomas. Poderíamos agora trabalhar em uma solução estritamente algébrica, deixando de lado a preocupação em manter a correlação com a teoria clássica.

Outros trabalhos seriam a definição da freqüência de cada tipo de operação nas diversas categorias de espécies, utilizar o formalismo algébrico na solução de diversos problemas

abertos na área de rearranjo de genomas, assim como, estudar o "poder" dos diversos tipos de operações e verificar a relação deste "poder" com a freqüência em que ocorrem em determinadas famílias de espécies.

Poderíamos ainda verificar a aplicabilidade do formalismo algébrico ao problema do genoma mediano. Em tal problema são dados três genomas como entrada e busca-se por um genoma cuja soma das distâncias deste em relação a cada um dos três seja mínima.

Apêndice A

Revisão Bibliográfica

Este apêndice apresenta resumos de trabalhos importantes em distância de reversão. Juntamente com o trabalho de Meidanis e Dias [17] foram a base deste trabalho.

Na Seção A.1 apresentamos o trabalho de Hannenhalli e Pevzner [11], onde mostraram a primeira solução polinomial $O(n^4)$ para o problema da distância de reversão com sinais. Simplificações na teoria levaram Kaplan e colegas [13] a um algoritmo $O(n^2)$, que é apresentado na Seção A.2. Na Seção A.3 apresentamos o trabalho de Bergeron [5], onde a teoria definida por Hannenhalli e Pevzner foi bastante simplificada. O trabalho de Meidanis e Dias foi apresentado no Capítulo 5.

A.1 Transforming Cabbage into Turnip: Polynomial Algorithm for Sorting Signed Permutation by Reversals

Esta seção apresenta um resumo do trabalho, cujo título é apresentado acima, desenvolvido por Hannenhalli e Pevzner [11]. Tal trabalho, que é uma versão estendida do trabalho publicado em 1995, apresenta o primeiro algoritmo polinomial para o problema de ordenar por reversões orientadas.

A.1.1 Introdução

Na natureza, freqüentemente genomas evoluem através de reversões $\rho(i, j)$. Tais reversões transformam uma seqüência de genes $\pi_1 \dots \pi_i \pi_{i+1} \dots \pi_{j-1} \pi_j \dots \pi_n$ na seqüência $\pi_1 \dots \pi_j \pi_{j-1} \dots \pi_{i+1} \pi_i \dots \pi_n$. A distância de reversão entre duas permutações π e σ, que representam dois genomas, é dada pelo número mínimo de reversões necessárias para transformar π em σ. Análises de rearranjo de genomas, na biologia molecular, tiveram

seu início na década de 1930 com Dobzhansky e Sturtevant [28]. Tais pesquisadores apresentaram um cenário de rearranjo com inversões entre espécies de *Drosophila*. Análises de evolução de genomas por inversões leva a um problema combinatorial de ordenação por reversões, que vem sendo estudado em detalhes recentemente. Ordenação por reversões orientadas modela adequadamente rearranjos em genomas pequenos tais como DNA de cloroplasto e mitocondrial.

Motivação e embasamento biológico

Na década de 1980, Jeffrey Palmer e colegas [22], compararam os genomas mitocondriais da *Brassica oleracea* (repolho) e da *Brassica campestris* (nabo). Eles perceberam que as seqüências dos genes destas duas espécies eram quase idênticas, muitos de seus genes são 99% a 99,9% idênticos. Porém a ordem em que estes genes aparecem na molécula de DNA é extremamente diferente. Este e diversos outros estudos mostram que rearranjo de genomas é uma forma comum de evolução do DNA de cloroplasto, mitocondrial, viral e bacteriano [3].

Em rearranjo de genomas, os genes são rotulados de 1 a n e sua ordem nas moléculas dos dois organismos é representada por permutações, tal como $\pi = (\pi_1\pi_2\ldots\pi_n)$ e $\sigma = (\sigma_1\sigma_2\ldots\sigma_n)$. Uma *reversão* $\rho(i,j)$ é a permutação

$$\begin{pmatrix} 1\ 2\ \ldots\ i-1\ i\ i+1\ \ldots\ j-1\ j\ j+1\ \ldots\ n \\ 1\ 2\ \ldots\ i-1\ j\ j-1\ \ldots\ i+1\ i\ j+1\ \ldots\ n \end{pmatrix}$$

A aplicação de uma reversão ρ em uma permutação π, $\pi\rho$, reverte a ordem dos genes $i\ i+1\ \ldots\ j-1\ j$. Se, no caso, estiver trabalhando com genomas orientados, onde a cada elemento da permutação é associado um sinal + ou −, a reversão, além de reverter a ordem dos genes, altera a orientação de cada um dos genes. Dado que numa molécula de DNA há duas fitas, uma complementar à outra. Cada gene i está presente nas duas fitas, só que em orientações opostas. Os sinais + e − são utilizados para simbolizar as duas orientações.

Dadas as permutações π e σ, uma solução para o *problema da distância de reversão* é uma série de reversões ρ_1, ρ_2, ..., ρ_t, tal que $\pi\rho_1\rho_2\ldots\rho_t = \sigma$ e t é mínimo. A *distância de reversão* entre π e σ é dada por t. Para *ordenar π por reversões* basta encontrar a distância de reversão $d(\pi)$ entre π e a permutação identidade $[1, 2, \ldots, n]$.

Resultados anteriores

Uma abordagem computacional baseada na comparação da ordem de genes ao invés da comparação de seqüências de DNA foi inicialmente estudada por Sankoff [24, 23, 25]. Kececioglu e Sankoff [15] formularam o problema da distância de reversão e definiram

os limites inferior e superior para a distância de reversão. Em seguida apresentaram o primeiro algoritmo de aproximação para ordenação por reversões, que apresentava solução exata para alguns instâncias difíceis. Posteriormente, Bafna e Pevzner [4] introduziram a noção de *diagrama de pontos de quebra* de uma permutação e apresentaram uma relação entre a *máxima decomposição em ciclos* de um diagrama e a distância de reversão.

Diagrama de pontos de quebra e decomposição em ciclos

A noção de *pontos de quebra* foi introduzida por Watterson e colegas [31] e Nadeau e Taylor [20], quando percebeu-se uma correlação entre a distância de reversão e o número de pontos de quebra.

Se $i \sim j$, então $|i - j| = 1$. *Estende-se* uma permutação $\pi = \pi_1 \pi_2 \dots \pi_n$ adicionando-se $\pi_0 = 0$ e $\pi_{n+1} = n + 1$. Diz-se que uma par de elementos (π_i, π_{i+1}), $0 \leq i \leq n$, de π é uma *adjacência* se $\pi_i \sim \pi_{i+1}$, e um *ponto de quebra* se $\pi_i \nsim \pi_{i+1}$. Como ordenar por reversões corresponde a eliminar pontos de quebra e toda reversão pode eliminar no máximo 2 pontos de quebra, segue que $d(\pi) \geq b(\pi)/2$, onde $b(\pi)$ é o número de pontos de quebra de π.

O *diagrama de pontos de quebra* de uma permutação π é um grafo colorido nas arestas $G(\pi)$ com $n + 2$ vértices $\{\pi_0, \pi_1, \dots, \pi_n, \pi_{n+1}\}$. Os vértices π_i e π_j são ligados por uma *aresta preta* caso (π_i, π_j) seja um ponto de quebra em π ($\pi_i \nsim \pi_j$ e $i \sim j$) e são ligados por uma *aresta cinza* caso (i, j) seja um ponto de quebra em π^{-1} ($\pi_i \sim \pi_j$ e $i \nsim j$). A Figura A.1 apresenta um exemplo de diagrama de pontos de quebra.

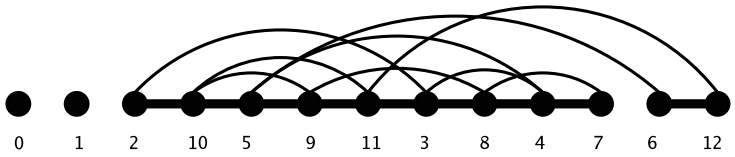


Figura A.1: Diagrama de pontos de quebra de uma permutação não sinalizada.

Um *ciclo* C em um grafo colorido nas arestas G é *alternante* se as cores de cada duas arestas consecutivas são distintas. No restante do texto, por ciclo entenda-se ciclo alternante. O *tamanho* de um ciclo C, $l(C)$, é o número de arestas pretas contidas nele. Diz-se que um ciclo C é *curto* caso $l(C) = 2$ e *longo* caso $l(C) > 2$. Uma permutação π é *simples* caso não possua ciclos longos.

A *máxima decomposição em ciclos* de $G(\pi)$ é o número máximo $c(\pi)$ de ciclos alternantes disjuntos nas arestas que podem ser obtidos em $G(\pi)$. Para a permutação $\pi = (0\ 1\ 2\ 10\ 5\ 9\ 11\ 3\ 8\ 4\ 7\ 6\ 12)$, na Figura A.1, $c(\pi) = 4$, pois $G(\pi)$ pode ser decomposto em três ciclos curtos $(2, 3, 11, 10, 2)$, $(5, 9, 8, 4, 5)$ e $(4, 7, 8, 3, 4)$ e um ciclo longo $(6, 12, 11, 9, 10, 5, 6)$. Bafna e Pevzner [4] mostraram que uma reversão pode alterar o

parâmetro $b(\pi) - c(\pi)$ por no máximo 1, e assim um limite melhor para a distância de reversão é:

$$d(\pi) \geq b(\pi) - c(\pi) \tag{A.1}$$

Novos resultados

Bafna e Pevzner [4] estenderam o conceito de diagrama de pontos de quebra para o caso de permutações sinalizadas. Kececioglu e Sankoff [14] observaram que a média da diferença entre o limite apresentado na Equação A.1 e a distância exata é menor que um para permutações aleatórias. Neste trabalho, Hannenhalli e Pevzner apresentam um terceiro parâmetro (número de *obstáculos* em π), o qual é responsável por torna a ordenação mais difícil. Neste trabalho é provado que

$$b(\pi) - c(\pi) + h(\pi) \leq d(\pi) \leq b(\pi) - c(\pi) + h(\pi) + 1 \tag{A.2}$$

O parâmetro $h(\pi)$ denota o número de obstáculos em π. Foi baseado neste resultado que os autores desenvolveram um algoritmo polinomial para ordenação por reversões orientadas. Tal algoritmo foi o primeiro polinomial para um modelo realístico de rearranjo de genomas.

A.1.2 Diagrama de pontos de quebra para permutações sinalizadas

Define-se uma transformação de uma permutação sinalizada π de ordem n para uma permutação não sinalizada π' de ordem $2n$ como segue: substitua um elemento positivo $+x$ em π por $2x - 1$, $2x$ em π' e elementos negativos $-x$ em π por $2x$, $2x - 1$ em π'. A Figura A.2 apresenta um exemplo de diagrama de pontos de quebra de uma permutação sinalizada. O diagrama de pontos de quebra para genomas orientados possui uma única decomposição em ciclos, que o torna mais fácil de ser tratado.

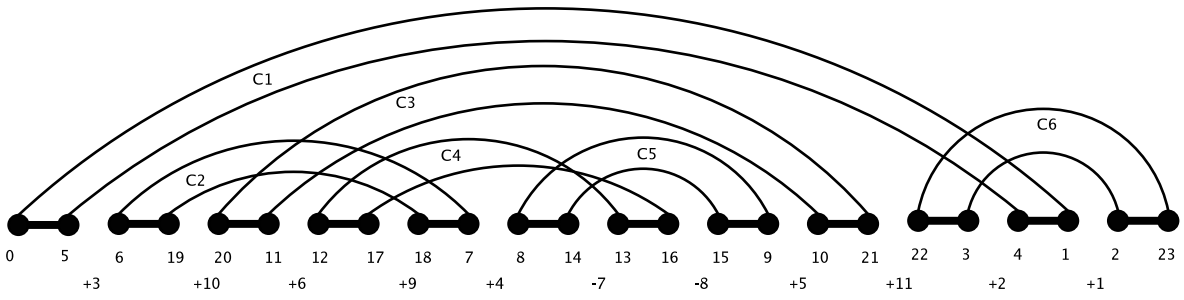


Figura A.2: Diagrama de pontos de quebra de uma permutação sinalizada.

O efeito de uma reversão $\rho(i+1, j)$ em uma permutação sinalizada π é imitado pela reversão $\rho(2i+1, 2j)$ na permutação não sinalizada π', isto implica que $d(\pi) = d(\pi')$ caso cortes entre π'_{2i-1} e π'_{2i} sejam proibidos [4].

Dada uma permutação π e uma reversão ρ, denota-se por Δb o incremento no número de pontos de quebra em π quando aplicada ρ e por Δc o incremento no número de ciclos em π quando aplicada ρ. Bafna e Pevzner [4] provaram que $\Delta(b - c) = \Delta b - \Delta c \geq -1$. Uma reversão é *própria* caso $\Delta(b - c) = -1$.

Sejam os pontos de quebra (π_{i-1}, π_i) e (π_j, π_{j+1}), diz-se que a reversão $\rho(i, j)$ *atua* nas arestas pretas (π_{i-1}, π_i) e (π_j, π_{j+1}). Uma reversão $\rho(i, j)$ *atua em um ciclo* C de $G(\pi)$ caso as arestas pretas (π_{i-1}, π_i) e (π_j, π_{j+1}) pertençam a C. Um aresta cinza g é *orientada* caso uma reversão atuando nas duas arestas pretas incidentes a g seja própria, e é *não orientada* caso tal reversão não exista. Por exemplo, no diagrama da Figura A.2, a aresta cinza $(12, 13)$ é orientada, pois uma reversão atuando nas arestas pretas $(12, 17)$ e $(13, 16)$ destrói dois pontos de quebra e um ciclo. Já a aresta cinza $(6, 7)$ é não orientada. O Lema 44 apresenta um modo mais intuitivo de dizer se uma aresta cinza é ou não orientada. Pode-se ler o Lema 44 ainda como: uma aresta cinza é orientada caso seja incidente a mesma extremidade em ambas as arestas pretas.

Lema 44. *Seja (π_i, π_j) uma aresta cinza incidente às arestas pretas (π_k, π_i) e (π_j, π_l). Então (π_i, π_j) é orientada se, e somente se, $i - k = j - l$.*

Um ciclo em $G(\pi)$ é *orientado* caso possua pelo menos uma aresta cinza orientada, caso contrário ele será *não orientado*. No diagrama apresentado na Figura A.2, o ciclo $C4$ é orientado, enquanto os ciclos $C1$, $C2$, $C3$, $C5$ e $C6$ são não orientados. Desta definição pode-se derivar que não existe reversões próprias atuando em ciclos não orientados e que uma permutação só possuirá uma reversão própria caso tenha um ciclo orientado.

A.1.3 Grafo de entrelaçamento e obstáculos

Duas arestas cinzas (π_i, π_j) e (π_k, π_t) em $G(\pi)$ estão *entrelaçadas* caso os intervalos $[i, j]$ e $[k, t]$ sobrepõem-se, mas um não contém propriamente o outro. Por exemplo, no diagrama da Figura A.2, as arestas $(6, 7)$ e $(12, 13)$ estão entrelaçadas enquanto as arestas $(20, 21)$ e $(12, 13)$ não estão. Dois ciclos C_1 e C_2 estão *entrelaçados* caso existam as arestas cinzas entrelaçadas $g_1 \in C_1$ e $g_2 \in C_2$. No exemplo da Figura A.2 os ciclos $C2$ e $C4$ estão entrelaçados.

Seja C_π o conjunto de ciclos no diagrama de pontos de quebra de uma permutação π. Define-se o *grafo de entrelaçamento* $H_\pi(C_\pi, I_\pi)$ de π com o conjunto de arestas

$$I_\pi = \{(C_1, C_2) : C_1 \text{ e } C_2 \text{ são ciclos entrelaçados em } G(\pi)\}.$$

Figura A.3: Grafo de entrelacamento

A Figura A.3 apresenta o grafo de entrelaçamento para o diagrama de pontos de quebra mostrado na Figura A.2. Dado que os vértices de H_π são ciclos em C_π, diz-se que um vértice é *orientado* caso o ciclo ao qual está relacionado é orientado, e é não orientado caso contrário. Um componente conexo de H_π é *orientado* caso tenha pelo menos um vértice orientado, e *não orientado* caso contrário. Para um componente conexo U, define-se as posições mais a esquerda e mais a direita de U como

$$U_{min} = min_{C \in U}\ min_{\pi_i \in C}\ i \text{ e } U_{max} = max_{C \in U}\ max_{\pi_i \in C}\ i$$

Seja $Estende(U)$ o intervalo $[U_{min}, U_{max}]$. Diz-se que um componente U *separa* os componentes U' e U'' em π caso exista uma aresta cinza (π_i, π_j) em U tal que $Estende(U') \subset [i, j]$, porém $Estende(U'') \cap [i, j] = \emptyset$.

Seja $\prec$ uma ordem parcial em um conjunto P. Um elemento $x \in P$ é *mínimo* em $\prec$ caso não exista um elemento $y \in P$ onde $y \prec x$. Um elemento $x \in P$ é máximo em $\prec$ caso $y \prec x$ para todo $y \in P$.

Seja U_π o conjunto de componentes não orientados em H_π e $\{U, W\} \subset U_\pi$, $U \prec W$ se, e somente se, $Estende(U) \subset Estende(W)$. Um componente não orientado $U \in U_\pi$ que é um elemento mínimo em $\prec$ é um *obstáculo*, chamado *obstáculo mínimo*. O componente não orientado $U \in U_\pi$ que é o elemento máximo em $\prec$ é um *obstáculo*, chamado *obstáculo máximo*, se U não separa dois obstáculos mínimos. A permutação π na Figura A.2 tem um componente não orientado e $h(\pi) = 1$. Já o exemplo na Figura A.4 apresenta $h(\pi) = 2$.

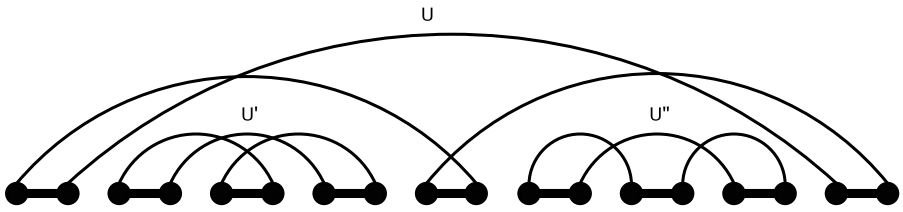


Figura A.4: Exemplo de obstáculo. Note que o componente não orientado U não é um obstáculo.

Teorema 2. *Para uma permutação (sinalizada) arbitrária* π, $d(\pi) \geq b(\pi) - c(\pi) + h(\pi)$.

A.1.4 Transformações de equivalência em permutações

Estudos anteriores mostram que complicadas estruturas de entrelaçamento de ciclos longos no diagrama de pontos de quebra dificultam a ordenação por reversão [4] e por transposição [3]. Tal problema é resolvido aqui através de transformações de equivalência, onde, se uma permutação $\pi \equiv \pi(0)$ tem um ciclo longo, ela é transformada em uma nova permutação $\pi(1)$ pela quebra deste ciclo longo em dois ciclos menores. Esta operação é aplicada sucessivamente $\pi \equiv \pi(0), \pi(1), \ldots, \pi(k) \equiv \sigma$ até que se alcance uma permutação simples.

Sejam $b = (v_b, w_b)$ uma aresta preta e $g = (w_g, v_g)$ uma aresta cinza pertencentes a um ciclo $C = \ldots, v_b, w_b, \ldots, w_g, v_g, \ldots$ no diagrama de pontos de quebra $G(\pi)$ de uma permutação π. Um (g, b)-corte de $G(\pi)$ é um novo diagrama $\hat{G}(\pi)$ obtido de $G(\pi)$ da seguinte forma:

- remove as arestas g e b
- adiciona dois novos vértices v e w
- adiciona duas novas arestas pretas (v_b, v) e (w, w_b)
- adiciona duas nova arestas cinzas (w_g, w) e (v, v_g)

Se $G(\pi)$ é um diagrama de pontos de quebra de uma permutação sinalizada π, então todo (g, b)-corte de $G(\pi)$ corresponde ao diagrama de pontos de quebra de uma permutação sinalizada *generalizada* $\hat{\pi}$ tal que $\hat{G}(\pi) = G(\hat{\pi})$.

Um permutação generalizada $\pi = (\pi_1 \pi_2 \ldots \pi_n)$ é uma permutação de reais arbitrários distintos. Nesta seção quando se falar em permutação, entenda-se permutação generalizada. A extensão de uma permutação $\pi = (\pi_1 \pi_2 \ldots \pi_n)$ é realizada adicionando-se $\pi_0 = min_{1 \leq i \leq n} \pi_i - 1$ e $\pi_{n+1} = max_{1 \leq i \leq n} \pi_i + 1$. Os elementos π_j e π_k de π são *consecutivos* se não existe um elemento π_l tal que $\pi_j < \pi_l < \pi_k$ para $1 \leq l \leq n$. Os elementos π_i e π_{i+1} de π são *adjacentes* para $0 \leq i \leq n$. O *diagrama de pontos de quebra de uma permutação generalizada* $\pi = (\pi_1 \pi_2 \ldots \pi_n)$ é definido sobre os vértices $\{\pi_0, \pi_1, \ldots, \pi_n, \pi_{n+1}\}$ com arestas pretas entre elementos adjacentes mas não consecutivos e arestas cinzas entre elementos consecutivos mas não adjacentes.

Sejam $b = (\pi_{i+1}, \pi_i)$ uma aresta preta e $g = (\pi_j, \pi_k)$ uma aresta cinza pertencentes a um ciclo $C = \ldots, \pi_{i+1}, \pi_i, \ldots, \pi_j, \pi_k, \ldots$ em um diagrama de pontos de quebra $G(\pi)$. Sejam ainda $\Delta = \pi_k - \pi_j$, $v = \pi_j + (\Delta/3)$ e $w = \pi_k - (\Delta/3)$. Um (g, b)-*preenchimento* de $\pi = (\pi_1 \pi_2 \ldots \pi_n)$ é uma permutação com $n + 2$ elementos obtida de π pela inserção de v e w depois do i-ésimo elemento de π $(0 \leq i \leq n)$:

$$\hat{\pi} = (\pi_1 \pi_2 \ldots \pi_i v w \pi_{i+1} \ldots \pi_n).$$

Note que v e w são consecutivos e adjacentes em $\hat{\pi}$.

Lema 45. *$\hat{G}(\pi) = G(\hat{\pi})$.*

Se g e b são arestas não incidentes de um ciclo longo C em $G(\pi)$, então o (g, b)-preenchimento quebra C em dois ciclos menores em $G(\hat{\pi})$. Desta forma, preenchimentos podem ser utilizados para transformar uma permutação arbitrária π em uma permutação simples. Note que $b(\hat{\pi}) = b(\pi) + 1$ e $c(\hat{\pi}) = c(\pi) + 1$.

Lema 46. *Seja um (g, b)-preenchimento em um ciclo C de $G(\pi)$ que exclui a aresta cinza g e adiciona duas novas arestas cinzas g_1 e g_2. Se g é orientada, então g_1 ou g_2 é orientada em $G(\hat{\pi})$. Se C é não orientada, então g_1 e g_2 são não orientadas em $G(\hat{\pi})$.*

Lema 47. *Seja um (g, b)-preenchimento que quebra um ciclo C de $G(\pi)$ em dois ciclos C_1 e C_2 de $G(\hat{\pi})$. Então C é orientado se, e somente se, C_1 ou C_2 é orientado.*

Lema 48. *Sejam g' e g'' duas arestas cinzas de $G(\pi)$ diferentes de g. Então g' e g'' estão entrelaçadas em π se, e somente se, g' e g'' estão entrelaçadas em um (g, b)-preenchimento de π.*

Lema 49. *Seja um (g, b)-preenchimento que quebra um ciclo C de $G(\pi)$ nos ciclos C_1 e C_2 de $G(\hat{\pi})$. Então todo ciclo D entrelaçado com C de $G(\pi)$ está entrelaçado com C_1 ou C_2 de $G(\hat{\pi})$.*

Lema 50. *Para toda aresta cinza g, existe um aresta cinza f entrelaçada com g em $G(\pi)$.*

Um (g, b)-preenchimento ϕ que transforma π em $\hat{\pi}$ é *seguro* caso atue em arestas não incidentes de um ciclo longo e $h(\pi) = h(\hat{\pi})$. Todo preenchimento seguro quebra um ciclo longo em dois ciclos menores.

Teorema 3. *Se C é um ciclo longo em $G(\pi)$, então existe um (g, b)-preenchimento seguro atuando em C.*

Uma permutação π é *equivalente* a uma permutação σ ($\pi \rightsquigarrow \sigma$) se existe uma série de permutação $\pi \equiv \pi(0), \pi(1), \ldots, \pi(k) \equiv \sigma$ tal que $\pi(i+1) = \pi(i)\phi(i)$ para um (g, b)-preenchimento seguro $\phi(i)$ atuando em π_i $(0 \leq i \leq k-1)$.

Teorema 4. *Para toda permutação, existe uma permutação equivalente simples.*

Uma seqüência de permutações $\pi \equiv \pi(0), \pi(1), \ldots, \pi(k) \equiv \sigma$ é uma *ordenação generalizada* de π se σ é a identidade e $\pi(i+1)$ é obtida de $\pi(i)$ por uma reversão ou por um preenchimento.

Lema 51. *Toda ordenação generalizada de π imita uma ordenação de π com o mesmo número de reversões.*

A.1.5 Reversões seguras em componentes orientados

Uma reversão ρ é *segura* caso $\Delta(b - c + h) = -1$. Nesta seção é apresentado que sempre há uma reversão segura atuando em um ciclo de um componente orientado. No restante da Seção A.1, por ciclo entenda-se ciclo curto e por permutação entenda-se permutação simples.

Seja $V(C)$ o conjunto dos ciclos entrelaçados com o ciclo C em $G(C)$, ou seja, vizinhos a C em H_π. Seja $E(C)$ o conjunto das arestas do subgrafo de H_π induzido por $V(C)$

$$E(C) = \{(C_1, C_2) : C_1, C_2 \in V(C) \text{ e } C_1 \text{ está entrelaçado com } C_2 \text{ em } \pi\}.$$

e seu complemento

$$\overline{E(C)} = \{(C_1, C_2) : C_1, C_2 \in V(C) \text{ e } C_1 \text{ não está entrelaçado com } C_2 \text{ em } \pi\}.$$

Um reversão ρ atuando em um ciclo orientado C irá "destruí-lo", ou seja, remover suas arestas em $G(\pi\rho)$. Os outros ciclos de $G(\pi)$ são mantido em um ciclo correspondente, com os mesmos vértices, em $G(\pi\rho)$, porém suas orientações e entrelaçamentos podem ser afetados. Tem-se que $H_\pi(C_\pi, I_\pi)$ irá se transformar em $H_{\pi\rho}(C_\pi \backslash C, I_{\pi\rho})$, ou em outras palavras, o complemento do subgrafo induzido por $V(C)$. As Figuras A.5 e A.6 mostram o que ocorre no diagrama de pontos de quebra e no grafo de entrelaçamento se aplicarmos uma reversão no ciclo orientado de $G(\pi)$ apresentado na Figura A.2.

Lema 52. *Seja ρ uma reversão atuando em um ciclo orientado C. Então*

- $I_{\pi\rho} = (\overline{I_\pi} \backslash E(C)) \cup \overline{E(C)}$
- *ρ altera a orientação de um ciclo $D \in C_\pi$ se, e somente se, $D \in V(C)$.*

O conjunto $\overline{I_\pi}$ é definido como $\overline{I_\pi} = I_\pi \backslash \{(C, D) : D \in V(C)\}$.

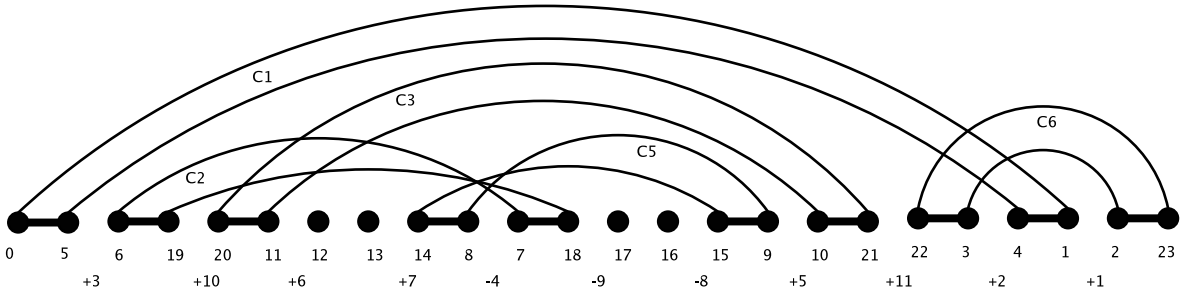


Figura A.5: Efeito de uma reversão, que atua em um ciclo orientado, sob o diagrama de pontos de quebra.

O conjunto $R(K)$ representa as reversões que atuam em ciclos orientados do componente orientado K de H_π.

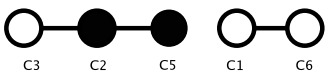


Figura A.6: Efeito de uma reversão, que atua em um ciclo orientado, sob o grafo de entrelaçamento.

Lema 53. *Todo componente não orientado do grafo de entrelaçamento (de uma permutação simples) contém pelo menos 2 vértices.*

Teorema 5. *Para todo componente orientado K em H_π existe uma reversão segura $\rho \in R(K)$ tal que todos os componentes resultantes $K_1(\rho), K_2(\rho), \ldots, K_n(\rho)$ são orientados em $H_{\pi\rho}$.*

A.1.6 Eliminando obstáculos

A Seção A.1.5 é concluída com o Teorema 5, onde é mostrado que se π tem um componente orientado, então existe uma reversão segura em π. Agora será realizada uma busca por reversões seguras na ausência de componentes orientados. Seja $\prec$ uma ordem parcial em um conjunto P e $\{x, y, z\} \subset P$. Diz-se que x está *coberto* por y caso $x \prec y$ e não existe um elemento $z \in P$ tal que $x \prec z \prec y$. O *grafo de cobertura* Ω de $\prec$ é um grafo não orientado com conjunto de vértices P e conjunto de arestas $\{(x, y) : x, y \in P$ e x está coberto por $y\}$.

Seja U_π o conjunto de componentes não orientados em H_π e seja $Estende(U)$, definido por $[U_{min}, U_{max}]$, o intervalo definido pelas posições dos elementos mais a esquerda e mais a direita de um componente não orientado $U \in U_\pi$. Define-se $\overline{U_{min}} = min_{U \in U_\pi} U_{min}$, $\overline{U_{max}} = max_{U \in U_\pi} U_{max}$ e $[\overline{U_{min}}, \overline{U_{max}}]$ o intervalo entre as posições dos elementos mais a esquerda e mais a direita levando em consideração todos os componentes não orientados em π. Define-se também um elemento (artificial) $\overline{U}$ relacionado ao invervalo $[\overline{U_{min}}, \overline{U_{max}}]$.

Seja $\overline{U_\pi}$ o conjunto de $|U_\pi|+1$ elementos consistindo dos elementos de U_π acrescido do elemento $\overline{U}$. Seja $\prec \equiv \prec_\pi$ uma order parcial em $\overline{U_\pi}$ regida por: $U \prec W$ se, e somente se, $Estende(U) \subset Estende(W)$ para $U, W \in \overline{U_\pi}$. Define-se Ω_π como a árvore representando o grafo de cobertura da ordem parcial $\prec_\pi$. Note que todo vértice, com exceção de $\overline{U}$, está relacionado a um componente não orientado U_π. Todas as folhas em Ω_π, que correspondem a um elemento mínimo em $\prec_\pi$, são obstáculos. O número de folhas em Ω_π coincide com o número de obstáculos $h(\pi)$, exceto nas seguintes situações:

- existe apenas um componente não orientado em π. Neste caso, Ω_π tem duas folhas, mas $h(\pi) = 1$
- existe um elemento máximo em $\prec_\pi$ que não é um obstáculo (separa dois obstáculos). Neste caso o número de folhas é igual a $h(\pi) + 1$.

Lema 54. *(Corte de obstáculo). Toda reversão ρ em um ciclo de um obstáculo K elimina a folha K de um grafo de cobertura de π, ou seja, $\Omega_{\pi\rho} = \Omega_\pi \backslash K$.*

Um obstáculo $K \in U_\pi$ *protege* $U \in U_\pi$, não obstáculo, se ao eliminar K de U_π transforma U em um obstáculo. Define-se *super obstáculo* como um obstáculo em π que protege um não obstáculo $U \in U_\pi$. Quando um obstáculo não é super, diz que ele é *simples*.

Lema 55. *Uma reversão atuando em um ciclo de um obstáculo simples é segura.*

Lema 56. *(Junção de obstáculos) Seja π uma permutação com grafo de cobertura Ω_π e seja ρ uma reversão atuando em arestas pretas de dois obstáculos L e M em π. Então, ρ elimina os obstáculos em $\pi\rho$ e, conseqüentemente, remove L e M de $\Omega_{\pi\rho}$.*

A operação de junção de obstáculos, definida no Lema 56, pode ser utilizada na busca por reversões seguras na ausência de componentes orientados e obstáculos simples.

Escreve-se $U < W$, para os obstáculos U e W, se $U_{max} < W_{max}$. Define-se o conjunto de obstáculos $Entre(L, M) = \{U(i) : l < i < m\}$ e $Fora(L, M) = \{U(i) : i \notin [l, m]\}$.

Lema 57. *Seja ρ uma reversão que realiza uma junção dos obstáculos L e M de π. Se ambos os conjuntos de obstáculos $Entre(L, M)$ e $Fora(L, M)$ não são vazios, então ρ é segura.*

Lema 58. *Se $h(\pi) > 3$, então existe uma reversão segura que realiza a junção de dois obstáculos de π.*

Lema 59. *Se $h(\pi) = 2$, então existe uma reversão segura que realiza a junção dos dois obstáculos em π. Se $h(\pi) = 1$, então existe uma reversão segura que corta o único obstáculo em π.*

Lema 60. *Seja ρ uma reversão em π que realiza a junção de dois obstáculos L e M. Então todo super obstáculo em π (diferente de L e M) serão mantidos como super obstáculos em $\pi\rho$.*

A.1.7 Fortalezas

Os Lemas 58 e 59 implicam que, a menos que Ω_π seja homeomorfo ao grafo 3-estrela (três arestas incidentes a um mesmo vértice), existe uma reversão segura em π. Tem-se que se existir pelo menos um obstáculo simples em π, pelo Lema 55 então existe uma reversão segura em π. Então, o único caso onde pode não existir uma reversão segura é quando Ω_π é homeomorfo do grafo 3-estrela com três super obstáculos, tal reversão π será chamada *3-fortaleza*.

Lema 61. *Se ρ é uma reversão que destrói uma 3-fortaleza π, então ρ não é segura.*

Lema 62. *Se π é uma 3-fortaleza, então $d(\pi) = b(\pi) - c(\pi) + h(\pi) + 1$.*

Uma permutação π é uma *fortaleza* se possuir um número ímpar de obstáculos e todos estes são super obstáculos.

Lema 63. *Se ρ é uma reversão que destrói uma fortaleza com $h(\pi)$-super obstáculos, então ou ρ não é segura ou $\pi\rho$ é uma fortaleza com $h(\pi) - 2$ super obstáculos.*

Lema 64. *Se π é uma fortaleza, então $d(\pi) = b(\pi) - c(\pi) + h(\pi) + 1$.*

Teorema 6. *Para toda permutação π,*

$$d(\pi) = \begin{cases} b(\pi) - c(\pi) + h(\pi) + 1, \textit{ se } \pi \textit{ é uma fortaleza} \\ b(\pi) - c(\pi) + h(\pi), \textit{ caso contrário.} \end{cases}$$

A.1.8 Algoritmo Polinomial

De acordo com os resultados apresentados nas seções anteriores segue o algoritmo *ReversalSort*.

```
    Algoritmo ReversalSort(π)
01. enquanto π não estiver ordenada
02.   se π tem um ciclo longo
03.     selecione um (g,b)-preenchimento seguro ρ de π
04.   senão se π tem um componente orientado
05.     selecione uma reversão segura ρ neste componente
06.   senão se π tem um número par de obstáculos
07.     selecione ρ segura que realize uma junção de dois obstáculos de π
08.   senão se π tem pelo menos um obstáculo simples
09.     selecione uma reversão segura ρ que corte este obstáculo de π
10.   senão se π tem uma fortaleza com mais que três super obstáculos
11.     selecione ρ segura que junte dois super obstáculos de π
12.   senão
13.     selecione ρ que junte dois super obstáculos de π
14.   π ← πρ
15. fim enquanto
16. imite a ordenação generalizada computada de π.
```

Teorema 7. *ReversalSort(π) ordena otimamente uma permutação $\pi = [\pi_1\pi_2 \dots \pi_n]$ em tempo $O(n^4)$.*

Uma análise mais cuidadosa do algoritmo leva a uma redução no tempo de execução para $O(n^3)$.

A.2 A Faster and Simpler Algorithm for Sorting Signed Permutations by Reversals

Esta seção apresenta um resumo do trabalho, cujo título é apresentado acima, desenvolvido por Kaplan, Shamir e Tarjan [13]. Em tal trabalho é apresentado um algoritmo polinomial que resolve o problema de ordenar por reversões permutações sinalizadas em tempo $O(n^2)$. Este algoritmo é mais eficiente que o primeiro algoritmo polinomial, apresentado por Hannenhalli e Pevzner [11], e a implementação otimizada, desenvolvida por Berman e Hannenhalli [7].

A.2.1 Introdução

Neste artigo é discutido o problema de ordenar permutações sinalizadas através de reversões. Uma *permutação sinalizada* é uma permutação $\pi = [\pi_1, \ldots, \pi_2]$ sobre os inteiros $\{1, \ldots, n\}$, onde a cada número está associado um sinal + ou −. Uma *reversão* $\rho(i, j)$ transforma π em

$$\begin{aligned} \pi' &= \pi\rho(i,j) \\ &= [\pi_1, \ldots, \pi_{i-1}, -\pi_j, -\pi_{j-1}, \ldots, -\pi_i, \pi_{j+1}, \ldots, \pi_n] \end{aligned}$$

A *distância de reversão* entre duas permutações, π e σ, é o número mínimo de reversões necessárias para transformar π em σ. O problema de *ordenar por reversões permutações sinalizadas* é encontrar uma seqüência mínima de reversões que transformem uma permutação π na permutação identidade $[+1, +2, \ldots, +n]$.

Hannenhalli e Pevzner [11] mostraram que ordenar por reversões permutações orientadas é um problema polinomial. Eles provaram um teorema que relaciona a distância de reversão a soma de três parâmetros e baseados neste teorema, e em alguns outros, apresentam um algoritmo que executa em tempo $O(n^4)$. Posteriormente, Berman e Hannenhalli [7], apresentaram uma versão otimizada do primeiro algoritmo, que executa em tempo $O(n^2\alpha(n))$, onde α é o inverso da função de Ackermann [1].

Este trabalho apresenta simplificações e otimizações nos trabalhos de Hannenhalli e Pevzner [11] e Berman e Hannenhalli [7] que levam a um algoritmo mais eficiente. Tal algoritmo executa em tempo $O(rn + n\alpha(n))$, dada uma distância de reversão r. A seguir são listadas as simplificações no algoritmo e nas estruturas alcançadas com este trabalho:

- O objeto básico utilizado no algoritmo é uma representação implícita do grafo de entrelaçamento de arcos. Este grafo, diferentemente do grafo de Hannenhalli e Pevzner [11], tem arcos como vértices e, desta forma, é mais simples de ser produzido e mantido durante a busca pelo resultado.

- Elimina a necessidade de trabalhar com permutações simples, deixando de lado as transformações de equivalência definidas por Hannenhalli e Pevzner [11].
- A noção de obstáculo é simplificada.
- A busca pelas reversões é simplificada e não requer estruturas de dados especiais.

A.3 A Very Elementary Presentation of the Hannenhalli-Pevzner Theory

Esta seção apresenta um resumo do trabalho, cujo título é apresentado acima, desenvolvido por Anne Bergeron [5]. Em tal trabalho são apresentadas simplificações no trabalho de Hannenhalli e Pevzner [11] que levam a um algoritmo que executa em tempo $O(n^2)$ ao invés de $O(n^4)$.

Convenciona-se que toda as permutações estão emolduradas por $\pi_0 = 0$ e $\pi_{n+1} = n+1$, ou seja, $\pi = [0, \pi_1, \pi_2, \ldots, \pi_n, n+1]$. Um *par orientado* (π_i, π_j) é um par de inteiros consecutivos ($|\pi_i| - |\pi_j| = \pm 1$) com sinais opostos. Por exemplo, os pares orientados da permutação:

$$\pi = [0, 5, -1, 2, 4, 6, -3, 7]$$

são $(0, -1)$, $(-1, 2)$, $(2, -3)$ e $(-3, 4)$.

Pares orientados induzem reversões que tornam elementos consecutivos, em outras palavras, inteiros consecutivos. Por exemplo, o par $(2, -3)$ induz a reversão $\rho(4, -3)$:

$$\rho\pi = [0, 5, -1, 2, 3, -6, -4, 7]$$

criando os inteiros consecutivos 2 3. A reversão induzida por um par orientado (π_i, π_j) será:

$$\rho(i, j-1), \text{ se } \pi_i + \pi_j = +1, \text{ e}$$
$$\rho(i+1, j), \text{ se } \pi_i + \pi_j = -1$$

As reversões induzidas por pares orientados são chamadas de reversões orientadas. Define-se a pontuação de uma reversão (orientada) como o número de pares orientados na permutação resultante. Por exemplo, para $\rho(4, -3)$ temos pontuação 6.

O primeiro algoritmo definido serve para os casos onde há pelo menos um par orientado. Algoritmo: se houver pares orientados, escolha a reversão orientada de pontuação máxima.

Bergeron mostra que tal reversão com pontuação máxima é segura. Quando as permutações possuem apenas elementos positivos e são reduzidas (não contém elementos consecutivos) o primeiro algoritmo não funciona. Em tal situação é necessário um algoritmo mais elaborado, que será mostrado a seguir.

Assume-se que a permutação é circular, ou seja, o sucessor de $n+1$ é 0. Define-se intervalo emoldurado em π como um intervalo da forma:

$$i, \pi_{j+1}, \pi_{j+2}, \ldots, \pi_{j+k-1}, i+k$$

tal que todos inteiros entre i e $i+k$ pertencem ao intervalo $i \ldots i+k$. Por exemplo, considere a permutação

$$[0, 6, 1, 3, 2, 4, 5, 7].$$

A permutação inteira é um intervalo emoldurado. Mas há também $[1, 4]$ e $[4, 1]$ devido a circularidade.

Se π é reduzida, um obstáculo em π é um intervalo emoldurado que não contém outro intervalo emoldurado menor. Tal definição é equivalente a de Hannenhalli e Pevzner [11]. Uma vez estabelecido o conceito de obstáculos, são definidas duas operações sobre obstáculos (corte e junção). Um corte de obstáculo consiste em reverter o segmento entre i e $i+1$ de um obstáculo. Uma junção de obstáculos consiste na reversão que atua nos terminais de dois obstáculos. Um obstáculo simples é um obstáculo que quando cortado decrementa o número de obstáculos. Um obtstáculo que não é simples é um super obstáculo.

Introduzidos os conceitos acerca de obstáculos é definido um segundo algoritmo, este é capaz de resolver o problema de ordenar por reversões orientadas para uma permutação que não possui pares orientados. O algoritmo é definido como segue: se uma permutação tem $2k$ obstáculos, $k > 2$, junte qualquer dois obstáculos não consecutivos. Se uma permutação tem $2k+1$, $k \geq 1$, então se tem um obstáculo simples, corte; Se não tem, junte dois obstáculos não consecutivos, ou consecutivos se $k = 1$.

Bergeron mostra que tal estratégia sempre encontra reversões seguras. Os dois algoritmos juntos resolvem otimamente qualquer permutação com sinais.

Dado um genoma π 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 π' substituindo-se cada elemento positivo x em π por $2x-1$ $2x$ em π' e cada elemento negativo $-x$ em π por $2x$ $2x-1$ em π'. O elemento 0 de π matém-se em π' e o elemento $n+1$ em π é substituído por $2n+1$ em π'.

2. Toma-se os elementos da permutação π' como vértices, totalizando $2n+2$ vértices.

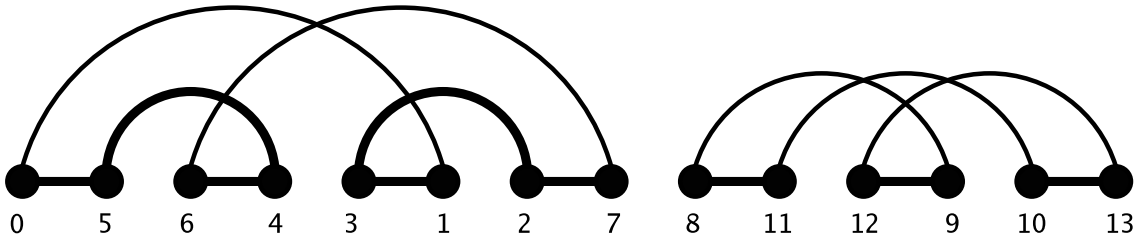


Figura A.7: Diagrama de pontos de quebra de π em relação a identidade.

3. As arestas são definidas e classificadas da seguinte forma:

   Arestas retas: elementos consecutivos [1] de π' par sim, par não, começando do 0.

   Arestas curvas (ou arcos): inteiros consecutivos par sim, par não, começando de 0.

Por exemplo,

$$\pi = [0\ 3\ -2\ 1\ 4\ 6\ 5\ 7]$$

torna-se:

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

A Figura A.7 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 (π) e para definir as arestas curvas precisamos olhar para o genoma destino (σ). 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 π, de tamanho n, em relação a um genoma σ 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 (σ) a identidade.

O suporte de um arco é dado pelo intervalo de elementos de π' entre, e incluindo, seus terminais. Por exemplo, no diagrama da Figura A.7, 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

[1] Dado o elemento $x = \pi'(i)$, seu consecutivo é $y = \pi'(i + 1)$.

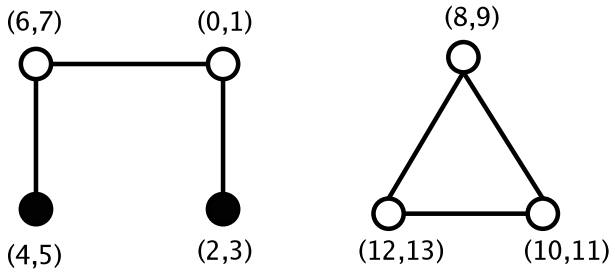


Figura A.8: Grafo de entrelaçamento de arcos.

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 A.7 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 (π). 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 se entrelaçam se a intersecção dos suportes não é vazia e um deles não está contido propriamente no outro. O grafo de entrelaçamento 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 entrelaçados.

A Figura A.8 apresentado o grafo de entrelaçamento de arcos correspondente ao diagrama de pontos de quebra da Figura A.7.

Um componente conexo do grafo de entrelaçamento 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 entrelaçamento 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 entrelaçados.

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

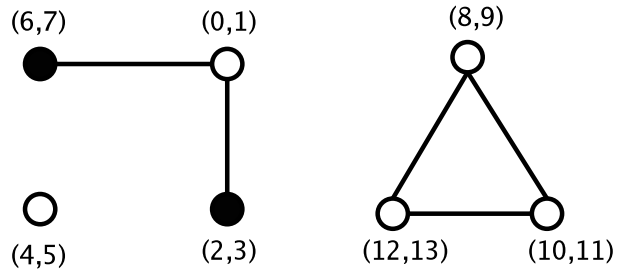


Figura A.9: Grafo de entrelaçamento de arcos resultante da aplicação da reversão induzida pelo vértice $(4, 5)$.

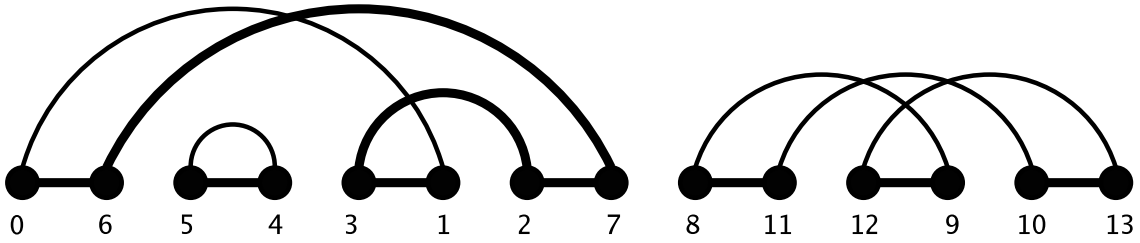


Figura A.10: Diagrama de pontos de quebra resultante da aplicação da reversão induzida pelo vértice $(4, 5)$.

Ao aplicar uma reversão correspondente a um vértice orientado v, o efeito no grafo de entrelaçamento de arcos será o complemento do subgrafo de v e seus vértices adjacentes. A Figura A.9 apresenta o grafo de entrelaçamento 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 A.10 podemos visualizar o que ocorre no diagrama de pontos de quebra. Uma reversão $\rho(i, j)$ em π é simulada por uma reversão $\rho(2i - 1, 2j)$ em π'. Tal reversão isola o arco que induz a reversão e inverte a ordem dos elementos que estão entre os terminais do arco em π'. Como efeito a reversão ainda altera a orientação do 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.

A Equação A.3 mostra como pontuar uma reversão correspondente a um vértice orientado v através do grafo de entrelaçamento de arcos.

$$T + U - O - 1 \tag{A.3}$$

Na equação, T é o número total de vértices 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

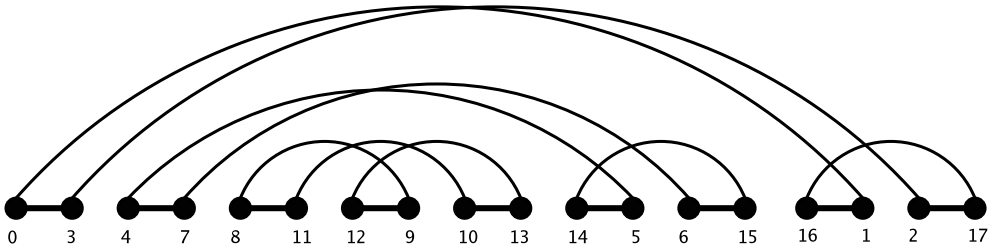


Figura A.11: Diagrama de pontos de quebra de π em relação a identidade.

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 A.8, 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 componentes 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 entrelaçamento de arcos é o intervalo mínimo que contém, na ordem circular, todos os intervalos de vértices em X. Por exemplo, considere o genoma π = [0 2 4 6 5 7 3 8 1 9]. A Figura A.11 apresenta o diagrama de pontos de quebra de π e a Figura A.12 apresenta o grafo de entrelaçamento de arcos. Os componentes conexos de seu grafo de entreçamento de arcos têm coberturas:

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

Segundo Hannenhalli e Pevzner [11], obstáculos são componentes não orientados que são minimais no que diz respeito a inclusão de cobertura ou, em outras palavras, deve ser não orientado e sua cobertura deve conter apenas os vértices do componente. São exemplos de obstáculos os componentes de cobertura $[8, 13]$ e $[16, 3]$.

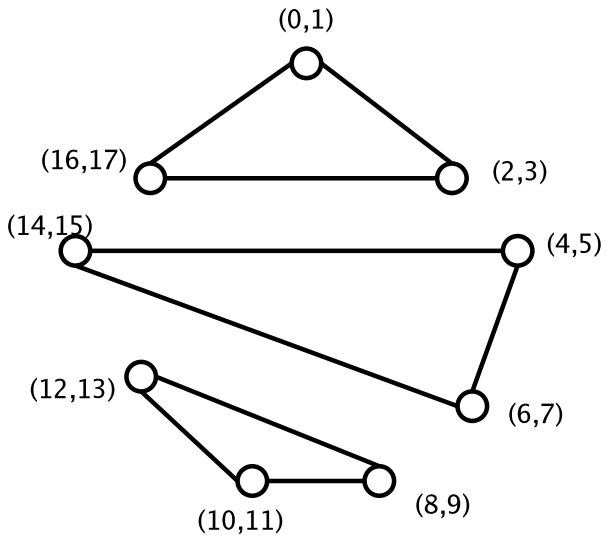


Figura A.12: Grafo de entrelaçamento de arcos.

Bibliografia

[1] W. Ackermann. Zum hilbertschen aufbau der reellen zahlen. *Mathematische Annalen*, 99:118–133, 1928.

[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. Sorting by reversals: Genome rearrangements in plant organelles and evolutionary history of X chromosome. *Molecular Biology and Evolution*, 12:239–246, 1995.

[4] V. Bafna and P. A. Pevzner. Genome rearrangements and sorting by reversals. *Society for Industrial and Applied Mathematics, Jornal on Computing*, 25:272–289, 1996.

[5] A. Bergeron. A very elementary presentation of the Hannenhalli-Pevzner theory. *Discrete Applied Mathematics*, 146(2):134–145, 2005.

[6] A. Bergeron, J. Mixtacki, and J. Stoye. Reversal distance without hurdles and fortresses. In *Combinatorial Pattern Matching*, volume 3109 of *Lecture Notes in Computer Science*, pages 388–399. Springer Berlin / Heidelberg, 2004.

[7] P. Berman and S. Hannenhalli. Fast sorting by reversals. In *Combinatorial Pattern Matching, Proceedings of the 6th Annual Symposium (CPM'96)*, Lecture Notes in Computer Science, pages 168–185, Berlin, Germany, 1996. Springer-Verlag.

[8] T. A. Brown. *Genomes*. Wiley-Liss, 1999.

[9] A. Caprara. Sorting permutations by reversals and eulerian cycle decompositions. *SIAM Journal on Discrete Mathematics*, 12(1):91–110, February 1999.

[10] Z. Dias. *Rearranjo de Genomas: Uma Coletânea de Artigos*. PhD thesis, Instituto de Computação - Universidade Estadual de Campinas, 2002.

[11] S. Hannenhalli and P. A. Pevzner. Transforming cabbage into turnip: polynomial algorithm for sorting signed permutations by reversals. *Journal of the Association for Computing Machinery (ACM)*, 46(1):1–27, 1 1999. Preliminary version appeared in Proceedings of the Twenty-Seventh Annual ACM Symposium on the Theory of Computing 1995 (178–189).

[12] N. Jacobson. *Basic Algebra*. W. H. Freeman and Company, New York, 1985.

[13] H. Kaplan, R. Shamir, and R. Tarjan. A faster and simpler algorithm for sorting signed permutations by reversals. *Society for Industrial and Applied Mathematics, Jornal on Computing*, 29(3):880–892, 1999.

[14] J. Kececioglu and D. Sankoff. Efficient bounds for oriented chromosome inversion distance. In Springer-Verlag, editor, *Combinatorial Pattern Matching, Proceedings of the 5th Annual Symposium (CPM'94)*, volume 807 of *Lecture Notes In Computer Science*, pages 307–325, Berlin, Germany, 1994.

[15] J. Kececioglu and D. Sankoff. Exact and approximation algorithms for the inversion distance between two permutations. *Algorithmica*, 13:180–210, 1995.

[16] S. MacLane and G. Birkhoff. *Algebra*. The Macmillan Company, London, sixth edition, 1971.

[17] J. Meidanis and Z. Dias. An alternative algebraic formalism for genome rearrangements. In D. Sankoff and J. H. Nadeau, editors, *Proceedings of the Gene Order Dynamics, Comparative Maps and Multigene Families (DCAF'2000)*, Le Chantecler, Canada, Sep 2000.

[18] 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, 1997.

[19] L. H. J. Monteiro. *Elementos de Álgebra*. Livros Técnicos e Científicos Editora, Rio de Janeiro, Brasil, 1974.

[20] J. H. Nadeau and B. A. Taylor. Lengths of chromosomal segments conserved since divergence of man and mouse. In *Proceedings of the National Academy of Sciences*, volume 81, pages 814–818, USA, 1984.

[21] The National Center for Biotechnology Information, August 2006. http://www.ncbi.nlm.nih.gov.

[22] 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.

[23] D. Sankoff. Edit distance for genome comparison based on non-local operations. In *Combinatorial Pattern Matching, Proceedings of the 3rd Annual Symposium (CPM'92). Lecture Notes In Computer Science*, volume 644, pages 121–135, Berlin, Germany, 1992. Springer-Verlag.

[24] D. Sankoff, R. Cedergren, and Y. Abel. *Genomic divergence through gene rearregement. In Molecular Evolution: Computer Analysis of Protein and Nucleic Acid Sequences*, chapter 26, pages 428–438. Academic Press, 1990.

[25] D. Sankoff, G. Leduc, N. Antoine, B. Paquin, B. F. Lang, and R. Cedergren. Gene order comparisons for philogenetic inference: Evolution of the mitochondrial genome. In *Proceedings of the National Academy of Sciences*, pages 9575–9579, USA, 1992.

[26] J. C. Setubal and J. Meidanis. *Introduction to computational molecular biology.* PWS Publishing Company, 1997.

[27] T. F. Smith and M. S. Waterman. Identification of common molecular subsequences. *J. Mol. Biol.*, 147:195–197, 1981.

[28] A. H. Sturtevant and T. Dobzhansky. Inversions in the third chromosome of wild races of drosophila pseudoobscura, and their use in the study of the history of the species. In *Proceedings of the National Academy of Sciences*, volume 22, pages 448–450, 1936.

[29] E. Tannier and Marie-France Sagot. Sorting by reversals in subquadratic time. In *Combinatorial Pattern Matching*, volume 3109 of *Lecture Notes in Computer Science*, pages 1–13. Springer Berlin / Heidelberg, 2004.

[30] M. E. M. T. Walter. *Algoritmos para problemas em rearranjo de genomas.* PhD thesis, University of Campinas, Brazil, 1999. In Portuguese.

[31] G. A. Watterson, W. J. Ewens, T. E. Hall, and A. Morgan. The chromosome inversion problem. *Journal of Theoretical Biology*, 99:1–7, 1982.

[32] D. B. West. *Introduction to graph theory.* Prentice Hall, second edition, 2001.

[33] S. Yancopoulos, O. Attie, and R. Friedberg. Efficient sorting of genomic permutations by translocation, inversion and block interchange. *Bioinformatics*, 21(16):3340–3346, 2005.