

Improvements in the score matrix calculation in multiple sequence alignments

Geraldo F. D. Zafalon
Evandro A. Marucci
José R. A. Amazonas
Liria M. Sato
José M. Machado

Sumário

- 1 Introdução
- 2 Definições
- 3 Problemática
- 4 Algoritmos e análises
- 5 Resultados
- 6 Conclusões

- 1 Introdução
- 2 Definições
- 3 Problemática
- 4 Algoritmos e análises
- 5 Resultados
- 6 Conclusões

- Grande necessidade do uso de técnicas computacionais para resolução de problemas biológicos cada vez mais complexos;
- Busca por regiões que possam identificar pontos de cura de doenças ou bloqueio de doenças;
- Grandes perspectivas de aplicações em outras áreas.

- 1 Introdução
- 2 Definições**
- 3 Problemática
- 4 Algoritmos e análises
- 5 Resultados
- 6 Conclusões

Alinhamentos de Sequências

- Técnicas que objetivam tratar dessas sequências de caracteres;
- O tratamento dado por esses alinhamentos visa comparar as sequências e buscar entre elas um rearranjo, tentando obter um número máximo de coincidências entre esses caracteres;
- Inferências podem ser realizadas sobre as partes similares encontradas;

Alinhamentos Múltiplos

- Um dos algoritmos de alinhamentos múltiplos mais utilizado é o Progressivo [1, 2];

Dividido em três fases:

- Alinhamento par-a-par;
- Construção da árvore guia;
- Alinhamento múltiplo;

Alinhamentos Múltiplos

- Execução sequencial do algoritmo se torna inviável com o aumento do número de sequências;
- Utilização de técnicas de programação paralela;
- Fase de alinhamento par-a-par é a mais crítica e a mais custosa computacionalmente.

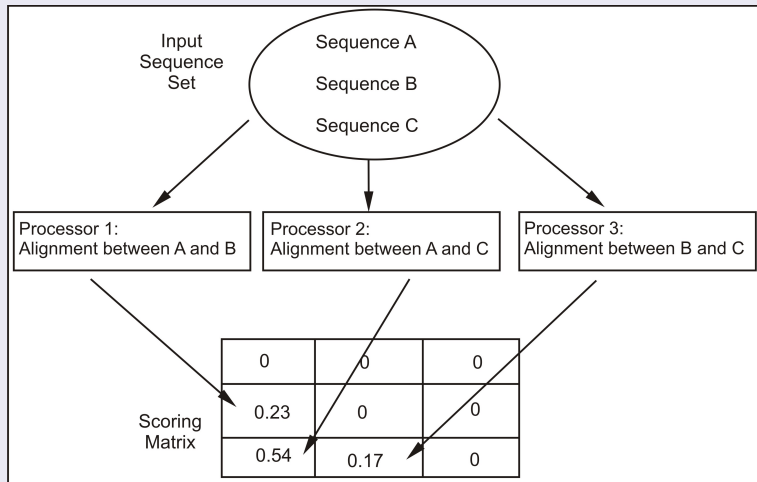
Sumário

- 1 Introdução
- 2 Definições
- 3 Problemática
- 4 Algoritmos e análises**
- 5 Resultados
- 6 Conclusões

Proposta do trabalho

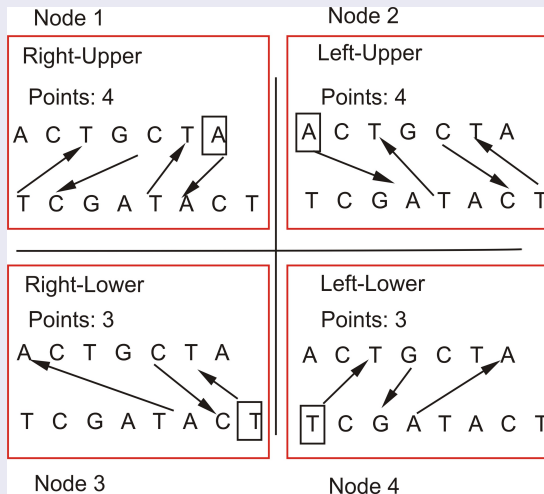
- Otimizar a execução da fase de alinhamento par-a-par;
- Utilização de um algoritmo progressivo com programação dinâmica, adequando o seu paralelismo para grão-grosso;
- Implementação do método de estimativa de pontuação [3] paralelizado, para a fase de alinhamento par-a-par;
- Avaliação do desempenho das duas estratégias, objetivando a redução da complexidade computacional.

Programação Dinâmica Paralela



Algoritmo com Estimativa de Pontuação

Estimativa de Pontuação Paralela

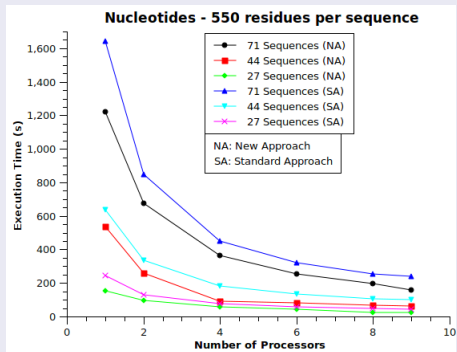


Sumário

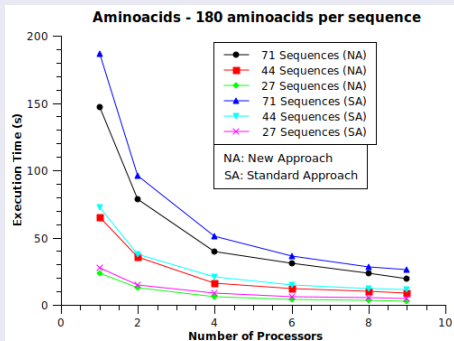
- 1 Introdução
- 2 Definições
- 3 Problemática
- 4 Algoritmos e análises
- 5 Resultados**
- 6 Conclusões

- Ambas as abordagens, foram executadas sob o mesmo conjunto de testes;
- As demais fases do alinhamento múltiplo, que são a construção da árvore filogenética e o alinhamento múltiplo, foram providas pelas mesmas estratégias para ambas;
- Testes executados em cluster Beowulf de Athlon XP 2100+, com 1GB de memória por nó, com 9 nós disponíveis.

Nucleotídeos



Aminoácidos



Estimativas de Pontuação (EP) X Programação Dinâmica (PD)




| Genes | Num. Seq. | EP | PD |
|--------------|------------------|-----------|-----------|
| 1ad3_ref1 | 4 | 0.929 | 0.899 |
| 1cpt_ref1 | 9 | 0.787 | 0.845 |
| 1fmb_ref4 | 9 | 0.882 | 0.852 |
| 1tis_ref5 | 18 | 0.898 | 0.867 |
| 1ppn_ref5 | 26 | 0.906 | 0.893 |
| 1gdoA_ref5 | 42 | 0.877 | 0.823 |
| 1ubi_ref3 | 48 | 0.797 | 0.786 |
| 1thm_ref5 | 49 | 0.794 | 0.775 |
| kinase_ref4 | 62 | 0.802 | 0.804 |

- 1 Introdução
- 2 Definições
- 3 Problemática
- 4 Algoritmos e análises
- 5 Resultados
- 6 Conclusões**

- Os resultados apresentados mostram um ganho de desempenho da estratégia de estimativas de pontuação, em relação à programação de dinâmica;
- O ganho de desempenho foi da ordem de 15% no teste com nucleotídeos e 12% no teste com amino-ácidos.
- Trata-se de um ganho significativo.

- Nos testes de qualidade, abordagem de estimativas de pontuação também apresentou melhores resultados;
- Em quase todos os testes, a sua pontuação foi superior ao da programação dinâmica;
- Na média, apresentando melhores resultados.

Referências Bibliográficas

-  R. C. Edgar and S. Batzoglou, "Multiple sequence alignment," *Current Opinion in Structural Biology*, no. 16, pp. 368–373, 2006.
-  I. M. Wallace, G. Blackshields, and D. G. Higgins, "Multiple sequence alignment," *Current Opinion in Structural Biology*, no. 15, pp. 261–266, 2005.
-  Y. Chen, Y. Pan, J. Chen, W. Liu, and L. Chen, "Partitioned optimization algorithms for multiple sequence alignment," in *Proceedings of the 20th International Conference on Advanced Information Networking and Applications (AINA '06)*, 2006.

OBRIGADO PELA ATENÇÃO!