

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Uma abordagem do algoritmo do montador de DNA SOAPdenovo2

Evaldo B. Costa¹

Programa de Pós-Graduação em Informática, PPGI, UFRJ, Rio de Janeiro, RJ

Gabriel P. Silva, Marcello G. Teixeira²

Departamento de Ciência da Computação, DCC, UFRJ, Rio de Janeiro, RJ

Resumo. Em bioinformática, existem vários programas disponíveis para montagem de sequência de DNA [1]. Isso geralmente é uma tarefa muito demorada, uma vez que essas sequências de DNA podem ser muito longas e complexas [2]. A escolha correta de um montador implica diretamente na quantidade de recursos computacionais utilizados para este processo. Neste trabalho, faremos uma abordagem do novo algoritmo utilizado no montador SOAPdenovo2.

Palavras-chave. Montador, DNA, SOAPdenovo1, SOAPdenovo2, Algoritmo, desempenho

1 Introdução

O montador SOAPdenovo2 [3] é um programa de montagem NGS (Sequenciamento de Nova Geração) com suporte a computação paralela de multithreaded. Os montadores NGS utilizam o método de grafo de De Bruijn.

A versão SOAPdenovo1 conseguia montar muitos genomas, porém existiam alguns problemas em relação a continuidade, precisão e cobertura. Uma das principais diferenças entre essas duas versões está na maneira pela qual os gráficos de Bruijn são construídos.

2 Ambiente de Testes

Os testes foram realizados em um servidor com dois processadores Intel Xeon E5-2650 (2,00 GHz, 8 núcleos cada, 20 MB de cache), com 64 GB de memória. Ambos os montadores foram compilados usando o Intel C++ Compiler Suite XE 13.1 Update 2 para prover um melhor desempenho.

Para a execução dos testes foi utilizado o genoma do Cromossomo Humano 14. Os dados foram armazenados em um disco SSD (unidade de estado sólido) local de alta velocidade. O sistema operacional utilizado foi a distribuição Linux CentOS de 64 bits versão 6.3.

¹evaldo.costa@ppgi.ufrj.br

²gabriel@dcc.ufrj.br, marcellogt@dcc.ufrj.br

2

3 Resultados

Para os resultados apresentados foram realizadas três séries de testes e calculado o tempo médio de execução dos montadores SOAPdenovo1 e SOAPdenovo2.

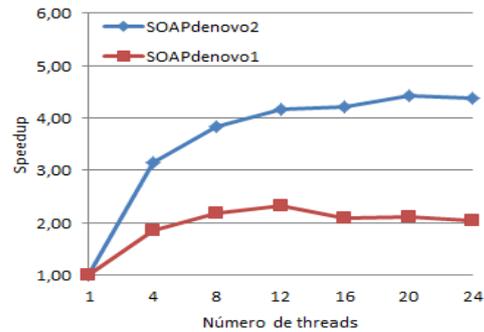


Figura 1: Speedup dos montadores.

Na Figura 1 é visto o desempenho dos montadores. O SOAPdenovo2 tem um ganho de desempenho 2 vezes maior em relação ao SOAPdenovo1. Em relação ao uso de memória, SOAPdenovo2 utilizou aproximadamente 60% a menos de memória.

4 Conclusão

Como visto nos testes a nova versão do algoritmo utilizada no SOAPdenovo2 tem uma melhor utilização dos recursos computacionais. A leitura dos dados de entrada e a utilização de memória são as principais melhorias no algoritmo.

Agradecimentos

Os autores agradecem à Microway Incorporated por fornecer os recursos computacionais utilizados para a realização dos experimentos apresentados neste trabalho.

Referências

- [1] E. B. Costa, G. P. Silva e M. G. Teixeira, Avaliação de Desempenho Paralelo de Montadores de DNA Velvet e SOAPdenovo2 XXXV Congresso Nacional de Matemática Aplicada e Computacional, Natal, XXXV CNMAC, Vol. 1, (2014).
- [2] R. Li, H. Zhu, J. Ruan, W. Qian, X. Fang and J. Wang, De novo assembly of human genomes with massively parallel short read sequencing, *Genome Res*, (2010).
- [3] R. Luo, B. Liu, Y. Xie, W. Huang, G. He, J. Tang and Y. Shi, Software and supporting material for SOAPdenovo2: an empirically improved memory efficient short read de novo assembly, *GigaScience*, (2012).