



PROGRAMAÇÃO CONCORRENTE E PARALELA EDIÇÃO DE GENES

Debora Rohden¹, Eduarda de Lima dos Anjos¹

¹Instituto Federal Catarinense - Campus Rio do Sul

deboraroh@gmail.com

delimadosanjoseduarda@gmail.com

Abstract. *In many software, it is very common that certain tasks have a low time performance due to the amount of parts that are necessary to execute. In this work, a comparison between execution in serial and with concurrent and parallel threads will be presented to solve the problem of Gene Editing. According to tests performed, the use of threads to solve the problem is satisfactory compared to serial execution.*

Key-words: *Edition; Genes; Threads.*

Resumo. *Em muitos softwares, é muito comum que determinadas tarefas tenham uma baixa performance de tempo devido a quantidade de partes que são necessárias executar. Neste trabalho será apresentado uma comparação entre execução em serial e com threads concorrentes e paralelas para resolução do problema de Edição de Genes. Conforme testes realizados, a utilização de threads para a resolução do problema é satisfatória comparado a execução em serial.*

Palavras-chave: *Edição; Genes; Threads.*

1. Introdução

Em muitos softwares, é muito comum que determinadas tarefas tenham uma baixa performance de tempo devido a quantidade de partes que são necessárias executar. Devido a isso, muitos programadores optam por processar essas tarefas a partir da criação de threads concorrentes, onde cada uma, é associada a uma determinada atividade, onde as mesmas acontecem ao mesmo tempo, assim, aumentando a performance do programa.

Neste trabalho será apresentado uma breve fundamentação teórica a respeito de Threads e concorrência, além da problemática, por fim, será descrito o desenvolvimento e a comparação entre o uso de threads para auxiliar na resolução do problema e a resolução do problema em serial.

2. Thread

Thread é uma parte do código que pode ser executada independentemente do programa principal (KITAMURA, 2021). É uma forma de um processo se dividir em duas ou mais tarefas que podem ser executadas concorrentialmente (WIKIPEDIA, 20).

2.1 Concorrência

Concorrência é relatado como algo que lida com várias coisas ao mesmo tempo. Ou seja, concorrência refere-se à várias threads em execução simultânea, podendo estarem rodando em núcleos distintos ou intercalando as execuções em um único núcleo (BECK, 2020).

2.2 Paralelismo

Paralelismo é a execução paralela de tarefas, ou seja, execução de mais de uma tarefa por vez, de forma simultânea, sempre dependendo da quantidade de núcleos presentes no processador (TEDESCO, 2020).

3. Edição de Genes

A Engenharia Genética é um conjunto de técnicas que manipulam o DNA por meio da sua recombinação, com o objetivo de fabricar organismos melhorados (CARREIRA, 2020).

A edição de genes começou a ser desenvolvido desde 1990, permitindo a eliminação e substituição de trechos específicos de DNA (INFOENEM, 2020). Uma das técnicas mais utilizadas é o alinhamento de sequencias, que consiste em uma forma de organizar estruturas primarias de DNA, RNA ou proteína para identificar regiões similares (WIKIPEDIA, 20?), muitas vezes é possível encontrar gaps(-) entre alguns caracteres para que seja alinhado com a sequência que será comparada. Ela é amplamente utilizada para identificação de genes, identificação de mutações, onde é feita a comparação de sequencias já conhecidas com as novas sequencias.

AAB24882	TYHMCQFHCERYVNNHSGEKLYECNERSKAFSCPSHLQCHKRRQIGETHEHNQCGKAFPT	60
AAB24881	-----YECNQCGKAFQHSLLKCHYRTHIGEPYECNQCGKAFSK	40
	: .: * *:*** * :***. : * ****. .	
AAB24882	PSHLQYHERTHTGKPYECHQCGQAFKKCSLLQRHKRTHTGKPYE-CNQCGKAFQ-	116
AAB24881	HSHLQCHKRTHTGKPYECNQCGKAFSQHGLLQRHKRTHTGKPYMNVINMVKPLHNS	98
	*** *:*****:***:***: . : *****: : * . : :	

Alinhamento de Sequencia
Fonte: Wikipedia

4. Extensão FASTA

O arquivo .FASTA é um formato de arquivo utilizado amplamente entre a bioinformática. Este arquivo é utilizado para descrever sequencias de ácidos, DNA e proteínas nucleicos, permitindo sequencias de nomes e comentários precedendo as sequencias(WIKIPEDIA, 2011).

Seu formato consiste inicialmente com uma descrição em uma única linha, seguidamente de linhas com as sequencias, sendo que a linha de descrição se defere pelo símbolo “>” (maior-que), e as demais, serão as sequencias, podendo conter mais de um “>”, indicando o início de uma nova sequência.

```
>gi|5524211|gb|AAD44166.1| cytochrome b [Elephas maximus maximus]
LCLYTHIGRNIYYGSYLYSETWNTGIMLLLITMATAFMGYVLPWQMSFWGATVITNLFSAIPYIGTNLV
EWIWGGSVDKATLNRFFAFHFILPFTMVALAGVHLTFLHETGSNNPLGLTSDSDKIPFHPYYTIKDFLG
LLILILLILLALLSPDMLGDPDNHMPADPLNTPLHIKPEWYFLFAYAILRSVPNKLGGVLALFLSIVIL
GLMPFLHTSKHRSMMLRPLSQALFWTLTMDLLTLTWIGSQPVEYPYTIIGQMASILYFSIILAFPLIAGX
IENY
```

Arquivo .FASTA
Fonte: Wikipedia

5. Desenvolvimento

Para realizar as comparações de sequencias, são retirados os gaps(-) iniciais e finais das sequencias de uma determinada entrada. Na segunda entrada, é retirado no início e no final da sequência a quantidade de gaps que estava presente na primeira entrada, fazendo com que as sequencias sejam iguais e possam ser comparadas. Após a edição das duas entradas, é realizada a geração de um novo arquivo para inserir as saídas correspondentes as entradas.

Para fins de comparação, todos os testes foram executados na mesma máquina para utilizar a mesma configuração, sendo elas:

- Intel i5-4460 CPU @ 3.20GHz
- 8 GB de RAM

5.1 Comparação Serial x Concorrente

Para realizar as comparações de performance entre serial e concorrente, foram desenvolvidas duas maneiras de processar a edição de genes, a partir da linguagem Java:

1. Fazendo a utilização de threads concorrentes para realizar a edição de vários arquivos ao mesmo tempo;
2. Fazendo execução em serial, onde deverá aguardar uma edição finalizar para iniciar outra.
3. Fazendo a criação de duas Threads, cada uma responsável pela edição de uma parte do arquivo.

Utilizando o quadro 1, é possível verificar que com a quantidade de 1.000 execuções o tempo permanece muito próximo, a medida de incremento das execuções, a diferença do tempo entre execução serial e execução em thread torna-se mais considerável.

Execuções	Serial	Thread Concorrente 1
1000	2 segundos	1 segundo
5000	12.431 segundos	5.913 segundos
10000	22.720 segundos	9.946 segundos

Quadro 1 – Tempo de execução
Fonte: Autor

No quadro 2, temos a comparação entre a execução em serial do arquivo, onde somente é executado uma vez, com a utilização de duas threads para edição do arquivo. Como é possível verificar, a utilizar de threads para somente uma execução é mais demorada que uma execução em serial.

Execução	Serial	Thread Concorrente 2
1	13 milissegundos	23 milissegundos

Quadro 2 – Tempo de execução 2
Fonte: Autor

5.2 Comparação Serial x Paralelo

Para realizar as comparações entre serial e execução em paralelo, foi desenvolvido na linguagem C++ utilizando a biblioteca OpenMP, a qual é amplamente utilizada para execuções de Threads paralelas. As duas maneiras de execução foram:

1. Fazendo a edição de um arquivo em serial, onde é necessário esperar que uma função acabe para executar outra;
2. Fazendo a edição de um arquivo com duas Threads paralelas, onde cada uma é responsável pela edição de uma parte do arquivo.

Utilizando o quadro 3, temos a comparação de tempos entre a execução em serial e a execução em paralela, podendo perceber que a execução em paralela se obteve um melhor resultado de execução, editando o arquivo 1 segundo a menos que na execução em serial.

Execução	Serial	Thread Paralela
1	2.299 segundos	1.329 segundos

Quadro 3 – Tempo de execução Serial x Paralela
Fonte: Autor

6. Conclusão

Conforme os testes realizados com diferentes quantidades de execuções, nota-se que para o caso descrito, a utilização threads para resolve-lo é mais eficaz comparado a execução em serial, mas não pode-se dizer que a utilização de threads é sempre a opção mais inteligente para situações gerais.

7. Referências

BECK, FILIPE. **Assincronismo, concorrência e paralelismo**. Disponível em: <<https://dev.to/filipebeck/assincronismo-concorrencia-e-paralelismo-19hk>>. Acesso em: 26 nov. 2021.

CARREIRA. **Engenharia Genética**: tudo sobre a manipulação dos genes. Disponível em: <<https://www.stoodi.com.br/blog/carreira/engenharia-genetica/>>. Acesso em: 26 nov. 2021.

INFOENEM. **Engenharia genética:** A edição de genes. Disponível em: <<https://infoenem.com.br/engenharia-genetica-a-edicao-de-genes/>>. Acesso em: 26 nov. 2021.

KITAMURA, C. **O que é Thread.** Disponível em: <<https://celsokitamura.com.br/o-que-e-thread/>>. Acesso em: 25 nov. 2021

TEDESCO, K. **Concorrência, Paralelismo, Processos, Threads, programação síncrona e assíncrona.** Disponível em: <<https://www.treinaweb.com.br/blog/concorrencia-paralelismo-processos-threads-programacao-sincrona-e-assincrona>> . Acesso em: 8 dez. 2021.

WIKIMEDIA. **Alinhamento de sequências.** Disponível em: <https://pt.wikipedia.org/wiki/Alinhamento_de_sequ%C3%Aancias> . Acesso em: 26 nov. 2021.

WIKIMEDIA. **Formato FASTA.** Disponível em: <https://pt.wikipedia.org/wiki/Formato_FASTA> . Acesso em: 6 dez. 2021.