



Universidade Federal de Pernambuco  
Centro de Informática  
Graduação em Engenharia da Computação

**Representação eficiente em memória de Grafos de de Bruijn para uso  
com *Data Streams***

**Aluno:** Augusto Sales de Queiroz (asq@cin.ufpe.br)

**Orientador:** Paulo Gustavo Soares da Fonseca (paguso@cin.ufpe.br)

**Área:** Algoritmos e Estruturas de Dados

Recife,  
22 de Fevereiro de 2022

# 1 Resumo

O uso de memória causa um dos maiores *bottlenecks* nos *softwares* de reconstrução de genomas (*assemblers*). Um dos principais alvos de otimização nesse tipo de programa é a representação de um Grafo de de Bruijn (GdB), uma vez que essa estrutura é amplamente utilizada por esses *assemblers*, apesar de requerer um alto custo em memória (na ordem de GBs). Este trabalho se propõe a analisar soluções atuais para esse problema, e propor um novo modelo de representação de GdBs eficiente em memória. Além disso, será dado um foco em estruturas de dados baseadas em *sketches*, permitindo que as leituras não precisem estar todas disponíveis simultaneamente, reduzindo, também, os requisitos de armazenamento e, potencialmente, tempo necessário para a reconstrução do genoma.

## 2 Possíveis Avaliadores

- Nivan Roberto Ferreira Junior (nivan@cin.ufpe.br)
- Gustavo Henrique Porto de Carvalho (ghcp@cin.ufpe.br)

### 3 Cronograma

Atividade	Fev	Mar	Abr	Mai
Revisão Bibliográfica	X			
Desenvolvimento do Projeto	X	X	X	
Análise experimental comparativa de desempenho		X	X	
Escrita do Texto			X	X
Preparação da Defesa				X

## Referências

- [1] P. Ghosh and A. Kalyanaraman, “A fast sketch-based assembler for genomes.” Association for Computing Machinery, Inc, 10 2016, pp. 241–250.
- [2] R. Chikhi and G. Rizk, “Space-efficient and exact de bruijn graph representation based on a bloom filter,” p. 22, 2013. [Online]. Available: <http://www.almob.org/content/8/1/22>
- [3] T. C. Conway and A. J. Bromage, “Succinct data structures for assembling large genomes,” *Bioinformatics*, vol. 27, pp. 479–486, 2 2011.

## 4 Assinaturas

---

Paulo Gustavo Soares da Fonseca  
**Orientador**

---

Augusto Sales de Queiroz  
**Orientando**