

6 Conclusão

Neste capítulo é fornecido um resumo da pesquisa realizada. As contribuições deixadas por esta dissertação e a listagem dos trabalhos futuros, divididos em dois tópicos. Primeiro sobre as melhorias desejadas para o VelvetH-DB e o segundo sobre a importância de continuar o estudo Velvet, aprofundando o conhecimento do módulo VelvetG.

6.1. Resumo

Foi apresentado no capítulo 1 a introdução deste trabalho, trazendo um pouco do contexto, a definição do problema de pesquisa, os objetivos deste trabalho e a estrutura construída nesta dissertação.

No capítulo 2 explicamos o contexto onde a pesquisa foi realizada, na área da genômica, explicando em particular dois métodos de sequenciamento, Sanger, o mais tradicional, e o Illumina, principal método da nova geração de sequenciadores (NGS).

O capítulo 3 continua discutindo sobre a montagem de fragmentos, gerados na fase de sequenciamento. É explicado o método De Novo de montagem, onde não são utilizados genomas de referência.

Discute-se especificamente os problemas computacionais existentes na montagem de fragmentos, derivados pela característica da NGS, pelo volume de dados a ser processado e o alto consumo de memória RAM.

No capítulo 4 é descrita a implementação do Velvet. Trata-se de um conjunto de algoritmos criados para fazer a montagem de fragmentos curtos, que são a saída da NGS, no método De Novo. Esta implementação está dividida em dois módulos: VelvetH, foco de estudo deste trabalho, responsável pelo tratamento dos dados e remoção de erros, e o VelvetG, onde o grafo de Bruijn é construído e a montagem dos fragmentos é realizada.

Além de aprofundar o entendimento do módulo VelvetH, explicando o funcionamento das suas duas fases em detalhes.

O capítulo 4 termina com uma análise de consumo de memória principal na fase de geração das anotações e conclui que o trabalho a ser realizado deve ser a inclusão de uma camada de gerenciamento de memória principal nesta aplicação, pois a quantidade de dados a ser processado tende a ser maior com o barateamento dos sequenciadores e apenas fazer melhorias pontuais não é a solução definitiva para o problema de consumo de memória.

No capítulo seguinte descrevemos a proposta de migração da solução VelvetH, de uma aplicação com alocação dos dados em memória principal exclusivamente para uma solução em banco de dados, buscando robustez e eficácia no gerenciamento de grandes volumes de dados.

O capítulo 5 segue descrevendo o fluxo para a geração das anotações proposto, apresenta o esquema relacional e como se deu a implementação das funções em PL/pgSQL dentro do PostgreSQL, um banco de dados em código aberto, possibilitando com esta característica uma modificação estrutural no banco caso seja necessário em trabalhos futuros.

6.2. Contribuições

Há dois itens que podemos destacar como as principais contribuições desta dissertação de mestrado.

O detalhamento e entendimento do subprograma VelvetH é algo importante a ser estudado, pois trata-se de uma especificidade desta implementação. Pouco se encontra sobre o seu funcionamento e benefícios para todo o processo de montagem de fragmentos. Estudá-lo foi essencial para a identificar o porquê e como se dá a contribuição para a redução do consumo de memória na fase seguinte, a construção do grafo. Entender a sua estrutura de funcionamento, o formato dos arquivos de saída que ele gera e também o seu gargalo no consumo de memória RAM foi a primeira das contribuições desta dissertação.

A segunda contribuição deste trabalho está na reflexão e implementação da versão do VelvetH baseado em banco de dados. Por construção é robusta e eficaz para grandes volumes de dados. Mesmo que até o momento a implementação não

esteja eficiente, sendo mais lenta que a aplicação original, para conjuntos de dados menores, vemos como ponto de partida essa mudança na abordagem do problema, permitindo com trabalhos futuros a busca pela eficiência na geração dos Roadmaps através da solução VelvetH-DB.

6.3. Trabalhos Futuros

Sugere-se, dois tipos de trabalhos futuros: O primeiro cobre melhorias potenciais para a implementação do VelvetH-BD. O segundo trata do estudo e implementação de uma aplicação em banco de dados para a geração das outras fases do Velvet, que estão contidas no módulo VelvetG e envolve a construção do grafo.

Em relação ao primeiro caso, podemos destacar os seguintes pontos:

- Implementação da padronização dos diversos dos arquivos de sequências, não sendo mais necessária uma etapa anterior fora da aplicação
- Estudar estruturas de diferentes para as tabelas e índices do PostgreSQL
- Criar uma interface visual para melhorar a usabilidade da solução gerada

Em relação ao segundo, podemos destacar os seguintes pontos:

- Estudar e mapear o funcionamento do módulo VelvetG
- Identificar oportunidades de melhoria no funcionamento da aplicação e integração das fases para uma implementação em banco de dados
- Implementação de uma solução em banco de dados para a geração do grafo de Bruijn integrada à solução proposta do VelvetH-DB