

DESENVOLVIMENTO DE FERRAMENTAS PARA AUXILIAR NA VISUALIZAÇÃO E MONTAGEM DE GENOMAS

**Luciano Antonio Digiampietri, Camila Izidio Costa, Vivian Mayumi
Yamassaki Pereira**

Escola de Artes, Ciências e Humanidades, USP
{digiampietri, camila.izidio.costa, vivian.pereira}@usp.br

Objetivos

A montagem de um genoma consiste de, através da sobreposição de pequenos pedaços de DNA, se obter pedaços cada vez maiores e, por fim, o genoma completo. Muitas vezes este processo é comparado à montagem de um quebra-cabeça onde há peças repetidas, peças extremamente parecidas e, eventualmente, há peças faltando. O objetivo deste projeto é desenvolver ferramentas para auxiliar na visualização de genomas ou de partes dele, focando em trechos de montagem duvidosa. Além disso, algumas ferramentas foram desenvolvidas para encontrar evidências se estes trechos estão ou não montados corretamente.

Métodos/Procedimentos

A metodologia foi iniciada com a identificação, junto aos especialistas do domínio, das principais dificuldades quanto à montagem e validação da montagem de genomas de bactérias. Com base no estudo realizado foram definidas algumas ferramentas a serem desenvolvidas de forma a facilitar a visualização e encontrar evidências da corretude ou não de trechos específicos de uma montagem. As ferramentas desenvolvidas visam a adicionar funcionalidades a outras ferramentas existentes ou simplesmente facilitar a visualização de dados sem a necessidade da instalação de programas locais (através da disponibilidade dos mesmos em formato gráfico via internet).

Resultados

Até o momento, ferramentas com os seguintes objetivos foram desenvolvidas: ferramenta para a visualização do alinhamento de várias sequências contra um genoma completo; ferramenta para visualização na web de alinhamentos múltiplos; ferramenta para

identificar indícios da corretude ou problemas de trechos específicos da montagem.

A Figura 1 apresenta um exemplo de dúvida sobre um trecho da montagem de um plasmídeo de *Mycobacterium*. A região de w a x foi identificada numa montagem, porém no banco de genes GenBank esta região não existia. Obtivemos as sequências originais usadas na montagem e descobrimos que o trecho de w a x realmente existia, assim, muito provavelmente a montagem do plasmídeo depositada no GenBank está incorreto. Este tipo de problema é muito comum em montagens onde existem regiões de repetição com mais de 500 nucleotídeos (regiões R marcadas na figura).

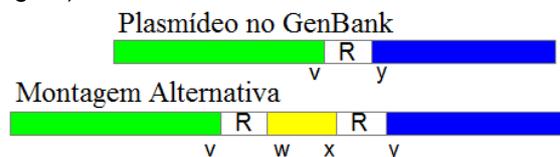


Figura 1 – Montagens alternativas avaliadas pelas ferramentas desenvolvidas

Conclusões

Este projeto está contribuindo para facilitar o processo de verificação de montagens de genoma através do desenvolvimento de ferramentas gráficas que facilitem a visualização das regiões de interesse e identificar indícios de corretude ou não de trechos específicos da montagem. Até o momento, já conseguimos a identificação de prováveis erros de montagem.

Agradecimentos

Agradecemos ao Programa de Educação Tutorial (MEC/SESu), a FAPESP e a Pró-reitoria de Pesquisa da USP que financiaram esse trabalho.