

Ciência da Computação / Sistemas de Informação

DESENVOLVIMENTO DE UM SISTEMA DE ANÁLISE DE REDES REGULATÓRIAS DE GENES EM PLANTAS

Luiz Victor Soriano da Conceição - 8º módulo de Ciência da Computação, UFLA, bolsista PIBIC/CNPq.

Lorrana Verdi Flores - Coorientadora, Pós-graduanda no Programa de Biotecnologia Vegetal, UFLA.

Gustavo José Rodrigues Pereira - 5º módulo de Sistemas de Informação, UFLA.

Renato Ramos da Silva - Professor do Departamento de Computação Aplicada, UFLA.

Joaquim Quinteiro Uchôa - Professor do Departamento de Computação Aplicada, UFLA.

Luciano Vilela Paiva - Professor do Departamento de Química, UFLA, Orientador. - Orientador(a)

Resumo

Experimentos de RNA-Seq geram vastos dados de sequenciamento que exigem processamento em múltiplas etapas para estimativas de expressão confiáveis. Porém, distintos scripts e etapas manuais dificultam sua expansão e a sua reprodução, enquanto alinhamentos incorretos podem deixar a inferência da rede enviesada. Para enfrentar esses desafios, o ATLAS foi desenvolvido como um fluxo de trabalho Nextflow orquestrado por uma plataforma web baseada em FastAPI-Celery, permitindo a execução assíncrona e desacoplada da análise de RNA-Seq. A metodologia começa com a extração automatizada de versões de anotação e montagem através de consultas que associam espécies escolhidas aos seus respectivos IDs e versões de genoma, as quais, combinadas com tokens de sessão, impulsionam o download dos arquivos de anotação correspondentes. Os dados de RNA-Seq são sistematicamente obtidos do SRA usando IDs de biblioteca, conseguidos por meio de scripts de pré-consulta, e baixados para processamento a jusante. A pipeline analítica central integra a indexação construída pelo Salmon com a construção de 'decoys' para um quasi-mapping preciso, o Trimmomatic para remoção adaptativa de adaptadores, e a quantificação feita pelo Salmon para produzir dados sobre TPM e suas contagens absoluta, complementada por relatórios completos de qualidade realizados pelo FASTQC, agregados pelo MultiQC. A operação da pipeline ATLAS é feita assincronamente, através de uma arquitetura desacoplada orquestrada pelo Celery; o Redis serve como broker de mensagens para sincronizar a comunicação entre os serviços. A pilha completa (FastAPI, Redis, workers Celery e banco de dados) é empacotada em containers e orquestrada pelo Docker Compose, proporcionando portabilidade e ambientes reproduzíveis em cada implantação. Essa abordagem integrada impulsiona futuros esforços visando integrar o ATLAS com outros serviços de bioinformática para facilitar a descoberta de genes biotecnologicamente relevantes em plantas.

Palavras-Chave: Nextflow, RNA-Seq, pipeline.

Instituição de Fomento: CNPq

Link do pitch: <https://youtu.be/KnRzLwHVTUs>