

Framework para Integração de Web Services em um Pipeline de Anotação de Sequências de Nova Geração

Milene Ferro¹, Wélliton Souza¹, Erik A. Antonio², Maurício Bacci¹

¹Laboratório de Evolução Molecular (LEM), Universidade Estadual Paulista “Júlio de Mesquita Filho” – UNESP/Rio Claro.

²Departamento de Ciências da Computação, Universidade Federal de São Carlos – UFSCar/São Carlos.

milenef@gmail.com, mbacci@rc.unesp.br

RESUMO

Análises genéticas usando sequências EST de formigas cortadeiras *Atta laevigata*, são fundamentais para a caracterização de espécies crípticas, castas ou para busca de alvos moleculares para controle como praga agrícola. A tecnologia dos sequenciadores de Nova Geração tornou capaz a produção de milhões de sequências de DNA em curto intervalo de tempo e com baixo custo [Carvalho e Silva 2010]. No entanto, processar esse grande volume de informação exige a integração de programas de bioinformática de forma automatizada. Em nosso laboratório está em desenvolvimento um *framework* flexível que representa e integra componentes para busca e tradução conceitual de ORFs [Ferro 2012], pesquisa de similaridade em banco de dados públicos usando BLAST, busca de domínios protéicos e termos GO com InterProScan, peptídeo sinal com SignalP e domínio transmembranar usando TMHMM. Inicialmente foi definida a arquitetura do *framework* e a forma de acoplamento de cada componente no *pipeline*, sendo que cada um foi elaborado como um serviço *web* em REST [Richardson e Ruby 2007]. A comunicação entre a aplicação cliente e o serviço *web* é coordenada pelo sistema de *pipeline*. Os resultados de cada componente no *pipeline* são integrados e armazenados no formato de intercâmbio XML/JSON, podendo estes ser acessados via *web*. Para validar o *pipeline* foram utilizados 32 mil transcritos obtidos pelo sequenciador de nova geração Illumina para a formiga *Atta laevigata*. O modo de acoplamento do *framework* contribui com a forma de reutilização dos componentes de bioinformática em um *pipeline* e serve de modelo para a elaboração de novos *pipelines* para outros organismos.

REFERÊNCIAS

- Carvalho, M.C.C.G. and Silva, D.C.G. (2010). “Sequenciamento de DNA de nova geração e suas aplicações na genômica de plantas”. *Ciência Rural*, Santa Maria, v.40, n.3, p.735-744.
- Ferro, M. (2012). WORF: Web Service for search and conceptual translation of ORFs. <http://omega.rc.unesp.br:8083/worfsite/>, abril.
- Richardson, L. and Ruby, S. (2007). *RESTful Web Services*, O'Reilly, p. 454.

APOIO FINANCEIRO: FAPESP