

Software para recuperação automatizada de anotações funcionais em estudos de associação genômica ampla

Rafael Câmara de Almeida¹

Gabriel Bueno de Oliveira¹

Roberto Hiroshi Higa²

O objetivo dos estudos de análise de associação genômica ampla, Genome-Wide Association Study (GWAS) é identificar marcadores moleculares do tipo Single Nucleotide Polymorphisms (SNP) associados com o fenótipo de interesse (ZIEGLER et al., 2008). Enquanto em estudos com humanos, o fenótipo considerado é a manifestação de doenças ou distúrbios com causa genética; no melhoramento animal, os fenótipos são características de interesse econômico como resistência a endo e ectoparasitas, maciez da carne, ausência/presença de chifre, etc. Esse tipo de análise envolve a utilização de métodos estatísticos e computacionais que manipulam uma grande quantidade de dados para encontrar um conjunto de SNPs que expliquem a variação observada.

Após a identificação de um conjunto de SNPs associados com o fenótipo estudado, a análise que se segue consiste em mapear esses SNPs no genoma de referência do organismo, identificar os genes presentes nas respectivas vizinhanças e procurar por informações sobre funções e vias biológicas relacionadas a eles, em bancos de dados público. Em geral, essa análise é feita de forma manual, acessando diferentes sítios na Internet, sendo os de principal interesse em genética animal o Kyoto Encyclopedia of Genes and Genomes (KEGG) (KYOTO ENCYCLOPEDIA OF GENES AND GENOMES, 2013), o National Center for Biotechnology Information (NCBI) (NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION,

¹ Universidade Estadual de Campinas - rcamara87@gmail.com; gabriel.oliveira@colaborador.embrapa.br

² Embrapa Informática Agropecuária - roberto.higa@embrapa.br

2013), o Gene Ontology (GO) (GENE ONTOLOGY, 2013) e o animal QTLdb (ANIMAL QUANTITATIVE TRAIT LOCI DATABASE, 2013). Neste trabalho, os bancos de dados disponíveis nestes sítios são utilizados para implementar um processo automático de mapeamento dos SNPs no genoma de referência, identificação de genes em sua vizinhança e busca por informações sobre funções e vias biológicas relacionadas a eles.

O processo implementado baseou-se na linguagem de programação Python (PYTHON, 2013) e banco de dados postgresql (POSTGRESQL, 2013). Ele possui uma etapa offline em que os dados de interesse são trazidos dos bancos de dados públicos e os organiza em um banco de dados local que relaciona informações de SNPs, genes, vias biológicas e anotação funcional. Esse banco de dados fica, então, disponível para consulta utilizando como query a lista de SNPs que resulta da análise de GWAS. O resultado da query é um arquivo HTML contendo uma tabela que relaciona os SNPs com genes próximos, suas respectivas anotações, incluindo links para os sítios originais para visualização de informações detalhadas.

Desta forma, espera-se contribuir para os pipelines de GWAS em genética animal, mantendo o foco do pesquisador que analisa os dados na interpretação das informações biológicas relacionadas ao estudo.

Agradecimento

Conselho Nacional de Desenvolvimento Científico e Tecnológico -(CNPq).

Referências

ANIMAL QUANTITATIVE TRAIT LOCI DATABASE **Animal QTLdb**. 2013. Disponível em: <<http://www.animalgenome.org/cgi-bin/QTLdb/index>>. Acesso em: 27 set. 2013.

GENE ONTOLOGY. 2013. Disponível em: <<http://geneontology.org/>>. Acesso em: 27 set. 2013.

KYOTO ENCYCLOPEDIA OF GENES AND GENOMES. 2013. Disponível em: <<http://www.genome.jp/kegg/>>. Acesso em: 27 set. 2013.

NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION. **NCBI resources** - gene database. 2013. Disponível em: <<http://www.ncbi.nlm.nih.gov/gene>>. Acesso em: 27 set. 2013.

POSTGRES SQL GLOBAL DEVELOPMENT GROUP. **PostgreSQL**: the world's most advanced open source database. Disponível em: < <http://www.postgresql.org/>>. Acesso em: 27 set . 2013.

PYTHON. **Python programming language** – official website. Disponível em: <<http://www.python.org/>>. Acesso em: 27 set. 2013.

ZIEGLER, A.; KÖNIG, I. R.; THOMPSON, J. R. Biostatistical Aspects of Genome-Wide Association Studies. **Biometrical Journal**, Berlin, v. 50, n. 1, p. 8-28. 2008.