

RA

SP 6185 P. 203
2013
SP-PP-6185

Obtenção de Metadados de Sequências Polimórficas em Banco de Dados Remoto

Leonardo Carvalho Nápolis Costa^{1*}, Leonardo Mariano Gravina Fonseca¹, Camillo de Léllis Falcão da Silva², Wagner Arbex¹⁺

¹Empresa Brasileira de Pesquisa Agropecuária – Embrapa
R. Eugênio do Nascimento, 610 – 36038-330 – Juiz de Fora – MG – Brasil

²Universidade Federal de Juiz de Fora – UFJF
R. Rua José Lourenço Kelmer, s/n – 36036-900 – Juiz de Fora – MG – Brasil

leonardocnc@gmail.com, leonardo.gravina@embrapa.br,
camillofalcao@gmail.com, wagner.arbex@embrapa.br

RESUMO. Polimorfismos de base única (*single nucleotides polymorphism* – SNPs) são as formas mais comuns de variações em sequências de DNA e definem-se como alterações de um único nucleotídeo em uma dada sequência, quando comparada com outras sequências alinhadas. Polimorfismos estão relacionados com variabilidade genética entre indivíduos e caracterizam-se como SNPs quando se encontram presentes em uma frequência alélica mínima de 1% em uma população [Brookes, 1999] [Arbex, 2009]. O estudo dessas alterações é importante para levar a compreensão do impacto dessas variações nas populações, para o entendimento de doenças, dentre outras. Existem diversas ferramentas computacionais cuja finalidade é auxiliar a pesquisadores no processo de busca por informação em diferentes áreas da bioinformática. Entretanto, mesmo com a evolução dos métodos e modelos que dão suporte à busca, pesquisadores ainda encontram dificuldades em agrupar e organizar os dados encontrados. Percebe-se que não há muita clareza no processo de recuperação e identificação dos dados, o que pode levar a incerteza ou até mesmo induzir o pesquisador ao erro. No caso do National Center for Biotechnology Information (NCBI) (<http://www.ncbi.nlm.nih.gov/>), o excesso de campos destinados à busca e a falta de conhecimento prévio sobre a ferramenta, dificultam a recuperação e a organização da informação que se deseja obter. Sob esse aspecto, esse trabalho trata da apresentação de uma ferramenta computacional que se destina a identificar e recuperar dados sobre SNPs de forma ágil e fácil, proporcionando um mecanismo de busca mais eficiente para o pesquisador, com a intenção de que se torne um futuro aliado no processo de validação dos resultados obtidos. A ferramenta em questão é eficaz para a investigação de informações sobre sequências de nucleotídeos sem que se tenha conhecimento prévio sobre elas – muitas vezes, conhecendo-se apenas de qual organismo são provenientes – e integra o *pipeline* SNPspy, que é um protocolo estabelecido a partir de descoberta de conhecimento em bases de dados, fundamentado em técnicas de aprendizado de máquina. Para sua implementação, juntamente com a linguagem Perl (<http://www.perl.org/>), utilizou-se a *tool box* BioPerl (<http://www.bioperl.org/>), que foi empregada para executar chamadas remotas a procedimentos. No caso, são executados procedimentos remotos de BLAST, disponibilizado no *website* do NCBI, automatizando a identificação das sequências e a recuperação de seus metadados sobre SNPs.

PALAVRAS-CHAVE. Polimorfismo de base única, SNP, SNPspy, NCBI, Bioinformática.

REFERENCIAS BIBLIOGRÁFICAS

Arbex, W. A. (2009) *Modelos computacionais para identificação de informação genômica associada à resistência ao carrapato bovino*. Doutorado em Engenharia de Sistemas e Computação, UFRJ. Trabalho vencedor do prêmio SBIAgro 2009 - Categoria Tese de Doutorado.

Brookes, A. J. (1999) The essence of SNPs. *Gene*, 2(234):177–186.

AGRADECIMENTOS: Os autores agradecem à Embrapa Gado de Leite, onde esse trabalho foi desenvolvido, e à FAPEMIG pela concessão da Bolsa de Iniciação Científica e Tecnológica

* Bolsista de IC da Fundação de Amparo à Pesquisa do Estado de Minas Gerais – FAPEMIG
+ Autor correspondente.

SP 6185
P. 203

