BANCO DE DADOS DE SNPs PARA FEIJÃO

MARCELO GONÇALVES NARCISO[1], TEREZA BORBA[2]

INTRODUÇÃO: Polimorfismos de única base ou SNPs ("Single Nucleotide Polymorphisms") são as formas mais comuns de variação genética em genomas eucariotos. Por serem estáveis e facilmente analisados através de metodologias automatizadas, os SNPs estão rapidamente substituindo os marcadores microssatélite ou SSR ("Simple Sequence Repeats") como marcadores genéticos (RAFALSKI, 2002). Entre as ferramentas utilizadas para a identificação destes marcadores estão o re-sequenciamento e modernas metodologias de bioinformática. A disponibilização e organização de um banco de dados de informações derivadas de re-sequenciamento permite a exploração destes dados no desenvolvimento de novos marcadores (SNPs), que funcionarão como ferramentas na exploração da diversidade armazenada em Bancos de germoplasma, no mapeamento de QTLs, na condução de estudos de associação e, consequentemente, no descobrimento de novos genes. Desta forma, justifica-se a construção de um sistema de informação para organizar os dados e facilitar o acesso, tanto quanto à inserção de dados no sistema quanto à recuperação destes para análises diversas. Assim, foi construído o sistema de informação de SNPs para feijão, o qual contempla um banco de dados sobre SNPs e uma interface web de acesso a esses dados. Este sistema oferece uma variedade de opções de consulta aos dados de SNPs, com possibilidade de salvar a consulta em arquivo texto ou HTML. Além destas consultas, este sistema oferece uma interface para acesso "navegadores" de genomas GBrowse (GBROWSE, 2011) e fornece arquivos de entrada para o visualizador de genoma Flapjack (FLAPJACK, 2011). Este sistema oferece também uma interface para a validação e construção de chip de SNPs e sistema tem código aberto. Feitas estas colocações, o objetivo deste trabalho é mostrar o sistema de informação de SNPs para Feijão, o qual foi feito conforme requisitos obtidos junto a biólogos e pesquisadores da área de biotecnologia e melhoramento genético.

MATERIAL E MÉTODOS: O sistema foi construído usando-se a linguagem Java Server Page (JSP) e JavaScript, para construção de uma interface web, e um sistema gerenciador de banco de dados MySQL. Através da interface web, o usuário faz diversas consultas ao banco de dados. JSP, JavaScript e MySQL são software livre. Para inserir dados na base de dados sobre SNPs, um programa feito em linguagem C é usado. Este programa lê arquivos de dados do tipo "texto" ou "csv" (comma sepparated values) sobre SNPs e insere os dados destes arquivos na base de dados MySQL. A interface web e a base de dados ficam no servidor web da Embrapa Arroz e Feijão. O programa JavaScript, que faz parte da interface web, é executado no computador do usuário para montagem de formulário e a verificação do correto preenchimento dos dados de entrada. Uma vez preenchido o formulário, a consulta é realizada. Caso o usuário cometa algum erro de preenchimento dos dados para fins de consulta, o programa JavaScript notifica o usuário sobre o erro no preenchimento dos dados. A Figura 1 ilustra a interface web. Além da interface web possuir formulário para preenchimento de dados para consulta, possui também opções para visualização de dados de cromossomo e de SNPs. Para isto, existe uma interface gráfica, que foi adotada como padrão recentemente na Embrapa, cujo nome é Gbrowse (GBROWSE, 2011). Com esta interface, que também é software livre, podem ser vistos diversos detalhes sobre um gene ou cromossomo. A interface Gbrowse é instalada e configurada no servidor web e esta necessita de um banco de dados PostgreSQL (POSTGRESQL, 2011) e também arquivos de dados de entrada no formato gff3, o qual contém as bases nitrogenadas componentes do genoma e informação sobre genes conhecidos do cromossomo. Gbrowse possibilita o download de parte ou todo o gene ou cromossomo, bem como diversas informações sobre genes conhecidos. Mais informações podem ser vistas em (GBROWSE, 2011). Para comparar visualmente SNPs de diferentes variedades, existe um visualizador cujo nome é FlapJack (FLAPJACK, 2011). Este visualizador deve

[1] Pesquisador da Embrapa Arroz e Feijão, Santo Antônio de Goiás, GO. E-mail: narciso@cnpaf.embrapa.br
[2] Pesquisadora da Embrapa Arroz e Feijão, Santo Antônio de Goiás, GO. E-mail: tereza@cnpaf.embrapa.br

ser instalado no computador do usuário (biólogo, pesquisador, analista, etc.) e utiliza os arquivos gerados pelo sistema de informação sobre SNPs de Feijão (interface web) como entrada de dados para então mostrar visualmente os SNPs. Um exemplo de uso deste sistema está descrito na Figura 3. Para a construção do chip (CHIP, 2011) sobre SNPs, última opção oferecida pela interface web, tal como ilustrado na Figura 1, foi usado o critério de avaliação das regiões flanqueadoras de cada SNP, conhecidas como upstream e downstream. Uma rede neural do tipo Perceptron (PERCEPTRON, 2011) foi usada para aprender a avaliação destas duas regiões e verificar se o SNP e as regiões flanqueadoras são de boa ou má qualidade (confiável). Outro fator de classificação foi a quantidade de bases C e G das duas regiões, que deve estar entre 30% e 60% do total de bases. Informações sobre como construir uma rede neural do tipo Perceptron e também como fazer o treinamento desta rede podem ser vistas em (PERCEPTRON, 2011).

RESULTADOS E DISCUSSÃO: Um dos grandes desafios dos pesquisadores é conseguir obter a informação organizada e acessível, de tal forma que se consiga os dados para análises diversas. Este sistema mantém as informações sobre SNPs organizadas e também possibilita a obtenção de dados sobre SNPs, além de poder visualizar os dados para análise usando-se os sistemas GBrowse e FlapJack. As consultas podem ser realizadas conforme a cultura de interesse (arroz ou feijão) e podem ser escolhidos diferentes genótipos da cultura escolhida, o cromossomo alvo, a representação gráfica dos diferentes cromossomos, o intervalo (distância) entre os SNPs e a verificação da presença de diferentes polimorfismos entre os genótipos. A Figura 1 ilustra as opções do sistema

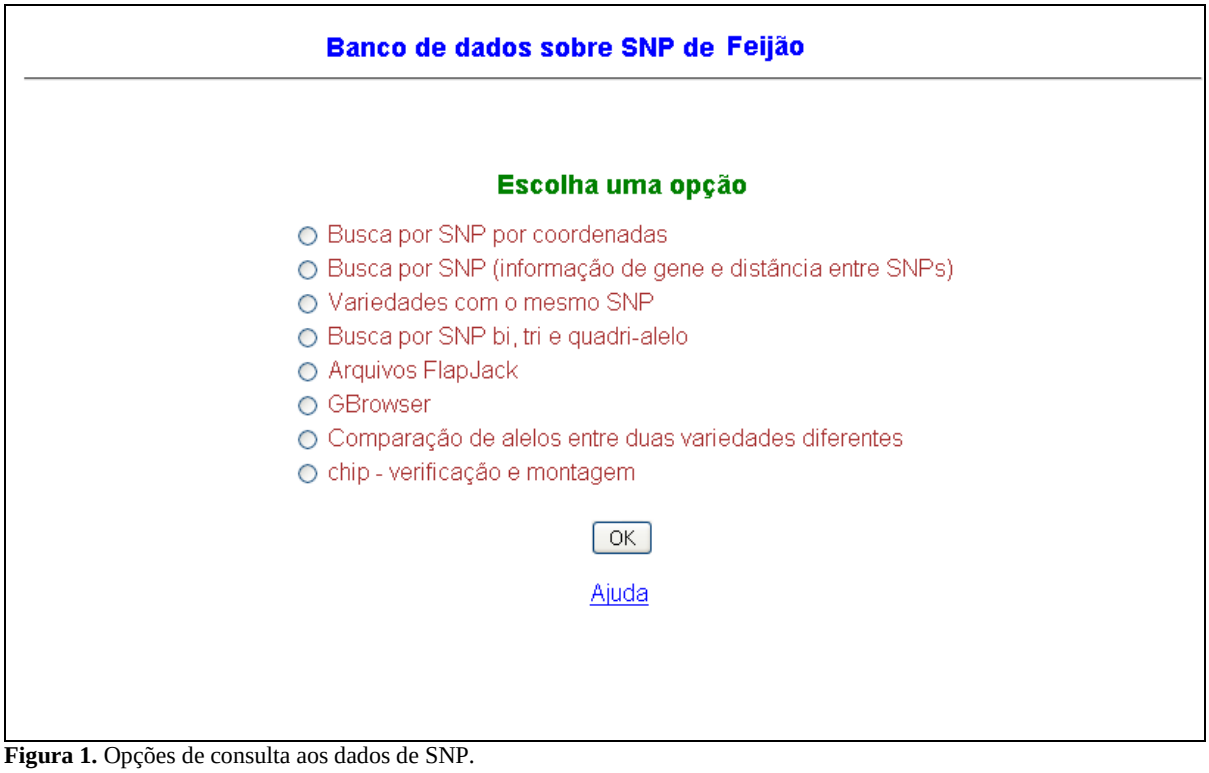

Figura 1. Opções de consulta aos dados de SNP.

A Figura 2 ilustra o resultado de uma busca por SNPs conforme coordenadas, para variedades AND 277, BAT 93, DOR 364. G 19833, Jalo EPP 558 e Ruda. Neste exemplo os dados estão em formato HTML. Existe também a opção de salvar este resultado em formato CSV, para download dos dados.

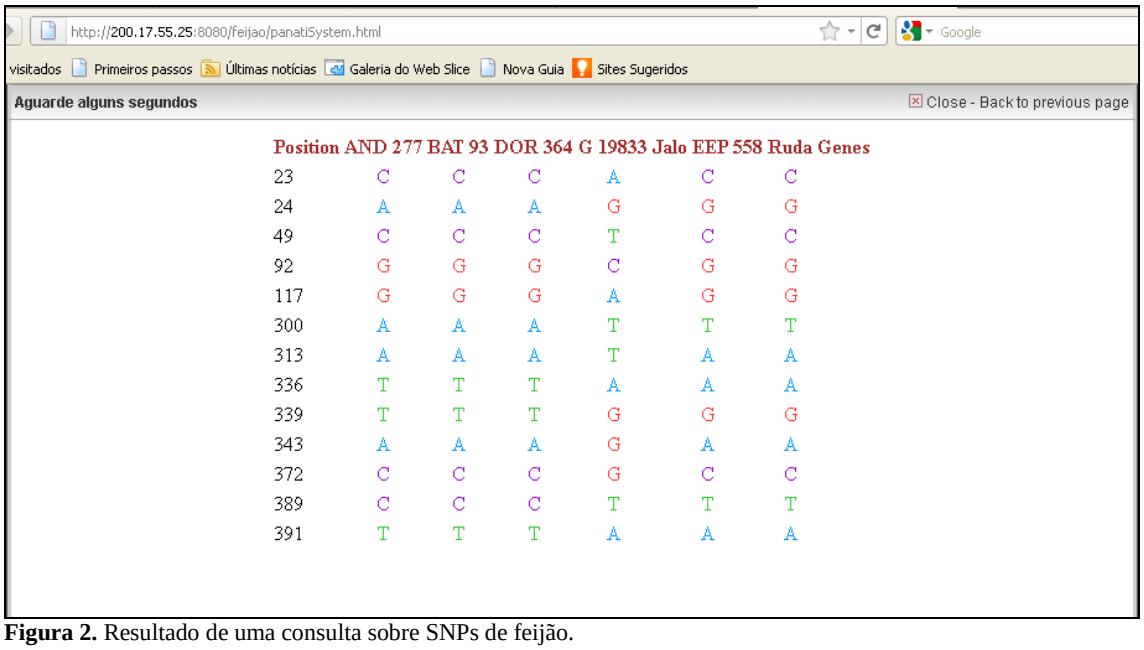

Figura 2. Resultado de uma consulta sobre SNPs de feijão.

A Figura 3 mostra um exemplo do sistema visualizador FlapJack em funcionamento para 3 cultivares de feijão: AND 277, BAT 93 e G 19833, para um dado cromossomo. No trecho deste exemplo é possível visualizar que as cultivares AND 277 e BAT 93 praticamente não apresentam polimorfismos de única base (SNPs) entre si, contudo a cultivar G 19833 apresenta polimorfismo (SNPs) expressivo quando comparada aos outros acessos. Assim, a ferramenta Flapjack evidencia as diferenças (polimorfismos) entre os diferentes genótipos.

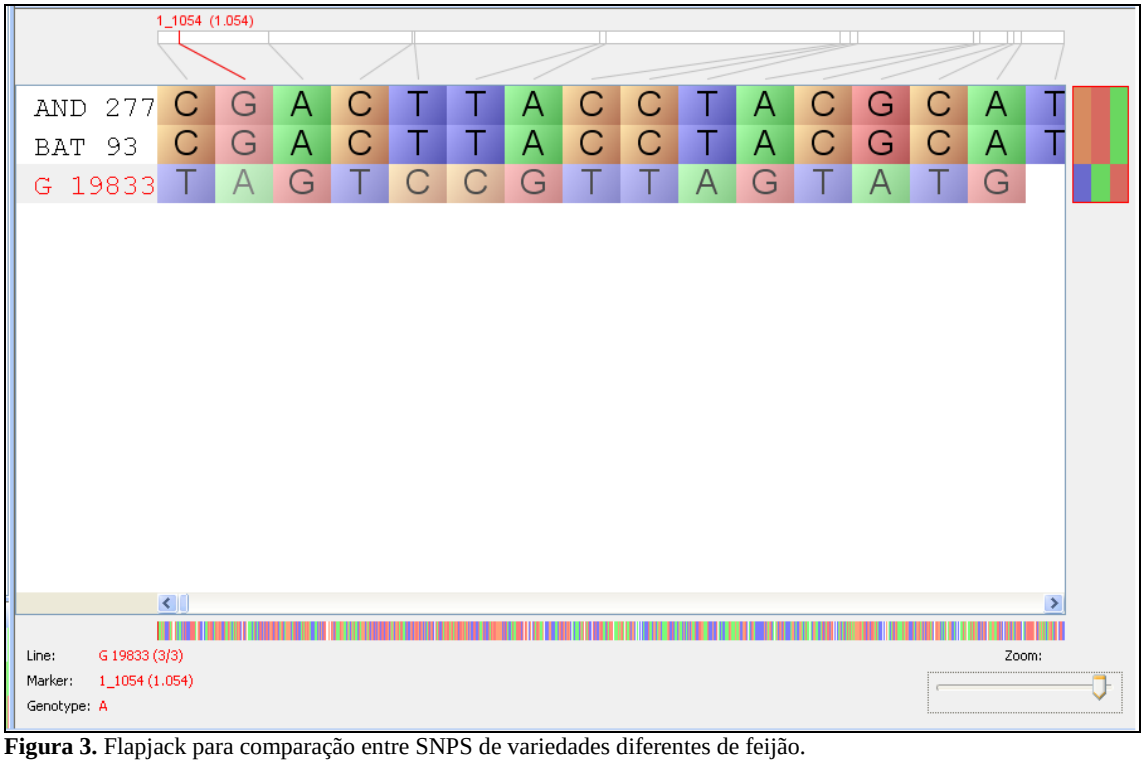

Figura 3. Flapjack para comparação entre SNPS de variedades diferentes de feijão.

Os dados resultantes das consultas relativas a SNPs (busca de SNPs por cromossomo, variedades, posição na cadeia ou coordenadas, chip) podem ser mostrados em formato HTML ou baixados em arquivo texto ou cvs. O limite da quantidade de SNPs que o sistema pode armazenar é conforme a capacidade física de armazenamento do servidor de banco de dados. Assim, à medida que novas

variedades de feijão vão sendo inseridas na base de dados sobre SNPs, o volume de dados vai aumentando, porém o desempenho do sistema pouco se altera. Foram feitas simulações para busca de SNPs em mais de 200 variedades na base de dados sobre SNPs e o desempenho foi considerado bom, com tempo de resposta máximo de 30 segundos (busca de todos os SNPs da base e tempo mínimo de 2 segundos (intervalo de 10.000 pares de bases). Isto acontece devido ao fato da base de dados ser indexada quanto à coordenada ou posição do SNP e cromossomo. Desta forma, o sistema tem bom tempo de resposta quando existem buscas ou consultas que retornam grande quantidade de dados de SNPs e referências diversas (base nitrogenada das variedades referência e outras, posição e cromossomo de cada SNP, referências ao gene, quando conhecido, que contém o SNP, etc.). Para finalizar, além de todas as ferramentas citadas, uma outra ferramenta, associada aos marcadores SNPs, bem como qualquer marcador, também faz uso da saída deste sistema. Esta ferramenta foi desenvolvida a fim de filtrar os alelos presentes somente naqueles genótipos que apresentam a característica de interesse (por exemplo, para eleger os marcadores potencialmente associados à resistência a seca). Embora não tenha força estatística, esta ferramenta é útil na escolha dos marcadores polimórficos que apresentam potencial de associação com a característica desejada. É software livre e pode ser obtida através de contato com os autores, bem como todas as ferramentas citadas neste artigo.

CONCLUSÕES: Existe um grande potencial quanto ao uso de marcadores SNP para a exploração da variabilidade genética das culturas de interesse bem como a associação de polimorfismos a caracteres de variação complexa ou de difícil mensuração. A disponibilização e organização de bancos de dados para o desenvolvimento e validação ensaios de SNPs customizados (personalizados) são extremamente valiosos a programas de melhoramento. Assim, uma vez, então, detectados e validados os SNPs ligados a um caractere de interesse, seria possível a seleção de genótipos com base no fenótipo do marcador.

REFERÊNCIAS

CHIP. Disponível em http://en.wikipedia.org/wiki/DNA_microarray. Visitado em 20/08/2011.

GBROWSE. Disponível em http://gmod.org/wiki/Gbrowse . Visitado em 20/08/2011.

FASSBENDER, H. W.; BORNEMISZA, E. **Química de suelos**: com énfasis en suelos de América Latina. 2.ed., San José: IICA, 1994. 420p.

FLAPJACK. Disponível em http://bioinf.scri.ac.uk/flapjack . Visitado em 20/08/2011.

MYSQL. Disponível em http://www.mysql.org/. Visitado em 20/08/2011.

PERCEPTON. Disponível em http://www.inf.unisinos.br/~jrbitt/annef. Visitado em 20/08/2011.

POSTGRESQL. Disponível em http://www.postgresql.org/. Visitado em 20/08/2011.

RAFALSKI, A. Applications of single nucleotide polymorphisms in crop genetics. **Current Opinion in Plant Biology**, v. 5, p. 94–100, 2002.