

SISTEMA PARA DETECÇÃO DE SNPS E INDELS. Vinícius Rodrigues da Cunha Perallis, Carlos Norberto Fischer, Darío Abel Palmieri, Delano Medeiros Beder. – Ciência da Computação – Bacharelado em Ciências da Computação - Departamento de Estatística, Matemática Aplicada e Computação – Instituto de Geociências e Ciências Exatas – Campus de Rio Claro.

Atualmente, as pesquisas relacionadas ao sequenciamento de DNA têm produzido um enorme volume de dados genômicos. Dentre estes resultados, estão informações relacionadas a polimorfismos em genomas. Assim, torna-se extremamente importante construir ferramentas computacionais auxiliares em processos de identificação e análise de polimorfismos.

Polimorfismos podem estar relacionados a alterações na sequência de nucleotídeos, eventos que podem ocorrer durante o processo de replicação do DNA. Dois tipos de polimorfismo de interesse são os chamados *SNPs* - *Single Nucleotide Polymorphisms* (polimorfismos referentes a apenas um nucleotídeo), que ocorrem especificamente devido à substituição de um nucleotídeo por outro, e os *InDels*, que são alterações no genoma provocadas por inserções e/ou deleções de um ou mais nucleotídeos (Bell, 2002).

SNPs constituem até 90% de todas as variações genéticas humanas e ocorrem, em média, a cada 1000 pares de bases nucleotídicas ao longo do genoma. Muitos projetos de pesquisa estão voltados para verificar a relação entre polimorfismos em DNA e como os indivíduos respondem, por exemplo, a doenças, patógenos e tratamentos. SNPs têm recebido muita atenção em pesquisas biomédicas e para o desenvolvimento de drogas terapêuticas (Bentley, 2000).

O Centro APTA Citros Sylvio Moreira (Centro de Citricultura - IAC), Cordeirópolis-SP, realiza vários tipos de pesquisa relacionados a citros, tendo uma grande coleção de *ESTs* (*Expressed Sequence Tags*) obtidas desse tipo de planta. Dentro dos interesses daquele centro estão os relativos a marcadores moleculares e polimorfismos. Com isso, pesquisadores do Grupo de Bioinformática de Rio Claro (GBIRC) e do Centro de Citricultura passaram a desenvolver, em conjunto, um sistema computacional que pode ser usado como ferramenta auxiliar em processos de detecção e análise de SNPs e InDels em bancos de *ESTs*, denominado SNP-ADApS.

O objetivo final no desenvolvimento do SNP-ADApS é criar um ambiente completo de análise que englobe a detecção de SNPs e InDels, a pré-seleção automatizada dos polimorfismos candidatos (de acordo com critérios estabelecidos pelo usuário), a geração dinâmica de tabelas e gráficos (dependendo do tipo de consulta definida sobre os resultados) e a especificação (o chamado “desenho”) dos *primers* correspondentes (*primers* são sequências flanqueadoras de uma região de interesse e são importantes para que cada SNP e InDel identificado possa ser validado em laboratório).

Este trabalho descreve o desenvolvimento de parte da ferramenta SNP-ADApS, especificamente no que se refere à implementação de um sistema computacional que detecta SNPs e InDels em bancos de *ESTs* e faz a pré-seleção automática dos marcadores identificados.

A ferramenta SNP-ADApS foi desenvolvida para o ambiente *web* (utilizando a linguagem PHP), o que permite acesso remoto por vários usuários dentro de um mesmo projeto. Basicamente, a ferramenta é um *pipeline* computacional que recebe como entrada sequências de bases nucleotídicas, através de arquivos em formato *fasta*. Estes arquivos devem ser carregados no sistema pelo usuário, sendo que cada um deles contém as sequências referentes a um dos genótipos, espécies ou populações para os quais se deseja identificar SNPs e InDels. A Figura 1 mostra a interface para carregamento desses arquivos.

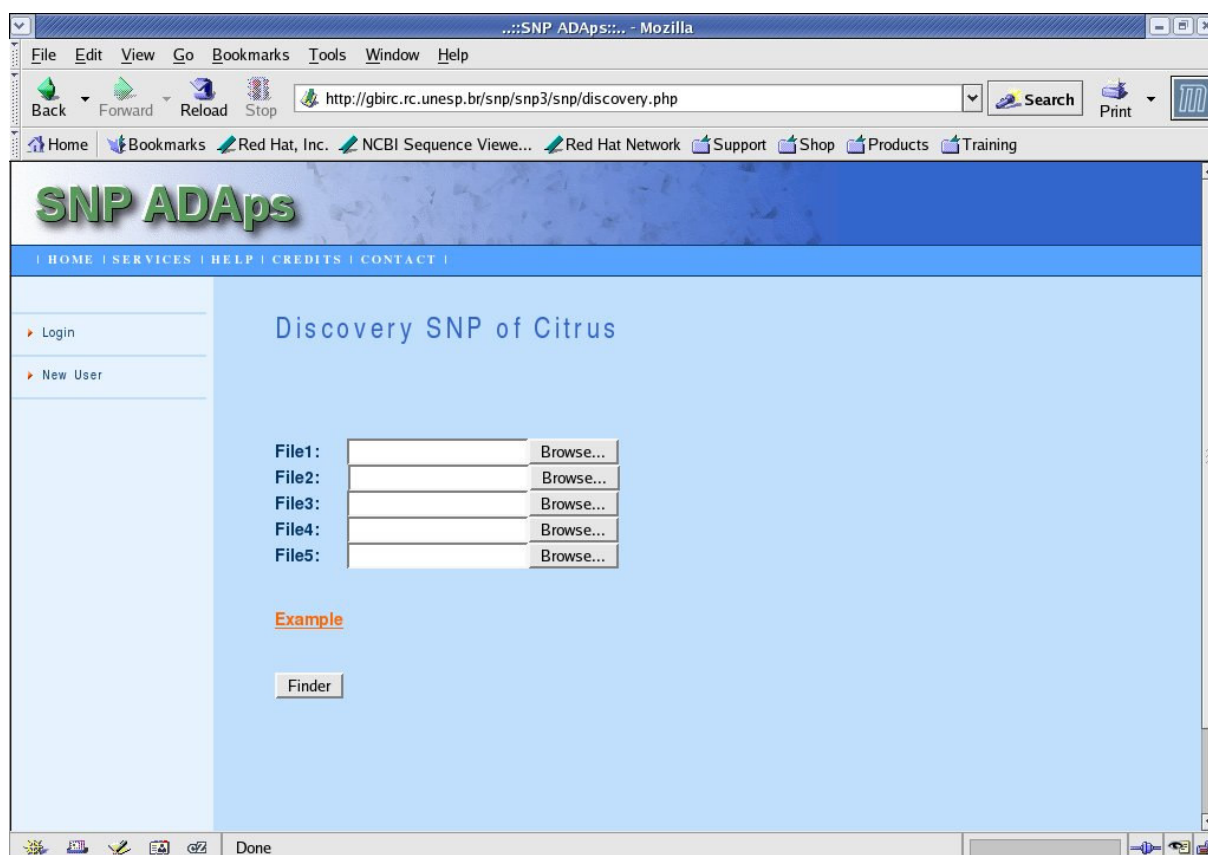


Figura 1 – Interface para carregar os arquivos de seqüências.

Inicialmente, o SNP-ADAPS processa os arquivos de entrada executando automaticamente os aplicativos *BLAST* e *CAP3* (ferramentas de acesso livre bastante conhecidas na área de genômica). Estes aplicativos são utilizados para alinhamento (*BLAST*) e montagem (*CAP3*) de seqüências de nucleotídeos, permitindo que possam ser realizadas comparações entre seqüências, com o objetivo de especificar onde há diferenças entre os indivíduos analisados. Esta abordagem (utilização dos aplicativos *BLAST* e *CAP3* em conjunto) foi a que melhor atendeu às necessidades de pesquisadores do Centro de Citricultura.

Dentro deste processo automatizado é gerado um arquivo tipo texto que contém informações sobre os alinhamentos das seqüências analisadas. Este arquivo é processado por outro programa, desenvolvido pelos autores (em linguagem Python), que, de acordo com as diferenças entre as seqüências de entrada (diferenças observadas através dos alinhamentos entre as mesmas), extrai todas as informações necessárias para a identificação, a caracterização e de posicionamento dos SNPs e InDels. As informações extraídas daquele arquivo são as relativas ao nome do *Contig* (nome dado ao conjunto de seqüências alinhadas) em que foi detectado o polimorfismo, o tipo do mesmo (se é relativo à Deleção - D, Inserção - I ou Substituição - S), à freqüência de ocorrência na seqüência do *Contig*, o seu tamanho e a localização do polimorfismo no contig. Os dados obtidos sobre os SNPs e Indels detectados são inseridos em uma base de dados relacional (construída utilizando o sistema MySQL), para que possam ser geradas consultas específicas sobre a mesma.

O SNP-ADAPS pode identificar um conjunto relativamente grande de SNPs e InDels, sendo que nem todos são de interesse. Assim, após a identificação preliminar de polimorfismos, torna-se necessário fazer uma pré-seleção dos mesmos, de modo a criar um conjunto mais restrito com os polimorfismos de real interesse. Para isso, o usuário deve especificar os parâmetros que serão usados para a seleção destes polimorfismos, como, por exemplo, o tipo de polimorfismo (D, I e S) desejado e tamanho do mesmo. A Figura 2 mostra a interface do SNP-ADAPS para a especificação desses parâmetros.

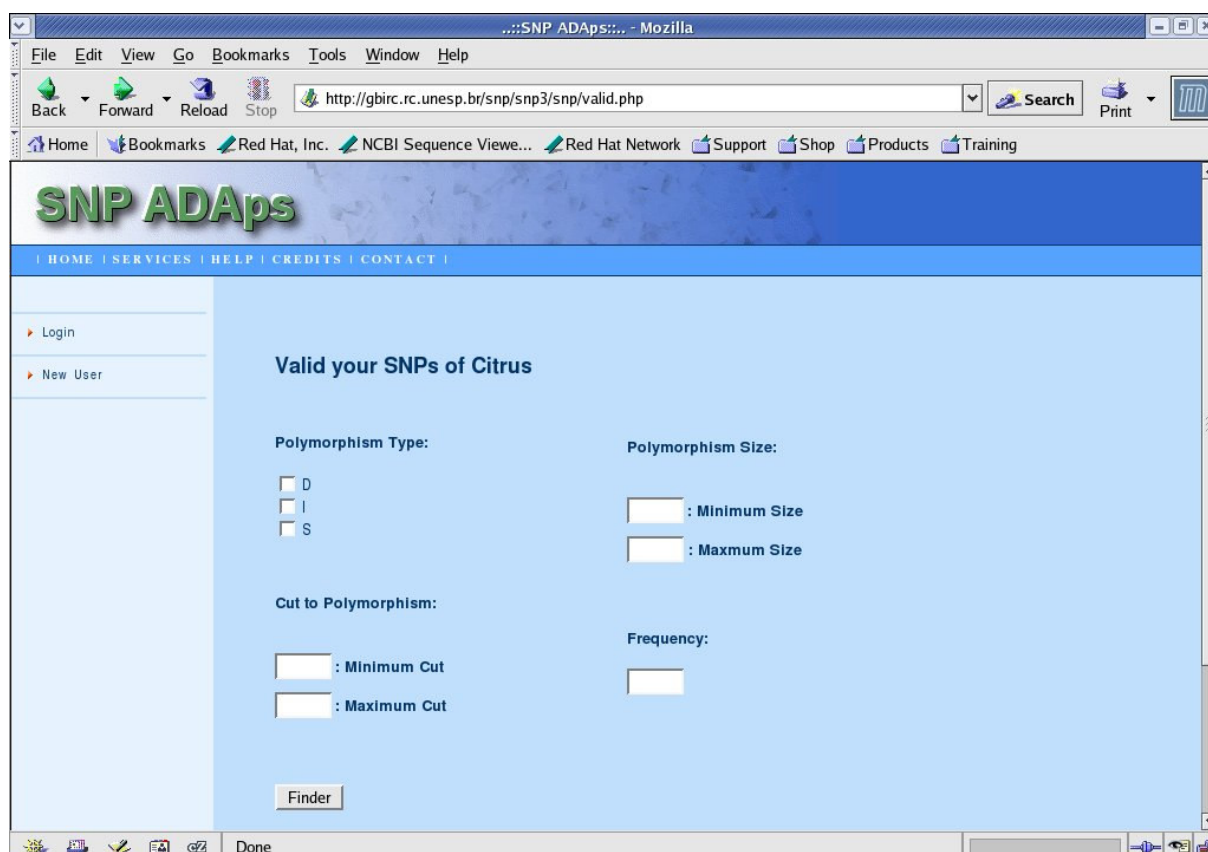


Figura 2 – Interface para a seleção dos parâmetros.

Baseado nestes critérios, o sistema seleciona os polimorfismos de interesse e os apresenta, ao usuário, juntamente com as respectivas informações. Um exemplo de polimorfismos selecionados pode ser visto na Figura 3.

A coluna “Generate” da interface mostrada na Figura 3 será utilizada para especificar os SNPs e Indels para os quais o usuário deseja desenhar *primers* específicos, visando a correspondente validação em laboratório. A descrição sobre o desenvolvimento das ferramentas correspondentes ao desenho de *primers* não é feita neste trabalho, pois, embora os programas computacionais que tratam disso já estejam em estágio adiantado de desenvolvimento, tais ferramentas fazem parte da etapa seguinte do projeto, assim como as ferramentas relativas à geração dinâmica de tabelas e gráficos, que deverão permitir a realização de análises mais detalhadas dos SNPs e InDels encontrados pelo SNP-ADAPS.

O sistema SNP-ADAPS está em fase final de testes pelo Centro de Citricultura e vem atendendo a todas as expectativas, o que deverá oferecer maior agilidade e precisão na identificação e caracterização de SNPs e InDels no banco de dados de citros.

SNP ADAPs

Page 1 of 721 Display 10 registers by page

Registers 1 to 10 of 7209

Contig	DIS	Cut	Size	Begin	End	Generate
Contig365	S	1/5	1	710	710	<input type="checkbox"/>
Contig365	S	1/5	1	727	727	<input type="checkbox"/>
Contig365	S	1/5	1	733	733	<input checked="" type="checkbox"/>
Contig365	S	1/5	1	754	754	<input type="checkbox"/>
Contig365	S	1/5	1	762	762	<input checked="" type="checkbox"/>
Contig365	S	1/5	1	767	767	<input type="checkbox"/>
Contig365	S	1/5	1	787	787	<input type="checkbox"/>
Contig365	S	1/5	1	806	806	<input checked="" type="checkbox"/>
Contig365	S	1/5	1	808	808	<input type="checkbox"/>
Contig365	S	1/5	1	825	825	<input type="checkbox"/>

Generate

Figura 3 – Exemplo de Polimorfismos identificados e selecionados.

Referências Bibliográficas

- Bell, J.I., 2002 - "Single nucleotide polymorphisms and disease gene mapping". ArthritisRes., 4 (Suppl 3), S273-S278.
- Bentley, D.R., 2000 - "The Human Genome Project--an overview". Med. Res. Rev., 20, 189-196.