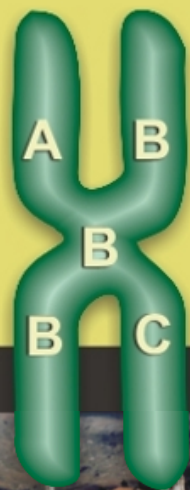6th International Conference of the Brazilian Association for Bioinformatics and Computational Biology

X-meeting

November 15 to 18, 2010

Parque Metalúrgico Centro de Convenções e Artes da UFOP , Ouro Preto, Minas Gerais, Brazil

november 16 to 18, 2010

Abstracts Book

Topic: **Databases and Bioinformatics Tools**
PI: **DBT06**

A BRAZILIAN SOYBEAN DATABASE

Nascimento L C[1], Costa G G L[1], Meyer L[1], Binneck E[2], Rodrigues F[2], Kulcheski F R[5], Margis R[5], Kido A[3], Marcelino F C[2], Nepomuceno A L[2], Abdelnoor R V[2], Pereira G A G[1], Carazzolle M F[1,4]

[1]*Laboratório de Genômica e Expressão - Departamento de Genética, Evolução e Bioagentes, Instituto de Biologia, Universidade Estadual de Campinas, Campinas - SP, Brazil.*
[2]*Embrapa Soja, Londrina - PR, Brazil.*
[3]*Universidade Federal de Pernambuco, Recife - PE, Brazil.*
[4]*Centro Nacional de Processamento de Alto Desempenho em São Paulo, Universidade Estadual de Campinas, Campinas - SP, Brazil*
[5]*Universidade Federal do Rio Grande do Sul, Porto Alegre - RS, Brazill*

Soybean is a legume with large economic importance in the international market, with a world production of almost two hundred and ten million tons in the 2008/2009 harvest. Brazil appears as the second largest producer, with about twenty-five percent of the world production. In 2007, the Brazilian Soybean Genome Consortium (GENOSOJA) was established with the goal of integrating several institutions currently working with soybean genomics in Brazil. The project has an initiative to search for new treats to improve the soybean production process, emphasizing in stresses that affect the national production, like the occurrence of droughts, pests attacks and the Asian Rust disease. Among the objectives of GENOSOJA is the creation of a relational database, integrating the results achieved by different methodologies utilized in the project.

In the GENOSOJA context, we created a brazilian soybean database, integrating: (1) public data consisting of genome and predicted genes from JGI, an assembly of 1,276,813 of *NCBI* ESTs from several cultivars and 4,712 full-length cDNA sequences from one japanese cultivar; and (2) private data consisting of (*i*) three cDNA libraries explored by SuperSAGE methodology, resulting in 4,373,053 tags of 26 bp, (*ii*) 22 cDNA subctrative libraries from several brazilian cultivars under different stresses and (*iii*) several libraries of soybean microRNAs under eight conditions and size between 19 and 24 bp. All these data were sequenced using Solexa/Illumina sequencing technology. This database offers to the users some features, including keywords searches, statics comparisons, automatic annotation, gene ontology classification and gene expression of the genes under certain conditions. All data are storage in a Fedora Linux machine, running the MySQL database server. The web interface is based in a combination of CGI scripts using Perl language (including BioPerl module) and the Apache Web Server.