Development and Use of a Tool for Automated Alignments of Genes in the Rice BAC's GenBank Card Against Other Species

P. Barre
INRA, France

G. Darrieutort
INRA, France

J. Auzanneau
INRA, France

B. Julier
INRA, France

Development and use of a tool for automated alignments of genes in the rice BAC's GenBank card against other species

P. Barre, G. Darrieutort, J. Auzanneau and B. Julier

INRA Poitou-Charentes, Unité de Génétique et d'Amélioration des Plantes Fourragères, route de Saintes, 86600 Lusignan France Email: barre@lusignan.inra.fr

Keywords : sequence alignment, BLAST, Sequence Tagged Site, *Lolium perenne*., homologous genes

Introduction In many cases, the analysis of the genetic bases of any trait requires molecular markers and if possible co-dominant PCR-based ones. In perennial fodder species, the number of publicly available markers (microsatellites and Sequence Tagged Site (STS)) is limited. Our goal is to use sequences from model grass species, i.e. rice, wheat, maize, barley, in *L. perenne* in order to develop STS markers in interesting regions such as under a QTL (Quantitative Trait Loci) or around a candidate gene,. As the genome sequence of rice is now available, the objective was to use the sequences of genes included in the BAC's GenBank card from rice. As there are almost no available sequences in *L. perenne*, we are designing consensus primers from an alignment of at least two different species. The problem is that for all the genes included in a BAC, just a few have their sequences known in at least two species. It is very laborious to check "by hand" if each gene has an homologous sequence known in another species.

Materials and methods We have developed a tool for automated sequence alignments of genes from the rice BAC's GenBank card against several species specific *Génoplante-info* data bases and screened the results. This tool includes two applications:

The first application consists of:
- a graphic interface for the user to enter BAC numbers and data base names on his local computer (PC);
- a procedure to transfer the files with BAC numbers and data base names from the local computer to the Génoplante-info computer;
- a procedure to perform sequence alignments between genes extracted from the BAC's GenBank Card and the *Génoplante-info* data bases selected by the user;
- a procedure to transfer the files with alignment results from the Génoplante-info computer to the local computer.

The second application consists of:
- a graphic interface for the user to enter his thresholds for parameters involved in the definition of homology between genes such as the e value, the bit score, the percent of homology and the length of the alignment;
- a procedure to screen all the results of various alignments using sorting parameters defined by the user ;
- a visualisation for each gene of species where homologous genes exist on an Microsoft Excel sheet.

Results In an analysis of the genetic basis of aerial morphogenesis in perennial ryegrass varieties, we are analysing linkage disequilibrium in the region of OsGAI gene (AY464568) in different varieties. For this purpose, we entered 13 rice BAC's Genbank card names from TIGR (including OsGAI and others around). The 232 genes present in the BACs were aligned against 6 species specific databases (1392 BLAST) in 1h30. We found 39 genes with homologous sequences in at least two species that were potentially interesting for STS development and STS marker primers were produced.

Conclusions i) Biologists who have used this tool find it convenient and time saving to select genes for which STS markers could be developed even if more automation of primer development is required ii) The fact that the program is split in two applications allows one to apply different thresholds on the alignment results and compare the screenings without again computing all the alignments. iii) This tool, developed for the Génoplante-info environment, could be easily adapted to other environments.