Introduction

Banana (Musa spp.) is cultivated in numerous tropical countries throughout the world, and in many of these countries its cultivation and marketing play very important roles, both economically and socially. In 2004, Brazil produced 6,602,750 tons of bananas in an area covering 484,981 ha [1]. The majority of the banana farmers are small-scale producers with the crop grown predominantly as a supplementary source of income. Developments in molecular biology have provided tools to enable insights into changes in the transcriptome that arise when a plant is submitted to different kinds of stresses. For the assessment of gene expression, methodologies such as large-scale single pass sequencing of cDNA clones to generate expressed sequence tags (ESTs) can be utilized. ESTs provide a quantitative method to measure specific transcripts that are transduced through a chain of signaling molecules that ultimately, affect regulatory elements of stress-inducible genes.

To help to identify genes related to biotic and abiotic stress in banana a web based tool called PHPH was used [4]. We describe a keyword-based search in the DATA database for genes known or related to biotic and abiotic stress, as well as the base calling, quality assignment and assembling of 20 candidate genes sequences using PHPH tool.

Materials and Methods

To identify genes related to the biotic and abiotic stress resistance in Musa acuminata a “virtual screening” was made in the transcriptome part of the DATA database (http://genoma.embrapa.br/musa/musa.html) [5]. The transcriptome part of this database consists of 5,571 M. acuminata EST Sequences (MaEST). Data was a result of a collaborative project sponsored by CNPq, and developed by Embrapa Genetic Resources and Biotechnology (Embrapa Genomas), Brasilia Catholic University (UCB) and the Agricultural Research for Developing Countries (CRAD) in France. These three institutions are also part of the Global Musa Genomics consortium (GMGC).

From the selected sequences retrieved from the DATA_Musa database, their correspondent electropherograms were analyzed using the PHPH tool. The sequences were submitted to a PHPH tool (version 4) using as an interface a web-browser. All the file manipulations and the calls for the analysis programs were developed using a PERL programming language [6] and a CGI interface. For the quality analysis a PHRED [7, 8] package was used. To mask out the vector parts that might be present within each sequence the CROSSMATCH [9] program was used. Optionally the user can run a CAP3 [10] program for the assembly, checking the sequences of interest (figure 3). A color code showing the sequence quality was used as shown in figure 4-A. A freely available chromatogram viewer (applet) [11] was used in order to show the trace (figure 4-C). This Applet can read SCF files, generated by PHRED (version 2 or 3) and ABI sample files. Also the user can save the coloured sequence in RTF format.

Results and Discussion

Using PHPH was possible to check the sequence quality automatically using the PHRED program. The user parameterized the PHRED quality and their sequences were grouped by CAP3 program resulting in contigs and singlets. Using the generated consensus sequence (figure 5), a BLAST [12] search was made against SwissProt database [13]. So far, using the PHPH tool was identified 20 genes that are related to biotic and abiotic stress in Musa acuminata, such as chitinase, pathogenesis-related protein (PR-10), germin-like protein, ascorbate peroxidase, glutathione peroxidase, selenium binding protein, heat shock proteins, polygalacturonase, peroxidiseoxyn, superoxide dismutase, salt tolerance protein, lectin, 14-3-3 protein among others.

Conclusion

The data presented in this study provide a first general overview of the genes related to biotic and abiotic stresses present in DATA, Musa database and the possibility to use PHPH tool as assembler for small EST projects, with the usual pipeline from the electropherogram analysis to sequence assembly, all built-in in a single run. The PHPH tool can also be used for rapid quality analysis of the sequences generated by the automatic sequencer. Thanks to BIOFOCO (http://www.biofoco.org) a group of researchers engaged in the bioinformatics multidisciplinary work, PHPH can be accessed in the different addresses:

http://adenina.biolor.unb.br/phph (since August 2001 at Brasilia University)
http://genoma.embrapa.br/phph (at Embrapa Genetic Resources and Biotechnology)
http://bioinformatica.ucb.br/phph (at Brasilia Catholic University)

BIOFOCO is a group of researchers engaged in the bioinformatics multidisciplinary work. The main field is the development of new tools for genomics using state of the art in information technology, and gather four institutions: Embrapa - Recursos Genéticos e Biotecnologia, UCB (Universidade Católica de Brasilia), UnB (Universidade de Brasilia) and UFMS (Universidade Federal de Mato Grosso do Sul).

Bibliography


Figures:

Figure 1 - A screen shot of the DATA_Musa web page. http://genoma.embrapa.br/musa

Figure 2 - PHPH initial screen and the result after submitting the sequence. From this screen the user can see the quality table, the vectors used for assembly, checking the sequences of interest and the best frame was chosen to be used.

Figure 3 - Created quality table. From this screen the user can see the individual sequence quality and select the sequences to execute the CAP3 program.

Figure 4 - The sequence quality screen. The sequences (A) are shown using the color code (B) depending on its quality. The chromatogram is shown using a IOXApplet (C).

Figure 5 - (A) shows the result after running the CAP3 program, (B) shows the result from the DATA_Musa database using the keyword: PR10. (C) NCLB hit list result from the assembled contig. This contig was annotated by the PEN InSite tool and the best BLAST search was chosen to be used.