Cry-Bt identifier: A biological database for PCR detection of Cry genes present in transgenic plants

Vinay Kumar Singh, Sonu Ambwani, Soma Marla*, Anil Kumar

Bioinformatics Centre, Molecular Biology & Genetic Engineering, CBS&H, G.B. Pant University of Agriculture & Technology, Pantnagar, Uttarakhand.263145, India: Soma Marla - E-mail: ssmarl@yahoo.com; *Corresponding author

Abstract:
We describe the development of a user friendly tool that would assist in the retrieval of information relating to Cry genes in transgenic crops. The tool also helps in detection of transformed Cry genes from *Bacillus thuringiensis* present in transgenic plants by providing suitable designed primers for PCR identification of these genes. The tool designed based on relational database model enables easy retrieval of information from the database with simple user queries. The tool also enables users to access related information about Cry genes present in various databases by interacting with different sources (nucleotide sequences, protein sequence, sequence comparison tools, published literature, conserved domains, evolutionary and structural data).

Availability: http://insilicogenomics.in/Cry-btIdentifier/welcome.html

Keywords: HTML, PERL, PHP, Access, MySQL, rice, Cry-Bt

Background:

Recent advances in molecular biology and genetic engineering help scientists to better understand the various mechanisms underlying plant development and growth. This in turn resulted in development of more productive crops possessing resistance to biotic and abiotic stresses by the introduction of foreign genes. Cry genes from *Bacillus thuringiensis* are important sources for development of insect/pest resistance in plant varieties [1]. Data from increased use of Cry genes in the development of transgenic plants by scientist world wide demands databases and tools for efficient storage/retrieval of required data. Hence, we describe the development of a method to use Cry gene database that enable identification of Cry genes present in transgenic crops [2-4]. This will further help in the design of primer sets for the PCR identification of specific Cry genes. The tool also enables users to access information related to Cry genes present in different online literature, nucleotide/protein sequence databases for sequence comparison and analysis [5-7]. The tool is designed based on relational database model that interacts with different sources of information located in various online-databases.

Methodology:

Tool Design and implementation:

Cry-Bt identifier was developed and deployed using an open source software system. The system used is MySQL database to store gene and protein sequence data. We used GenBank to retrieve gene sequence data and the data was processed using a filtering algorithm**.** The filtering algorithm was implemented using a set of routine scripts written in Hypertext Preprocessor (PHP). A GenBank record is considered for retrieving a sequence if the Gene name contains `DNA` or `PROTEIN`. Additionally, the GENE record in the FEATURES should also contain terms describing CRY-BT (for example, 'Cry1Aa' or 'Cry1Ab' or 'Cry1Ac').

Programs written in PHP enabled database search using keywords like `organism name`, `Gene name`, `Gene Id.` and `Gene & Protein Sequences` of Cry-Bt genes. The database facilitates to retrieve results in `Text` or `Table` or `Graphical` formats. Detailed sequence and literature data can be obtained from the corresponding hyperlinks. Relational database schema employed in design of the tool provides representation of the Cry-bt database system with its filtering algorithm to store, retrieve data with out redundancy up to 3rd normalization form with no loss of functional information [8-11].
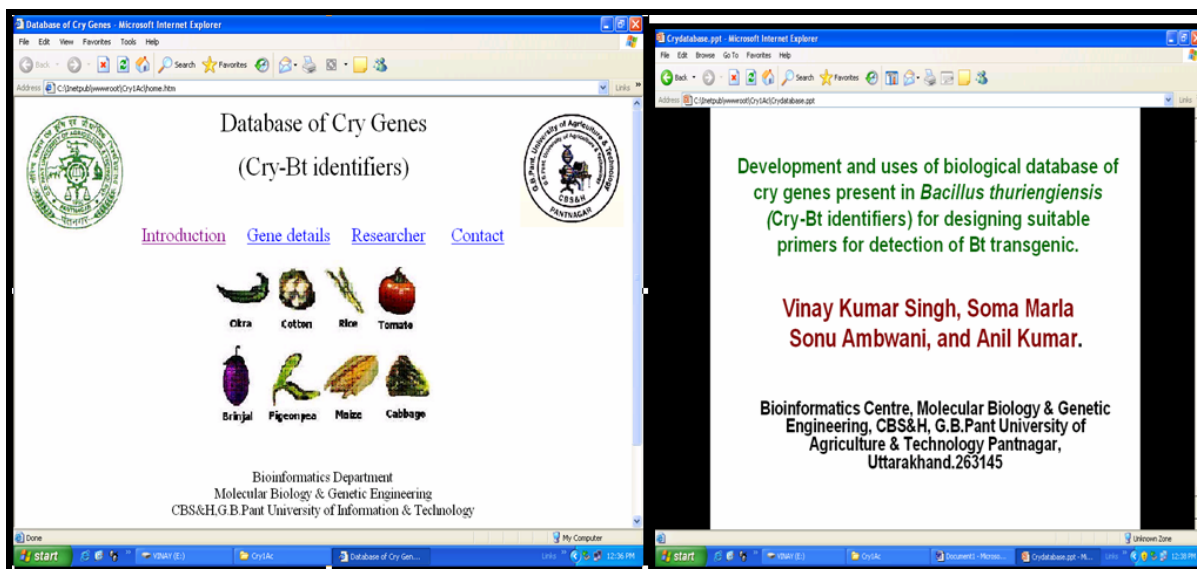

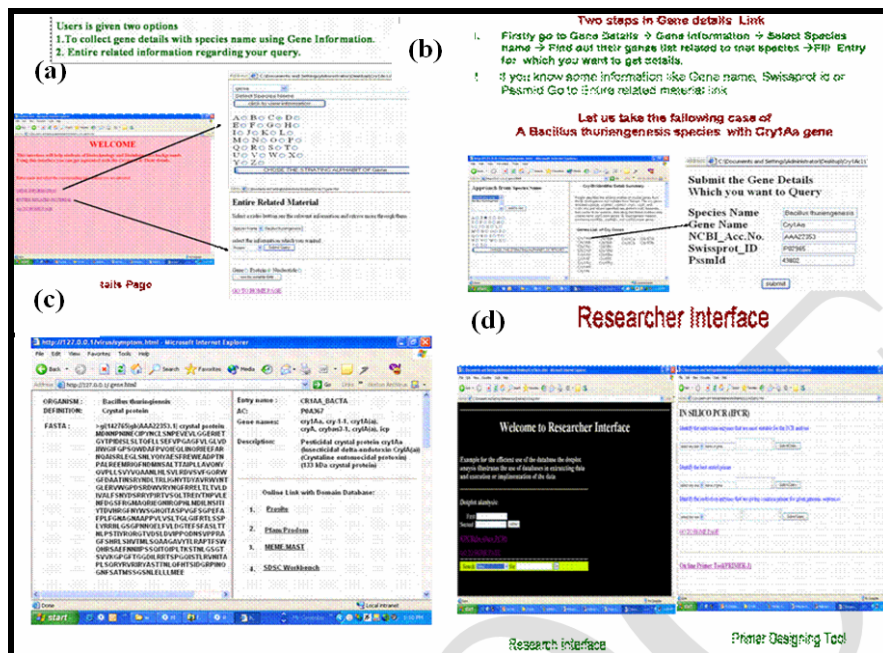**Figure 1:** Front view of Cry-Bt identifier.

Figure 2: Representation of Various features (a, b, c &d) that can be performed using Cry-Bt Identifier

The tool is also embedded with a primer design tool- E-PCR (Electronic PCR). EPCR employs a computational procedure that identifies sequence tagged sites (STS), present in the user fed DNA sequences. E-PCR initially looks for potential STS sites present in DNA sequences by searching subsequences that closely match to the PCR primers with correct order, orientation, and spacing that are used to generate known Sequence Tagged Sites.

Embedded E-PCR enables a choice of three options: (1) Identification of suitable restriction enzymes suitable for the given sequence. (2) Designing a Primer where at the starting positions are placed restriction sites for amplification. (3) Designing a common primer for two genome sequences for a given restriction enzyme. This feature enables to find out the best restriction enzyme for the both the homologous sequences. Designing a common primer can significantly cut the expenditure on primer design. Here we included information here about online linkage with our *Cry-bt identifier* tool to perform in bioinformatical research work. Pfam is a large collection of multiple sequence alignments and hidden Markov models covering many common protein domains and families, ProDom is a comprehensive set of protein domain families automatically generated from the SWISS-PROT and TrEMBL sequence databases, helps to discover motifs (highly conserved regions) in groups of related DNA or protein sequences using MEME and, search sequence databases using motifs using MAST.

Primer3 is a tool for picking primers from a DNA sequence. The Biology WorkBench is a web-based tool for biologists. The WorkBench allows biologists to search many popular protein and nucleic acid sequence databases. Database searching is integrated with access to a wide variety of analysis and modeling tools. Thus, the developed Cry-Bt identifier database will enable to provide useful information to users.

References:
[1] HC Sharma, KK Sharma, *Innovations in production, biotechnology, quality, and marketing*, pages 53 (2006)
[2] S Tounsi *et al., Biotechnol Lett*. **21**: 771 (1999) [PMID:]
[3] PA Kumar *et al., Advances in Applied Microbiology* **42**:1 (1996) [PMID: 8865583]
[4] R Bhalla *et al., FEMS Microbiology Letters* **243**: 467 (2005)
[5] GT Vilas-Bôas, MV Franco Lemos, *Can J Microbiol/Rev can microbiol.* **50**: 605 (2004)
[6] FC Chen *et al., Biotechnology Letters*, 247 (2006) [PMID:]
[7] CM Beron *et al., Appl Environ Microbiol.* **71**:761 (2005) [PMID:]
[8] Y Mohamed, *et al., CIDR.* 196-206 (2007)
[9] J Scaria *et al., Bioinformation* **1**: 5 (2005) [PMID:000000]
[10] F Rong *et al., Bioinformation* **1**: 16 (2005) [PMID:000000]
[11] J Dai *et al., Bioinformation* **1**: 214 (2006) [PMID:0000000]