

## PREDICTION OF BACTERIAL INFECTION IN COTTON USING EXPRESSED SEQUENCE TAGS

Neeta Maitre<sup>1\*</sup> and MM. Kshirsagar<sup>2</sup>

<sup>1</sup>Department of Computer Science and Engineering, G. H. Raisoni College of Engineering, Nagpur, Maharashtra, India.

<sup>2</sup>Department of Computer Technology, Yeshwantrao Chavan College of Engineering, Nagpur, Maharashtra, India.

### ABSTRACT

Expressed Sequence Tags (EST) are short sub-sequence of cDNA sequencing approach. Available ESTs can be used to tag almost 60% of the genes in a genome. ESTs can act like a powerful tool in hunt for genes involved in case of diseases. ESTs are short sequences and hence they are easy to compare. Comparing these ESTs for local alignment can be used to predict particular genetic configuration in a given genome assembly. This paper talks about the prediction of bacterial infections caused by *Pantoea Agglomerans* in Cotton breed, *Gossypium Raimondii*. BLAST is used to compare sequences and to generate its relationship with the actual genome mapping.

**Keywords:** Expressed Sequence Tags, gene prediction, BLAST, *pantoea agglomerans*

### INTRODUCTION

Cotton is a one of the important crop plants. Fibers of cotton are used to make cloth and seeds are processed to produce oil. There are 50 different species of cotton across America, Asia, Africa and Australia. Agricultural science talks about many of the infections, plants can have and importantly concerned about the growth phase of plant. There are various breeds of cotton available. Major four of them are *G. raimondii*, *G. arboreum*, *G. hirsutum* and *G. barbadense*. Like all other plants, this plant is also susceptible to diseases. Diseases could be bacterial, fungal or viral. Lint degradation is one of the symptom of bacterial disease and it results in internal necrosis and rot of immature cotton bolls. It causes premature death of cell and usually no external boll damage can be observed. Usually no external boll damage is observed but in sensitive cotton breeds, the damage can be significant. To address this problem, an approach of comparing genomic sequences can be adopted to have a speculation of bacterial infection. In order to generate this

comparison, genomic sequences causing infections can be compared with genomic sequences of cotton. This approach needs local similarity searches and calls for a time-consuming genomic comparison. Thus an efficient way of comparing expressed sequence tags (EST) can be used to reduce efforts in long comparisons. Expressed sequence tags are generated by shot sequencing of cloned cDNA/mRNA. ESTs are small pieces of DNA sequence (200-500 base pairs) and as these are from cDNA library, they lack introns and can directly be compared with DNA sequences. The idea used here is to find the chromosomes those may be affected by bacterial infection through sequence comparison of ESTs and genome by using a computer program. The results are then compared with genomic comparison of both for the same. The computer program which can do this sequence comparison is BLAST (Basic Local Alignment Search Tool). BLAST is extensively used in comparison of genomic sequences. We employed a technique to see similarity of the genetic constitution of cotton plant against the

genetic constitution of *Pantoea agglomerans*. It is an effort made to find traces of this disease in the cotton breed *G. Raimondii*.

#### **MATERIALS AND METHODS**

*G. Raimondii* is a D-genome diploid cotton breed having size of nearly 880Mb. Genome is all the DNA contained in an organism or a cell which includes chromosomes plus DNA in mitochondria. The genomic sequences of *G. Raimondii* and *Pantoea* are downloaded from NCBI (National Center for Biotechnology Information) alongwith the EST sequences of *Pantoea agglomerans*. As ESTs are cDNA sequences, they contain only exon-sequences which maps genes in an organism. The genome for cotton breed, *G. raimondii*, is analyzed for the occurrence of bacterial infection through a BLAST search against *Pantoea agglomerans*. BLAST search is preferred over global alignment search as coding regions are to be analyzed for the existence. BLAST is Basic Local alignment Search Tool. BLAST is generally used to identify unknown sequence, build homology tree for a protein and to map a sequence in a genome. For genomic comparison, datasets from National Center for Biotechnology Information are downloaded and then are analyzed to see the variation in results.

#### **RESULTS AND DISCUSSION**

The genome Vs genome comparison is made for *G. Raimondii* and *Pantoea Agglomerans*. The whole genome comparison resulted in query cover of "zero" value which lacks addressing of even a small tracing of disease. The other option chosen was to query *Pantoea Agglomerans* ESTs against the *G. Raimondii* genome. As ESTs are cDNA representation of genome it can directly be compared with the cotton genome. It lacks intronic region and hence can be mapped on to the location in chromosome using BLAST. We have BLAST individual *Pantoea* ESTs against

cotton genome. The results of BLAST shown query coverage lying in the range 3-40 with the similarity identity of 75% -100%. The results for maximum score values are shown in following diagrams.

The results shown traces of infection in specific set of chromosomes. For the given dataset, it has been seen that chromosome1 and 5 are affected more as compared to other chromosomes using EST similarity with cotton genome. The similar kind of result is evident from the genomic comparison of *Pantoea Agglomerans* and *G. Raimondii*. The speed of EST comparison is found to be high as compared to the genomic comparison due to small size of ESTs. It is found to be beneficial as ESTs deal with genetic material only.

#### **CONCLUSIONS**

All species of plants are subject to disease. Plant diseases are a normal part of nature and is considered to be one of many ecological factors that help keep the hundreds of thousands of living plants and animals in balance with one another. Cotton plant is not an exception to these situations. Cotton has a very high farm-gate value and is an essential commodity in the global economy. The advent of new genetic technology, increase in plant gene sequence data is dramatic. This data provides opportunities to get improvements in the yield, fiber quality and disease resistance. Many of these technologies have been designed to facilitate detection and understanding of gene expression that accompany differential development. EST (Expressed Sequence Tags) projects are very useful in quick access of gene sequences and expression information. Gene prediction, finding or discovery is assisted very well by EST projects. Genetic comparisons reveal possibility of disease causing genes in plants and can be judged for its quality.

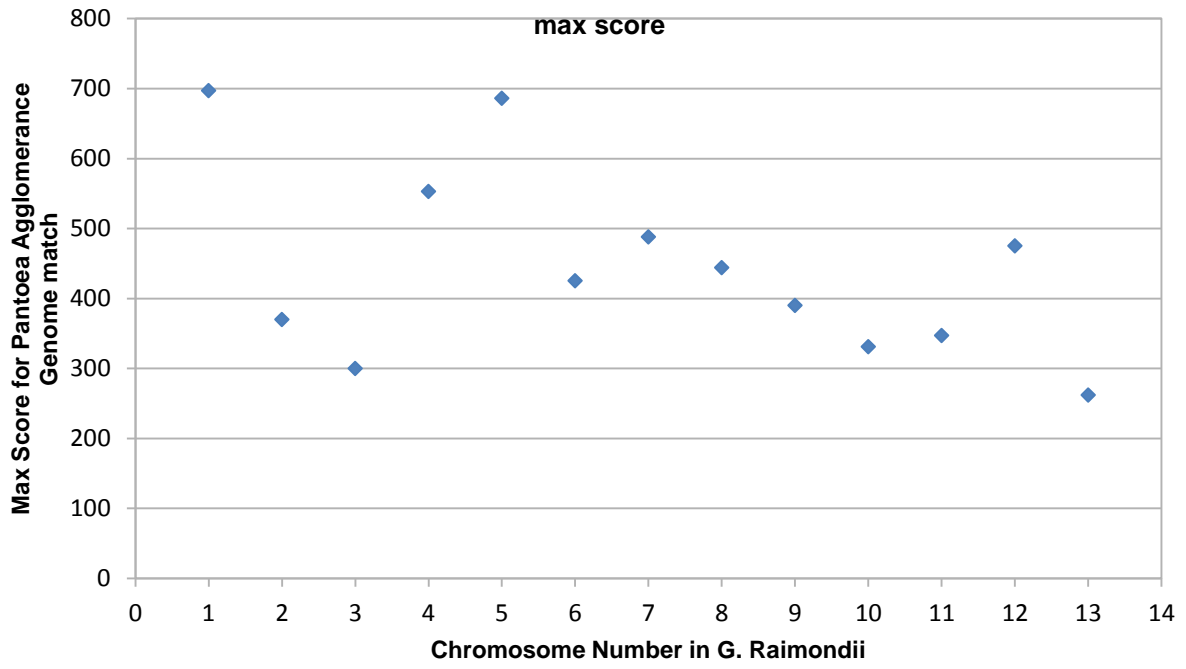


Fig. 1: Maximum score values for Pantoea genome against G.Raimondii genome

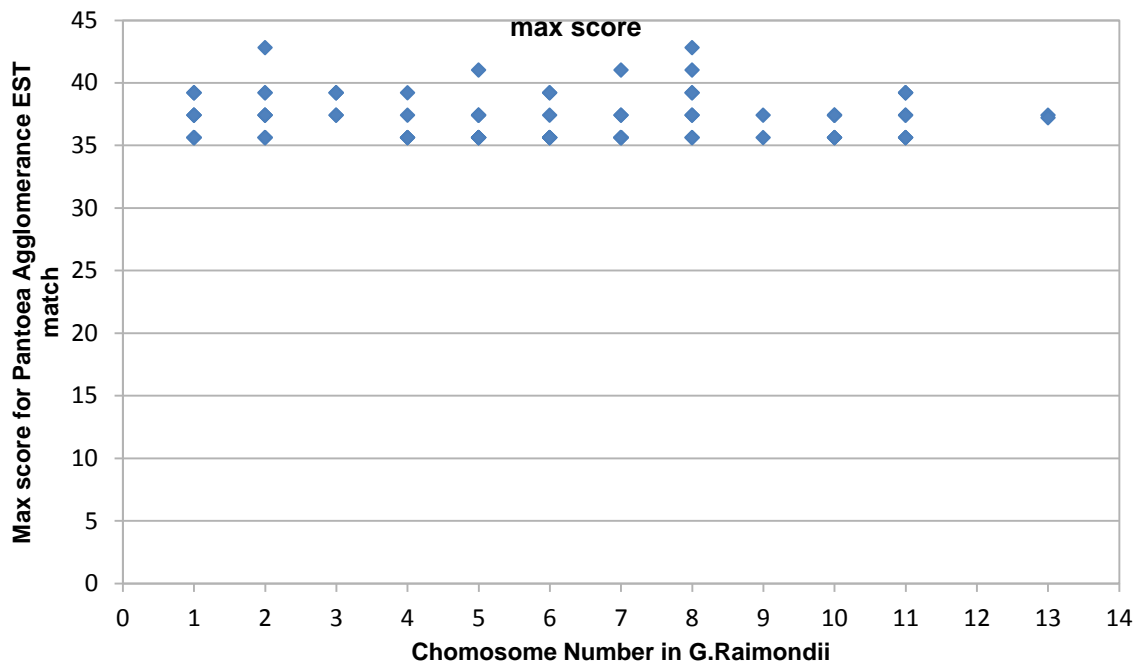


Fig. 2: Maximum score values for Pantoea ESTs against G.Raimondii genome

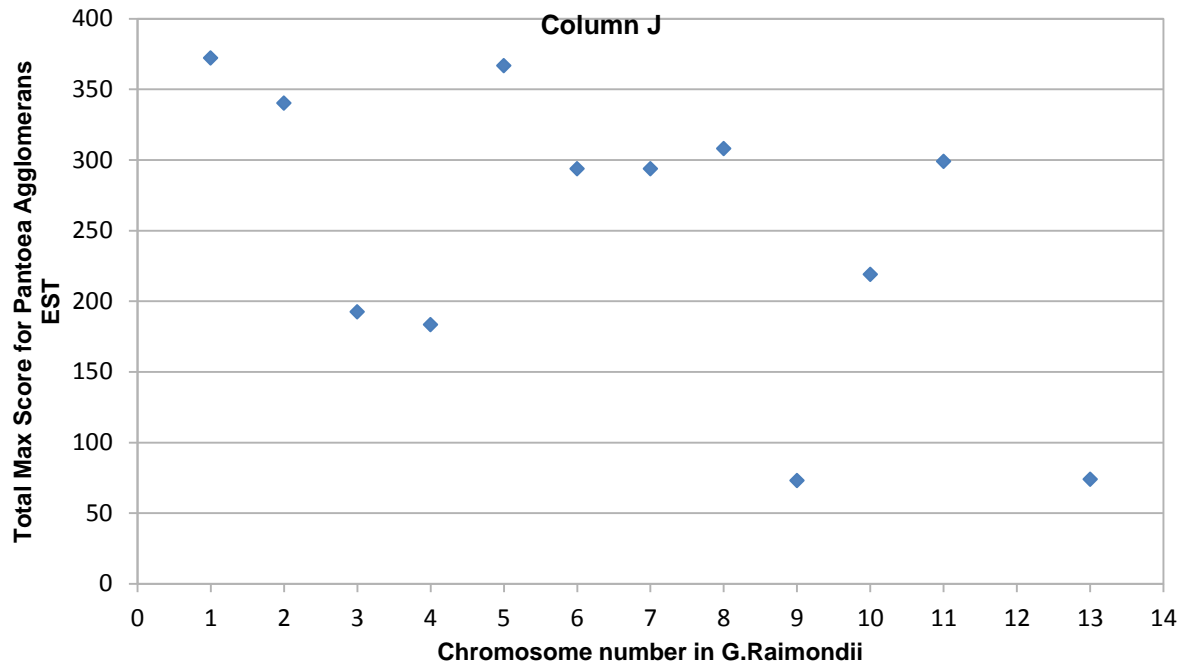


Fig. 3: Total Maximum score for Pantoea ESTs against G.Raimondii genome

#### REFERENCES

1. Glen L. Ritchie, Research Coordinator, Craig W. Bednarz, Philip H. Jost and Steve M. Brown, Cotton Growth and Development.
2. Catherine Mathe, Marie-France Sagot, Thomas Schiex and Pierre Rouze, Current methods of gene prediction, their strengths and weaknesses, Nucleic Acids Research, 2002, 30 (19):4103-4117.
3. B. Jayaram and Priyanka Dhingra, Bioinformatics for better tomorrow.