

## To Whom It May Concern

This is to certify that, *Rubaida Easmin*, MSSE0206, Institute of Information Technology, University of Dhaka, will work on the research topic “Finding Motifs for Producing Salt-Tolerant Variety of Rice” under my supervision until June 2015.

I wish her research turn out to be a success.



**Shah Mostafa Khaled**

Assistant Professor

Mobile: +8801812698789

Email: khaled@univdhka.edu

Institute of Information Technology

University of Dhaka

# **Finding Motifs for Producing Salt-Tolerant Variety of Rice**

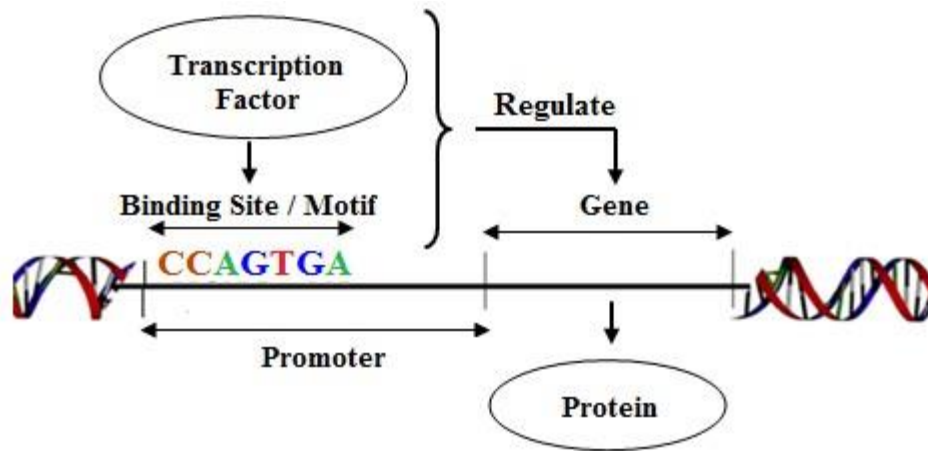
## **Introduction**

Despite having multiple problems, Bangladesh has made considerable progress in the food sector. Our country has achieved self-sufficiency in rice, the dominant staple food, increasing production from 11 million tons in 1971 to 33 million in 2012. However, recent changes in the climate hinder the continuity of this progress. Increasing abnormalities in climates patterns along with its associated hazards like sea level rise, cyclone and storm surge have been reducing soil fertility and amplifying salinity especially in the coastal region [10]. Thus it will eventually become hard to maintain the achievement in agriculture which Bangladesh has made during the last four decade. To continue the progress in rice production, measures has to be taken to utilize the coastal lands with high salinity and breed crop varieties with higher salt tolerance suited to saline environments. This research intends to contribute in the same direction.

## **Rationale**

Salinity is a major factor limiting plant productivity, affecting about 95 million hectares world-wide. High salinity is one of the most serious threats to crop production.

Salinity is a major constraint to rice cultivation as it limits the production and restricts the use of farming land. Generally it grows well in flooded fields with hot and humid climate. This is why South and Southeast Asia is the main region to produce rice. Because of environmental changes, this territory is highly affected by salinity which is minimizing the quantity of cultivable lands. Besides, overpopulation increases the consumption of rice. To handle all these situations improved rice varieties should be introduced, which may survive in coastal regions.



**Figure 1: Transcription Factor Binding Site (motif) in Gene Sequence**

In order to adapt in such adverse condition, rice evolve different metabolic activities [9]. These activities are controlled by certain genes [11]. Genes are often called the blue print of life as they give instruction to each of our cells about what to do and when to do it. These genes are activated by the signal of promoter, which is the upper stream regulatory region of gene. Whenever a regulation protein binds to this upstream region, the corresponding gene activates a transcription process. This binding site is known as Transcription Factor Binding Sites (TFBS) or motifs [4] which comprises of a set of similar oligomers (short part of gene sequences) containing some variations among them. The ultimate product of this transcription is a protein which helps an organism to respond in different environmental stress. Figure 1 gives an overview of this process.

Identification of these motifs is important to understand the regulatory aspects of the gene expression .Therefore, biologists and computer scientists have been very interested in identifying computational tools for motif finding from the gene sequences. Since these sequence data are large in size, it becomes extremely difficult to find such motifs and still mechanisms like reacting to saline environment are unknown. The objective of this research is to overcome that difficulty and identify salt-responsive oligomers from the gene sequence data which will help in finding desired motifs. The positive or negative impact of these motifs on gene regulation will also be investigated.

## Existing motif finding methods

Despite considerable efforts to date, DNA motif finding remains a complex challenge for biologists and computer scientists. Researchers have taken many different approaches in developing motif discovery tools and the progress made in this area of research is very encouraging.

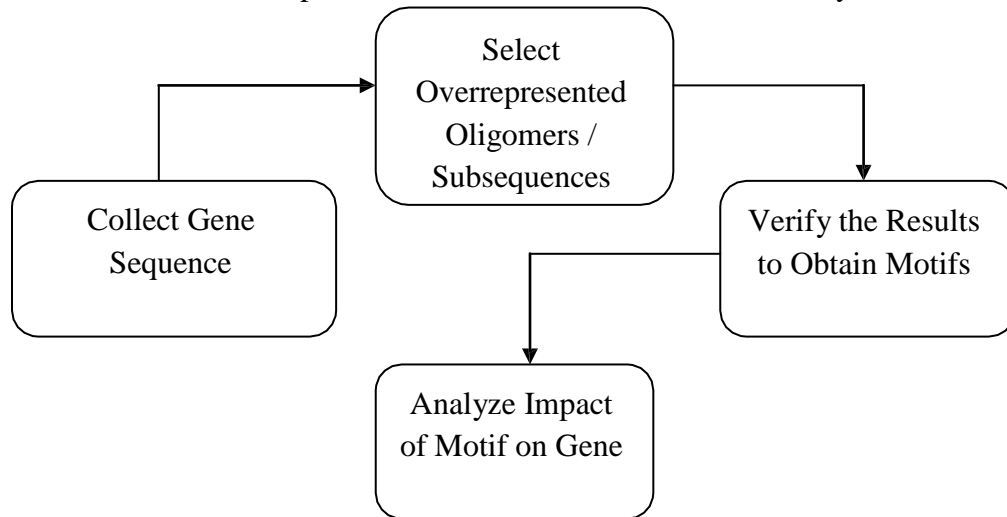
Gibbs sampling is a generalized probabilistic algorithm that generates a sequence of samples from a joint probability distribution of two or more random variables. This algorithm is presented in [12]. Gibbs sampler needed observable data that remains constant throughout the process and some parameters.

Another motif finding algorithm, named as Expectation-Maximization (EM) algorithm, is presented in [7]. However, the basic model assumption in this algorithm is that each sequence must contain at least one common site which is a lacking of this method. MEME [8] algorithm is an extension of EM algorithm, which includes a number of features to improve this task. The aim of MEME is to discover new motifs in a set of sequences where little is known in advance about any motifs that may be present.

Summary of the existing algorithm regarding motif identification is provided in this section. It is seen that a number of methods have been developed based on varied and complex motif models for different organisms [5]. However, most of those algorithms determine common motif pattern like TATA box or GCC box. Moreover, there is no existing method which only focuses on finding the salt tolerant motifs of rice. Thus it is required to propose method that will analyze the gene sequence of rice and identify the motifs.

## Proposed Method

The aim of this research is to identify salt responsive motif from the gene of rice. For this purpose we will use sequence data of rice gene. Initially these data will be collected from existing literature. Two different varieties of rice are selected for this purpose. One is called Pokkali – a unique saline tolerant rice variety that is mostly cultivated in the water-logged coastal regions and another one is Japonica which is a salt-sensitive rice variety.



**Figure 2: Proposed Model for finding salt-responsive motifs in rice**

At the initial level our task is to analyze promoter regions of these genes. Then motif discovery method will be implemented to find oligomers that are significantly enriched in these dataset. The obtained result will be justified by analyzing existing works whether they are motif or not. Our final task is to identify positive or negative impact of those motifs on the salt stressed gene. Figure 2 gives an overview of the proposed model of this research.

## Conclusion

Considering future eco-system, traditional rice varieties need to be improved. A number of tasks such as identification of corresponding salt tolerant genes, their functionalities etc. should have to be done and this project will persuade these activities by using gene sequence data. Knowing motifs, which may regulate stress responsive functionalities, will aid in taking appropriate steps to breed rice varieties.





## References

- [1] Lawrence, Charles E and Altschul, Stephen F and Boguski, Mark S and Liu, Jun S and Neuwald, Andrew F and Wootton, John C, “Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment”, American Association for the Advancement of Science, pp: 208-214, 1993.
- [2] Van Helden, Jacques and Andre, Bruno and Collado-Vides, Julio, “Extracting regulatory sites from the upstream region of yeast genes by computational analysis of oligonucleotide frequencies”, Journal of molecular biology, Elsevier, pp: 827-842, 1998.
- [3] Thijs, Gert and Marchal, Kathleen and Lescot, Magali and Rombauts, Stephane and De Moor, Bart and Rouz, Pierre and Moreau, Yves, “A Gibbs sampling method to detect overrepresented motifs in the upstream regions of coexpressed genes”, Journal of Computational Biology, Mary Ann Liebert, Inc, pp: 447-464, 2002.
- [4] Kilpatrick, Alastair M, “Regulatory motif discovery in magnetic bacteria”, Citeseer, 2009.
- [5] Hu, Jianjun and Li, Bin and Kihara, Daisuke, “Limitations and potentials of current motif discovery algorithms”, Nucleic acids research, Oxford Univ Press, pp: 4899-4913, 2005.
- [6] Roth, Frederick P and Hughes, Jason D and Estep, Preston W and Church, George M, “Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation”, Nature biotechnology, pp: 939-945, 1998.
- [7] Lawrence, Charles E and Reilly, Andrew A, “An expectation maximization (EM) algorithm for the identification and characterization of common sites in unaligned biopolymer sequences”, Proteins, pp: 41-51, 1990.
- [8] Bailey, Timothy L and Elkan, Charles, “Unsupervised learning of multiple motifs in biopolymers using expectation maximization”, Machine Learning, Springer. pp : 51-80, 1995.
- [9] Yokoi, Shuji and Bressan, Ray A and Hasegawa, Paul Mike, “Salt stress tolerance of plants”, JIRCAS working report, pp: 25-33, 2002.
- [10] Haque, SA, “Salinity problems and crop production in coastal regions of Bangladesh”, Pak J Bot, pp: 1359-1365, 2006.
- [11] He, Xinjian and Chen, Jianquan and Zhang, Zhigang and Zhang, Jinsong and Chen, Shouyi, “Identification of salt-stress responsive genes in rice (*Oryza sativa* L.) by cDNA array”, Science in China Series C: Life Sciences, Springer, pp: 477-484, 2002.
- [12] Eric C. Rouchka, “A Brief Overview of Gibbs Sampling”,



