

## Article:

<https://www.currentscience.ac.in/Volumes/117/03/0422.pdf>

## Abstract

HOX genes are a group of highly conserved genes in animals that dictates the animal's body plans. In this proposal we propose a project that aims to identify a highly conserved gene in the HOX gene, and analyze the gene for gene expression and similarities in animalia genome. The analysis would highlight differences between hox gene structure and expression allowing for future study of different organisms.

## Background/introduction

Genomic analysis is the identification, comparison of genetic features and their expression through the use of techniques such as DNA sequencing and bioinformatics. Genomic analysis is generally considered to be divided into two categories; structural genomics which identifies certain genomic structures in the genome and functional genomic analysis which analyses the expression of genes and their interactions often also called transcriptomes. Genomic analysis was made available through the prevalence and availability of gene sequencing. While the Human Genome Project sequenced the entire human genome most of the genomic analysis would not be possible without the further accessibility of sequencing using other techniques such as next-gen sequencing and whole-genome sequencing that does not depend on the isolation of cells in order to sequence the genes. Because these data that are generated from sequencing are generally put in online databases, these data are both available and able to be used for purposes that were not intended by the researchers who have originally read the sequence. The data set contains so many different information that a single sequencing can be the basis of several papers. Because of this, there is a data analysis bottleneck where there is so much data that needs to be analyzed but there is not enough time or computing power to analyze the entire genome has been a problem in the field.

Genomic analysis is the identification, comparison of genetic features and their expression through the use of techniques such as DNA sequencing and bioinformatics. Genomic analysis is generally considered to be divided into two categories; structural genomics which identifies certain genomic structures in the genome and functional genomic analysis which analyses the expression of genes and their interactions often also called transcriptomes. Genomic analysis was made available through the prevalence and availability of gene sequencing. While human genome project sequenced the entire human genome most of the genomic analysis would not be possible without the further accessibility of sequencing to the point where currently it is extremely affordable to have a genome sequenced, and other techniques such as next-gen sequencing and whole-genome sequencing that does not depend on the isolation of cells in order to sequence the genes. Because these data that are generated from sequencing are generally put online in databases, these data are both available and able used to be used for purposes that are not intended by the researchers who have originally made the sequence. The data set is so large that a single sequencing can be the basis of several papers. Because of this, there is a data analysis bottleneck where there is so much data that needs to be analyzed but there is not enough time or computing power to analyze the entire genome for all the information that it can yield. Because there are so many different venues that uses bioinformatics, there are a major need for people who are able to analyze the data that is available and easily accessible. Degrees in Bioinformatics is quickly becoming one of the highest-paid and available jobs on the market.

#### Main Idea:

The proposal in the article above highlights the analysis of a protein domain (BOP1) in many species, and creating a phylogenetic tree on the basis of taxonomy using program MEGA7. The co-expression of BOP1 protein was then mapped using a heat map. Absolute expression values

for BOP1 genes in normal and cancerous tissues were taken from the GENT database. These values between normal and cancerous tissues were then compared.

### Methods:

To choose what protein to analyze, we limited our decision to use highly conserved genes. This way, all species can be taken into account. The gene we decided to analyze is the **Hox gene**. This gene regulates the patterns of development in species.

- The Hox gene limits our subjects to organisms that belong in the Animalia family.

To analyze the Hox gene across different species:

- 1) Obtain homeodomain protein sequence in hox genes using CDART
- 2) Use MEGA7 to create a phylogenetic tree
- 3) Find protein-protein interaction and co-expression in animals with homeodomain protein using STRING.
  - a) Our reference organism will be homeosapiens
- 4) Create a heat map showing co-expression across the animalia family using R.
- 5) Make conclusions about our data.

\*specify domains of the Hox gene each group should study

\*specify conclusions each group should make

Specific Aims (300 words, 2 paragraphs):

Specific Aim 1: In preliminary work, we found that highly conserved genes are typically present in ribosomal DNA. HOX genes, on the other hand, were discovered relatively later and has less research done in comparison. We will now look into HOX genes to see if its expression is similar or different in Animalia species.

Specific Aim 2: The limited use of the Animalia family as our subjects will allow for easier comparison of the HOX gene, which is different from species to species. Characterization methods of our data include a phylogenetic tree and a heat graph of gene expression.

**Impact (500 words):**

Why studying the similarities and differences between hox genes is important. And why their expression is important. In the development of animals, there is a special period of embryonic similarity, during which not only the appearance of the embryos but also the expression of their genes is very conservative. The existence of this period is associated with the appearance in the evolution of stable building plans. In hox genes, the genes themselves are located on the chromosomes in order, therefore, strict, consistent activation of them is necessary in order for the body to form correctly. However, the mechanism and regulation of the Hox genes themselves remain unclear. It is clear that this process is very complex and multi-staged; non-coding RNAs are involved in it. To a complete understanding of all these processes in different animals, science is still very far away. When the basic mechanisms of gene regulation during early ontogenesis are deciphered, biologists will be able to closely address the following fundamental question, namely: why is all this necessary? Why do multicellular animals suddenly switch from maternal transcripts to their own? Incidentally, this problem has not only theoretical but also practical significance. The difficulties encountered in animal cloning are largely determined by the fact that we still know too little about the work of genes during early development. Cloning of animals is carried out by transplanting a nucleus from a somatic (non-sexual) cell of one animal into the egg of another. However, for normal development, it is necessary that the genes of the embryo are initially silent, and they are quite active in the donor nucleus. The egg cytoplasm must somehow reprogram the nucleus, turn off the genome, and then turn it on again at the right time. Until we know how to help her with this, it's hard to expect great success in cloning. Understanding the fact that all animals, from scolopendra to elephants, are formed at the direction of very similar blocks of Hox genes, greatly changes the picture of the world. Researchers, now armed with modern gene editing tools, including the CRISPR-cas9 system, can

begin to study various Hox genes in various animals. Such manipulations with animals can lead answering important questions on the evolution of various animals.