European Journal of Biotechnology and Bioscience
ISSN: 2321-9122; Impact Factor: RJIF 5.44
Received: 17-11-2018; Accepted: 20-12-2018
www.biosciencejournals.com
Volume 7; Issue 1; January 2019; Page No. 20-21

Identification of miRNA in Pan Troglodytes as genomics approach

Esha Dogra[1*], Prashant Singh[2]
[1, 2] Department of Biotechnology, Lovely Professional University, Phagwara, Punjab, India

Abstract
The last decade has witnessed huge effort in the understanding of the various mechanisms by which small non coding RNAs regulate the various life processes associated with a cell. Such efforts have led to the identification of multiple small non coding RNAs such as miRNA, siRNA, tasiRNA, piRNA, rasiRNA and many more of them microRNAs with their ubiquitous presence in all domains of life has become the molecule of choice of many research efforts and mechanisms involving their biogenesis and final function in the cell has been elucidated through these works. However, still certain grey areas exist in our understanding regarding the phenomenon of regulation of these non-coding wonders. As we all are aware that most microRNAs are differentially regulated during the entire lifespan of an organism, it becomes very clear that certain regulatory proteins and their interacting partners play important role in the process. With this background this work was performed where the precursor sequences of microRNAs of *Pan troglodytes* were considered as query and through position specific weight matrix evaluation and subsequent validation a large number of RNA regulatory elements were identified in those sequences.

Keywords: microRNA, precursor sequence, *Pan troglodytes*, homo sapiens

Introduction
RNA is a nucleic acid polymer which consists of nucleotide as monomers which allow RNA to encode genetic information. RNA is mostly single stranded molecule which folds to form secondary structure RNA. According to the central dogma of molecular biology RNA acts as a messenger between DNA and the protein synthesis. However this central dogma is getting challenged by the recent finding that tiny fragments i.e. the non coding RNA microRNA (miRNA) of 19-25 nucleotides in length present in the genomes of plants and animals are able to negatively regulate protein-coding genes by interfering with mRNA's original instructions Zhang *et al*. (2005) [4]. miRNAs are the highly conserved regions which suppresses the gene expression by imperfect base pairing to the 3′untranslated region (UTR) of target mRNAs which leads to the repression of protein production or mRNA degradation Bentwich *et al*. (2005) [1]. miRNA has several regulatory motifs which regulate the formation of the secondary structure by playing an essential role in transcriptional and post-transcriptional regulation of gene expression. These regulatory motifs activate under proper condition when the appropriate regulatory factor binds with it.

MicroRNAs are related to, but distinct from, short interfering RNAs (siRNAs). A key difference between siRNA and microRNA is that siRNA requires almost complete complementary to its targeting sequence for it to exert the silencing function, whereas a microRNA usually binds to its target genes through partial complementary binding. Because of this unique feature, a single microRNA has multiple target genes and, thus, could regulate a large number of protein-coding genes. This may explain why microRNAs play a fundamental role in regulation of diverse cellular processes. Increasing efforts to identify specific targets of microRNAs have led to the speculation that microRNAs may regulate at least 30% of human protein encoding genes. The first miRNA, lin-4, was identified by Lee *et al.* (1993) [3] in a genetic screen for mutants that disrupt the timing of post-embryonic development in *Caenorhabditis elegans* Lee RC and Ambros V (2001) [2]. The rapid progress in genome sequencing demands more comparative analysis to gain new insights into evolutionary, biochemical, genetic, metabolic, and physiological pathways. Comparative genomics is the direct comparison of complete genetic material of one organism against that of another to gain a better understanding of how species evolved and to determine the function of genes and noncoding regions in genomes. It includes a comparison of gene number, gene content, and gene location, the length and number of coding regions (called exons) within genes, the amount of non coding DNA in each genome, and conserved regions maintained in both prokaryotic and eukaryotic groups of organisms. Comparative genomics not only can trace out the evolutionary relationship between organisms but also differences and similarities within and between species.

MicroRNAs (miRNAs) are small non-coding RNAs, 19–24 nt long, which play a crucial regulatory role by inhibiting the translation of protein-coding mRNAs in various eukaryotic organisms. A miRNAs is processed from a longer transcript, referred to as the primary transcript (pri-miRNA). miRNAs can be located within the introns of protein-coding genes, outside of protein-coding genes entirely ('intergenic') or more rarely in a coding exon, untranslated region (UTR) or exon of a non-coding transcript. These tiny miRNAs inhibit the translation of a mRNA into protein through imperfect base pairing to one or more target sequences in the mRNA. The identification of animal miRNA targets is a challenging assignment for both experimental and computational groups.

Materials and Methods
Formulation of Hypotheses
1. Identification of the miRNAs of *Pan troglodytes* using sequence similarity approach by NCBI eukaryotic genomic BLAST with respect to the miRNAs of *Homo*

sapiens.
2. Formation of threshold for the prediction of miRNAs of *Pan troglodytes* after sequence similarity from already known miRNAs of *Pan troglodytes*.
3. Identification of the miRNAs by considering the flanking regions of the miRNAs having less similarity than that of the cut off.

Methodology

There are 1048 miRNAs of human present in miRBase database. So to find out the sequence similarities between human and chimpanzee various steps were followed. Collection of all the 1048 pre-miRNAs sequence of *Homo sapiens* from miRBase web server, Perform genomic BLASTn, threshold formation, flanking region study

Results and Discussion

Sequence similarity results

All the 1048 pre-miRNA fasta sequences of *Homo sapiens* were collected from miRBase web server and saved in a Microsoft word file. With each sequence NCBI genomic BLASTn was performed with respect to the genome of *Pan troglodytes* and saved the results in Microsoft excel sheet and analyzed the results by data filtering (Table 1 and 2).

Table 1: Data obtained by sequence similarity study of *Homo sapiens* with respect to *Pan troglodyte* (chimpanzee) genome

Maximum identity	No. of miRNAs
91%-100%	949
81%-90%	10
<=80%	0

Table 2: Data obtained by sequence similarity study of *Homo sapiens* with respect to *Pan troglodyte* (chimpanzee) genome considering query coverage

Maximum identity	Query coverage	No. of miRNAs
100%	100%	722
98%-99%	100%	2
95%-97%	100%	116
90%-99%	90%-99%	40
<90%	<90%	3
0%		89

Threshold formation

The 601 pre-miRNA fasta sequences of *Pan troglodytes* (chimpanzee) were collected from miRBase web server and saved in a Microsoft word file. With each sequence genomic BLASTn was performed with respect to the genome of *Homo sapiens* (human) and saved the results in Microsoft excel sheet and analyzed the results by data filtering (Table 3).

Table 3: Data obtained by sequence similarity study of *Pan troglodytes* with that of *Homo sapiens* genome

Maximum identity	Query coverage	No. of miRNAs
100%	100%	375
98%-99%	100%	125
95%-97%	100%	52
91%-94%	100%	2
85%-90%	100%	1
<85%		0
0%		33

Flanking region study

The flanking regions (left as well as right) of *Homo sapiens* and *Pan troglodytes* were identified for the sequences having <95% maximum identity analyzed in the first step. The length of the flanking sequence was taken to be 100bps nucleotide. It is known that the pre-miRNA sequence is 60-80bps in length; therefore the length of the flanking sequences for the flanking region similarity study was taken as 100bps nucleotides. The BLAST was then performed between the flanking sequence of human and chimpanzee (left as well as right) and the results were saved in a Microsoft excel sheet and analyzed by data filtering (Table 4)

Table 4: Data obtained for the flanking sequence similarity

Flanking Region	Maximum identity >80% and Query coverage >90%
Both left and right	22

Conclusion

The comparative genomics approach analysis helps us in the prediction of pre-miRNA in case of *Pan troglodytes.* Till now there are only 601 miRNAs in *Pan troglodytes* which are predicted according to the most reliable web server i.e. miRBase database. There are 1048 miRNAs present in *Homo sapiens*. There are possibilities of having more miRNAs in chimpanzee as it is the closest species to that of human. After performing sequence similarity of Human miRNA with that of chimpanzee genome 887 pre-miRNAs had been predicted which are similar to human miRNAs Similarly, flanking region study was done to predict further miRNA similarity and 22 more pre-miRNAs are identified. Therefore, 909 pre-miRNAs are predicted in chimpanzee. miRNA dysfunctioning is related to various diseases. So studying these miRNAs can help us to find proper treatment and valuable medication for trials. miRNA has been mostly conserved over varied number of species. Over 99% of MiRNA are conserved. This prediction will lead us to develop new thoughts for further prediction of unknown functions of miRNA. These miRNAs will further help in the evolutionary relationship among different organisms. miRNA research has been conducted for only a few years. There are still lots of unknown but exciting knowledge to be revealed about miRNAs.

References

1. Bentwich I, Avniel A, Karov Y, Aharonov R, Gilad S, Barad O, *et al*. Identification of hundreds of conserved and nonconserved human microRNAs. Nat. Genet. 2005; 37:766-770.
2. Lee RC, Ambros V. An extensive class of small RNAs in Caenorhabditis elegans. Scien. 2001; 294:862-864.
3. Lee RC, Feinbaum RL, Ambros V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell. 1993; 75:843-854.
4. Zhang LV, King OD, Wong SL, Goldberg DS, Tong AH, Lesage G, *et al*. Motifs, themes and thematic maps of an integrated Saccharomyces cerevisiae interaction network. J Biol. 2005; 4:6.