

PRTVBase: A Predicted promoter data base of Retro transcribing virus

RAGHUNATH SATPATHY⁺
*Department of Biotechnology,
MIRC Laboratory, MITS Engineering College,
Rayagada, Odisha, India-765017*

ANSHUMAN MAHAPATRA
*Department of IT,
MIRC Laboratory, MITS Engineering College,
Rayagada, Odisha, India-765017*

RASHMIRANJAN BEHERA
*Department of Biotechnology,
MIRC Laboratory, MITS Engineering College,
Rayagada, Odisha, India-765017*

RAJESH KUMAR GURU
*Department of Biotechnology,
MIRC Laboratory, MITS Engineering College,
Rayagada, Odisha, India-765017*

Abstract

Database having the information of viral promoters is highly required for molecular basis of virology study. In this work the data base of the predicted promoters of the all retro transcribing genome available in the NCBI has been constructed. The searching option of the gene in the data base includes both name of the virus and NCBI gene accession number and it provides the significant informations of various types promoters present, their strand position in the genome and specific nucleotide sequences etc. The data present in the data would be help the researchers for comparative analyse, to perform cloning study also gene regulation analysis of various deadliest retro transcribing virus. The database is currently managed by MIRC Laboratory of MITS Engineering College, Rayagada and freely accessed through a web interface (<http://www.mcb.i.gp/database.html>) which provides flexible and reliable searching options of retro transcribing virus promoters.

Keywords: PRTVBase; retro transcribing virus; promoter scan; NCBI; Promoter database.

1. Introduction

The members of the retro viruses belong to the family retroviridae are basically RNA viuses which replicates and then incorporated to genome of host by the reverse transcriptase and integrase enzymes [1]. In case of human cells have been invaded and co-inhabited by many retroviruses through the millennia; indeed, it is believed that 5 – 8% of the human genome is composed of so-called endogenous retroviruses [2]. The virions of retro viruses is about 100 nm in diameter and consists of two identical strand of RNA molecule each consists of 7-10 kilo bases long [3]. Retroviruses are a diverse group of RNA-containing viruses that contribute to several human deadliest diseases is caused by the members. The common diseases are different types of Cancers, AIDS (HIV) and T-cell leukemias, lymphomas, and spastic paraparesis (HTLV-1) etc [4-5]. By keeping view of the above stated diseases it is important to analyse the molecular basis genetic regulation pattern of the virus. So construction of data bases of accurate and comprehensive predicted of promoters of genes in case of retro transcribing virus is a great interest which leads to study their regulation pattern of the gene [6]. There is total 2567 genome records in NCBI from which 105 records are available for retro transcribing virus [7]. Many promoter prediction tools are available which are used frequently for finding promoters in case of viral genes with significant accuracy. So it is possible to construct a data base of predicted promoters of all retro transcribing virus genome available in NCBI.

2. Materials and methods

The genome records of retro transcribing virus were obtained from NCBI Virus genome data base further it was processed through the Promoter Scan tool [8]. Promoter Scan finds a putative promoter sequence, reports the sequence range in which the putative promoter is found. It then reports if a TATA box was found, and if so makes an estimate of the Transcription Start Site (TSS) position from the TATA position. The promoter and

non-promoter sequences are analyzed for the comparative density of each unique mammalian transcription factor binding site listed in the Transcription Factor Database. The combined individual density ratio of all binding sites is then built in a Promoter Recognition Profile. This profile, used in combination with a weighted matrix for scoring a TATA box, which is used by the program Promoter Scan to test the prediction of promoter sequences and the ability of the computer program to discriminate them from non-promoter sequences [9]. For the GUI interface the front end of the data base was constructed by HTML by using Microsoft Windows operating system. In the development and the back end for table creation for the data base was done by MySQL. The required connection was established by php code. The results obtained after processed by Promoter Scan tool were carefully examined and used as a record for data base construction. The basic arctecterial schema for construction of the data bases is given below figure 1.

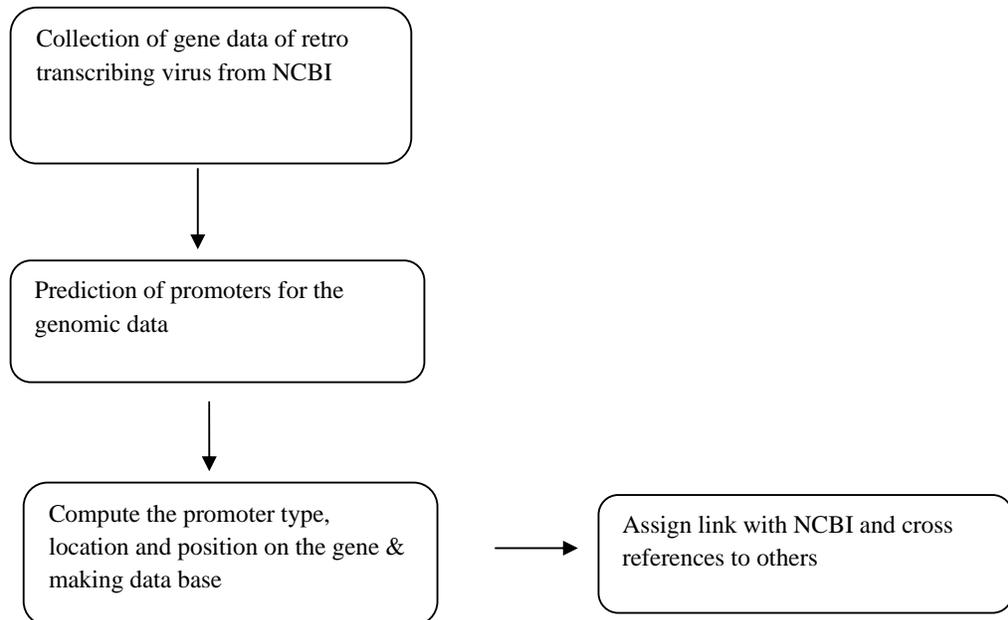


Fig.1: Showing the basic method for construction of PRTVBase

3. Results and discussions

The database having a web-based, flat-file type user interface with simple global search, specific database search, keywords help and with links to references in other external databases as shown in figure 2 and 3. The record entry contains the following information like name of the virus, Genbank accession number along with link to NCBI, GC content, Topology promoter information like type of promoter, sequences, location of gene on which it is present on gene as well as the DNA strand type, which can be accessed as result that has been shown in figure 4.



VIRUS NAME :

 GEN BANK ACC. NO. :

Fig. 2: Snap shot of search page of the data base



VIRUS NAME	GENBANK ACC. NO.	GC CONTENT	TOPOLOGY	MOLECULE	PROMOTER
Simian immunodeficiency virus SIV-mnd 2	AF367411	43%	linear	dsDNA	View

Fig. 3: Showing preliminary searching page of the NCBI gene accession number 'AF367411' or after searching" Simian immunodeficiency virus SIV-mnd2"by name.

VIRUS NAME	GENBANK ACC. NO	GC CONTENT	TOPOLOGY	MOLECULE
<i>Simian immunodeficiency virus SIV-mnd 2</i>	AF367411	43%	linear	dsDNA

Processed Sequence: 9518 Base Pairs
 Promoter region predicted on forward strand in 8107 to 8357
 TATA found at 8329, Est.TSS = 8359

Name	Sequence	Strand	Location
Sp1	CCCGCC	+	8107
GCF	SCGSSSC	-	8111
Sp1	GGCGGG	-	8112
T-Ag	GGGGC	+	8161
element_II_rs-3	TTTGCC	-	8209
TFIID	TATAWAW	+	8330
CTF	GATTGG	+	8355

Fig.4: Showing the promoter prediction (result page) for the above virus.

This data base is freely available which provides supplementary and useful information about promoter sequences. The database presently contains the promoter data for all 105 genome sequences that is available in NCBI and being updated according to sequence deposits in NCBI. The database is also useful for the scientific community and industries for a quick and informative access of promoter data of retro transcribing virus.

4. Conclusion

Databases constructions for the promoters are highly essential that provides information to study the gene regulation for aiming for a better treatment method. PRTVBase has been designed to manage and to explore the promoter data analysis. At present PRTVBase contains the basic predicted promoter data of different strains of retro transcribing virus. The data base has been developed by keeping pace with the progress of the availability retro transcribing virus genome sequences. User can search either by name of particular strain and also by the NCBI accession number of the viral genome. This database facilitates the retrieval of genomic data, promoter content of the gene, strand position for the location of promoter etc. This platform can be very useful for research and development sectors for the purpose of information sharing among various scientific communities.

5. Acknowledgment

We are thankful to Chief Executive of MITS Engineering College, Rayagada for providing us MIRC lab for computing facility.

References

- [1] Marcelle, W.; Wilhelm FX. (2006) Cooperation between Reverse Transcriptase and Integrase during Reverse Transcription and Formation of the Preintegrative Complex of Ty1. *Eukaryot Cell*, **5**, pp. 1760–1769.
- [2] Klaus, B.; *et al.* (2008) Human endogenous retrovirus HERV-K113 is capable of producing intact viral particles. *Journal of General Virology*, **89**, pp.567–572.
- [3] Christopher, S.; *et al.* (2006) Structure of an RNA switch that enforces stringent retroviral genomic RNA dimerization. *PNAS*, **103**, pp.13640-13645.
- [4] Robin, A.;(2001) Weiss Retroviruses and cancer. *Current Science* , **81**, pp. 528-534
- [5] Urnovitz, HB. Murphy, WH.; (1996) Human endogenous retroviruses: nature, occurrence, and clinical implications in human disease. *Clin Microbial Rev.*, **9**, pp.72-99.
- [6] Jessica, S.; *et al.* (2009) FANTOM4 EdgeExpressDB: an integrated database of promoters, genes, microRNAs, expression dynamics and regulatory interactions, *Genome Biol.*, **10**, R39.
- [7] <http://www.ncbi.nlm.nih.gov/genomes/GenomesHome.cgi?taxid=10239>
- [8] Pallavi, S.; *et al.* (2008) Identification and analysis of putative promoter motifs in *Flavivirus* genome .*Bioinformatics*, **3**, pp.162-167.
- [9] Prestridge, DS; (1995) Predicting Pol II promoter sequences using transcription factor binding sites. *J Mol Biol.*, **249**, pp.923-32.