Vegetable Science (2017) 44 (2) : 1-7

LTR retroelement in genes related to abiotic stress in *Capsicum annuum* L.

Shailesh K Tiwari[1±*], Yogesh Srivastava[1,2 ±], Vinay Kumar Singh[3], Major Singh[1] and B Singh[1]

Received: July 2017 / Accepted: November 2017

Abstract

Long terminal repeats (LTRs) are a type of retrotransposons possessing all commonly found structural features of long interspersed element (LINEs), including RNA polymerase III promoter, polyA tail, and flanking repeats. Here we report identification and characterization of LTR element in abiotic stress genes of solanaceae family along with an attempt to explain their functional importance through their secondary structure. We retrieved 366 abiotic stress genes of 5 species of solanaceae family– *Solanum lycopersicum, S. melongena, S. tuberosum, Nicotiana tabacum* and *Capsicum annuum* from NCBI nucleotide Database. *C. annuum*, accession HI543260.1 with the gene Id gi|311377312 possesses the LTR region. The gene reportedly has role in stress tolerance and high yield production. In gene masking, 16 simple repeats were detected, which best masked 569 bp out of 308042 bp, however, in protein masking 1 LTR (Ty1/copia) of 83 bp, 17 simple repeats of 603 bp and 15 low complexity regions of 514 bp were detected when masked 1200 bp out of 308042 bp. Secondary structure of the LTR region obtained, has low thermodynamic energy i.e. -12.40 kcal/mol and for the whole gene sequence it is -208.31 kcal/mol, denoting stability and its role in gene therapy and RNA processing. The findings shall be useful in gene delivery process, genome evolution, TE regulation, transposition mechanism, transgenesis, gene therapy and understanding of its biological structure and functions.

Keywords: Gene therapy; Heat stress; LTRs, Retroelements, RNA Processing; Solanaceae; Transposons

[1]Division of Crop Improvement, ICAR-Indian Institute of Vegetable Research, Jakhini (Sahanshahpur), Varanasi-221305, Uttar Pradesh, India
[2]Research Scholar, Genome Regulation Laboratory, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, 190 Kaiyuan Road, Guangzhou Science Park, Luogang District, Guangzhou 510530, China
[3]Centre for Bioinformatics, School of Biotechnology, Banaras Hindu University, Varanasi-221005, Uttar Pradesh, India
± Equal contribution in the research article
*Corresponding author, Email: tiwarishailu@gmail.com

Introduction

Bioinformatics and genomic approaches have become an integral part of large-scale studies on Transposable Elements (TEs) to extract information with pure *in silico* analyses or to assist wet lab experimental studies. TEs are mobile genetic elements forming major portion of most eukaryotic genomes and their dynamic presence continues to be a major force in diversify genomes by insertional mutagenesis (Kidwell and Lisch 2001). Computational analyses of TEs in genome sequences converge on filtering out "junk" sequences to make possible gene annotation, early effort on abundance and diversity of TEs in eukaryotic genomes transformed into the systematic genome-wide categorization and classification of TEs, (Janicki et al. 2011). RNA functions as an information carrier, catalyst and regulatory element, perhaps reflecting its importance in the earliest stages of evolution, (Baranov et al. 2005). Long terminal repeat (LTR) is a type of retrotransposons, that replicate through reverse transcription of a mRNA intermediate, (Hou et al. 2010). Because LTR retrotransposons and retroviruses are similar in their genetic organization and mechanism of replication, they are collectively referred to as retroelements, (Boeke and Stoye, 1997). The LTRs delimit retroelement insertions, and contain the promoters and transcription terminators required to produce a retroelement mRNA. Adjacent to the 52 and 32 end of LTRs are a primer binding site (PBS) and a polypurine tract (PPT), respectively, which serve as priming sites for reverse transcription. All retroelements have two genes in common: GAG encoding proteins that form virus or virus-like articles (VLPs) and POL encoding three enzymes, namely reverse transcriptase (RT), integrase (IN) and protease (PR), (Hou et al. 2010). The ability to catalyze cDNA integration makes retroelements a good tool for gene delivery (Anson 2004). Similar to their retrovirus relatives, retrotransposons rely on element-encoded reverse transcriptase to generate DNA copies of their own sequence from an RNA template.

The cDNA copy can subsequently be inserted by integrase into the host genome, (Tramontano et al. 2011).

Secondary structure of RNA is important to describe the function of RNA element. Structural RNAs are important as regulator, catalysts and structural component of cells. Their secondary structure must be known in order to understand the relationship between structure and function, (Engelen and Tahi 2007). The secondary structure is produced by the Watson-Crick pairings (AU and GC), Wobble pairing (GU), and non-canonical pairings, (Lescoute et. al. 2005). The concept of secondary structure was introduced by Doty et al. (1959). The structure is predicted using a loop-based energy model and the dynamic programming algorithm introduced by Zuker and Stiegler (1981). Prediction of the secondary structure of an RNA sequence is a classic problem of sequence analysis in bioinformatics, (Hamada et al. 2010). The importance of accurate predictions of secondary structures has increased due to the finding of functional non-coding RNAs whose functions are closely related to their secondary structures, (Griffiths-Jones et al. 2005). There are many tools and algorithms for secondary structure prediction like Mfold, RNAfold, RNAstructure, etc. (Ding et al. 2005) and prediction the minimum free energy (MFE) structure by using the Zuker's algorithm (Zuker and Stiegler 1981) being the most popular approach.

The current revolution in genome sequencing technology facilitates further progress in the existing frontiers of research and emergence of new initiatives. TEs can diversify gene regulation and potentially contribute to the evolution of gene expression, (Bureau and Wessler, 1992). A TE database can only keep pace with the ever increasing abundance of genome sequences updated with new information on TE families or super-families. New TEs are still being discovered using classical genetic approaches based on the mutagenic phenotypes caused by their insertions (Janicki et al. 2011). Among abiotic stress genes of Solanaceae family, the information on LTRs, Long Interspersed Nuclear Elements (LINEs) and Short Interspersed Nuclear Elements (SINEs) is very limited (Grandbastien et al. 2005; Casacuberta and González 2013; Alzohairy et al. 2014; Kriedt et al. 2014; Chénais et al. 2016). Here we predict, identify and characterize the LTR element along with its secondary structure in abiotic stress genes of Solanaceae family to describe its functional importance.

Methods

i. DATA Retrieval:

Solanaceae family is one of most important family of vegetables. Among 425 genes reportedly having role in abiotic stress available at NCBI nucleotide database we retrieved 366 genes from 5 species of Solanaceae family i.e. *Solanum lycopersicum, S. melongena, S. tuberosum, Nicotiana Tabacum* and *Capsicum annuum* [http://www.ncbi.nlm.nih.gov/nuccore/?term=solanaceae+(abiotic+stress)]

ii. Masking of Genomic data:

Sequence comparisons in RepeatMasker web server were performed by one of several popular search engines including, cross_match, ABBlast/WUBlast, RMBlast and Decypher (http://www.repeatmasker.org/). RepeatMasker compares the query sequence against one or more files of FASTA sequences. The sequences in the library provided by the RepeatMasker are consensus sequences derived from alignment of multiple copies of interspersed or satellite repeats (Kumari et al. 2013). For interspersed repeats, a consensus tends to approach the sequence of the transposable element from which the repeat is derived. Both cross match and WU-blast perform their Smith-Waterman (SW) alignments by first identifying exact word matches and restricting the alignment to a band or matrix surrounding the exact match(es) while merging the overlapping matrices. Generally, cross match and WU-blast report to the script only those matches which are less than 80-90% overlapped by a higher scoring match. The program accepts information only in FASTA format with errors and raw sequence files, but does not work with other formats like GenBank, Staden, etc. RepeatMasker makes a .masked file containing the query sequence(s) with all identified repeats and low complexity sequences masked. These masked sequences are listed and annotated in the .out file. The masked sequences are printed in the same order as they are in the input file, whereas the sequences are presented alphabetically in the annotation table. The .tbl file is a summary of the repeat content of the analyzed sequence. (http://www.animalgenome.org/bioinfo/resources/manuals/RepeatMasker.html)

iii. RNA Secondary Structure Prediction:

The Vienna RNA secondary structure server provides a web interface to the most frequently used functions of the Vienna RNA software package for the analysis of RNA secondary structures. It currently offers prediction of secondary structure from a single sequence, prediction of the consensus secondary structure for a set of aligned sequences and the design of sequences that will fold into a predefined structure. All three services can be accessed via the Vienna RNA web server

at http://rna.tbi.univie.ac.at. It provides programs, web services, and databases, including RNAfold, RNAcofold for the prediction of secondary structures of RNA. It predicts minimum free energy (MFE) structures and base pair probabilities from single RNA or DNA sequences and allowed predicting the structure of dimers that helps in prediction of function of retroelements on the basis of their comparative study (Hofacker 2003).

RNA fold server provides both the basic and most widely used function. Input consists of a single sequence that has to be typed or pasted into a text field of the input form. In the simplest case, the server predicts only the minimum free energy (mfe) structure of a single sequence using the classic algorithm of Zuker and Stiegler (1981). In addition to mfe folding the server calculates equilibrium base pairing probabilities via John McCaskill's partition function algorithm (McCaskill 1990). By default, the RNA energy parameters of the Mathews et al. (1999) are used, but single stranded DNA sequences can also be handled, by selecting the DNA parameter set (SantaLucia 1998). The fold server output consists of a static html page presenting the predicted mfe structure as a string in bracket notation and links to the plots generated for visualization. Three types of plots can be produced. Firstly, the predicted mfe structure is plotted as a conventional secondary structure graph using the naview layout method (SantaLucia 1998). Secondly, the pair probabilities can be visualized in a so-called 'dot plot': on a square grid of n_n we draw for each possible pair (i, j) a box with area proportional to its probability are produced. Finally, a mountain plot depicting both the predicted mfe and pair probabilities was obtained. A mountain plot is a xy-graph that plots the number of base pairs enclosing a sequence position (for pair probabilities the average number of enclosing pairs).

Results and Discussion

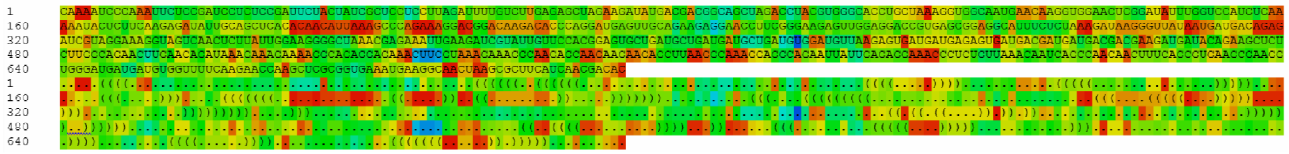
LTRs possess all the structural features commonly found in long interspersed elements including RNA polymerase III promoter, polyA tail, and flanking repeats. In gene masking the result are 16 simple repeats only it best masked 569 bp out of 308042 bp and in protein masking 1 LTR (Ty1/copia) of 83 bp, 17 simple repeats of 603 bp and 15 low complexity regions of 514 bp, it best masked 1200 bp out of 308042 bp. LTR region found in *C. annuum* responsible for polynucleotides, polypeptides encoded thereby, and methods of using same for increasing abiotic stress tolerance and/or biomass and/or yield in plants expressing the same (Kumar *et al*. 2012). Accession is HI543260.1 and the gene Id is gi|311377312 (Sequence 671 from Patent EP2183371 play role in stress tolerance and high yielding) other genes having only polyA tail, and flanking repeats. The secondary structure prediction result of LTR region and the concerning gene is showing the minimum free energy (MFE). Details of LTR/Copia sequence found in repeat masking were given in Table 2.

Here, 306 shows score of SW algorithm for the match found, 21.3% is substitutions in matching region compared to the consensus regions, 0.0 showing percentage of bases opposite a gap in the query sequences (deleted bp) 3.8 shows percentage of bases opposite a gap in the repeat consensus (inserted bp) Gi|311377312 shows gene identification number, 394 shows starting position of match in query sequence, 476 is ending position of match in query sequence, (238) is the number of bases in query sequence past the ending position of match, C is the match with complement of the consensus sequence in the database, ATCopia24I is the name of matching interspersed repeat element, LTR/Copia is the class of repeat, 1901 is the number of bases in (complement of) the repeat consensus sequence prior to beginning of match, 2806 is the starting position of

```
306   21.3  0.0  3.8  gi|311377312  394   476   (238) C ATCopia24I LTR/Copia  (1901)   2806   2727  28
ANNOTATION EVIDENCE:
306  21.28 0.00 3.75  gi|311377312  394   476    238 C ATCopia24I LTR/Copia        2727   2806   1901
  gi|311377312       394 TGCTGATGCTGATGATGCTGATGTGGATGTTAAGAGTGATGATGAGAGTG 443
                               v      v       v     vv    v i vii          ---
C ATCopia24I#LT     2806 TGATGATGCAGATGATGATGATGATGATGATGATGATGATGA---TG 2760
  gi|311377312       444 ATGACGATGATGACGACGAAGATGATACAGAAG 476
                                 i       i  i  v    i i   v
C ATCopia24I#LT     2759 ATGATGATGATGATGATGATGATGGTGCAGTAG 2727
Matrix = 20p43g.matrix
Transitions / transversions = 0.89 (8 / 9)
Gap_init rate = 0.01 (1 / 82), avg. gap size = 3.00 (3 / 1)
```

Sequence alignment by cross match, A. thaliana taken as model plant

i = Transition (G<->A, C<->T); v = Transversion (all other substitutions)



match in database sequence, 2727 is the ending position of the match in database sequence.

A brief summary of the 366 aligned sequences by RepeatMasker web server, information about retroelement and simple repeats were shown in Table 1. Details of aligned sequences by cross match program were shown in Supplementary file of Table 1.

Secondary structure prediction of Gene and LTR region

For Gene, gi|311377312,

Table 1: Brief Information about the aligned sequences by RepeatMasker web server

	Number of elements*	Length occupied	Percentage of sequence
Retroelements	1	83 bp	0.03
SINEs :	0	0 bp	0.00
Penelope	0	0 bp	0.00
LINEs:	0	0 bp	0.00
CRE/SLACS	0	0 bp	0.00
L2/CR1/Rex	0	0 bp	0.00
R1/LOA/Jockey	0	0 bp	0.00
R2/R4/NeSL	0	0 bp	0.00
RTE/Bov-B	0	0 bp	0.00
L1/CIN4	0	0 bp	0.00
LTR elements:	1	83 bp	0.03
BEL/Pao	0	0 bp	0.00
Ty1/Copia	1	83 bp	0.03
Gypsy/DIRS1	0	0 bp	0.00
Retroviral	0	0 bp	0.00
DNA transposons	0	0 bp	0.00
hobo-Activator	0	0 bp	0.00
Tc1-IS630-Pogo	0	0 bp	0.00
En-Spm	0	0 bp	0.00
MuDR-IS905	0	0 bp	0.00
PiggyBac	0	0 bp	0.00
Tourist/Harbinger	0	0 bp	0.00
Other (Mirage, P-element, Transib)	0	0 bp	0.00
Rolling-circles	0	0 bp	0.00
Unclassified:	0	0 bp	0.00
Total interspersed repeats:		83 bp	0.03
Small RNA:	0	0 bp	0.00
Satellites:	0	0 bp	0.00
Simple repeats:	17	603 bp	0.20
Low complexity:	15	514 bp	0.17

File name: RM2_solabioticstress.txt_1310374708; Sequences: 366; Bases masked: 1200 bp (0.39 %); Total length: 308042 bp (308042 bp excl N/X-runs); GC level: 42.89 %

* most repeats fragmented by insertions or deletions have been counted as one element. The query species was assumed to be Arabidopsis; RepeatMasker version open-3.3.0, default mode run with blastp version 3.0SE-AB [2009-10-30] [linux26-x64-I32LPF64 2009-10-30T17:06:09] **RepBase Update 20110419, RM database version 20110419**

Structures generated from Vienna Web Server, Figure 1A and 1B showing secondary structure of RNA on the basis of their positional entropy (range 0 to 2.5) and MFE (range 0 to 1) structure drawing encoding base-pair probabilities while The optimal secondary structure in dot-bracket notation with a minimum free energy of **-194.38** kcal/mol is given below **(Figure 1).**

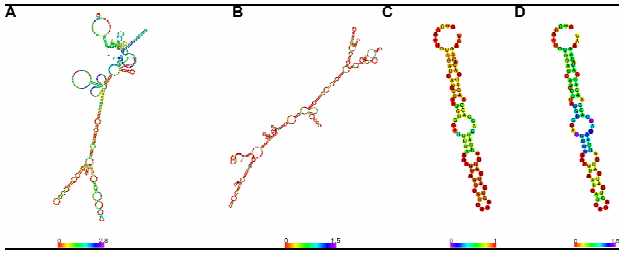


Figure 1: Secondary structure for gene **gi|311377312** and LTR region:
Structures generated from Vienna Web Server, A and B showing secondary structure of RNA on the basis of their positional entropy (range 0 to 2.5) and MFE (range 0 to 1) structure drawing encoding base-pair probabilities while C and D showing secondary structure of RNA on the basis of their positional entropy (range 0 to 2.5) and MFE (range 0 to 1) structure drawing encoding base-pair probabilities for LTR region found in the same gene.

Table 2: Details of LTR/Copia sequence found in repeat masking

Values	**Description**
306	Smith-Waterman score of the match, usually complexity adjusted. The SW scores are not always directly comparable. Sometimes complexity adjustment has been turned off, and a variety of Scoring-matrices are used.
21.3%	Substitutions in matching region compared to the consensus.
0.0	% of bases opposite a gap in the query sequence (deleted bp)
3.8	% of bases opposite a gap in the repeat consensus (inserted bp)
Gi|311377312	Gene Id of query sequence.
394	starting position of match in query sequence
476	ending position of match in query sequence
(238)	no. of bases in query sequence past the ending position of match
C	match is with the Complement of the consensus sequence in the database
ATCopia24I	name of the matching interspersed repeat
LTR/Copia	the class of the repeat
1901	no. of bases in (complement of) the repeat consensus sequence prior to beginning of the match
2806	starting position of match in database sequence
2727	ending position of match in database sequence

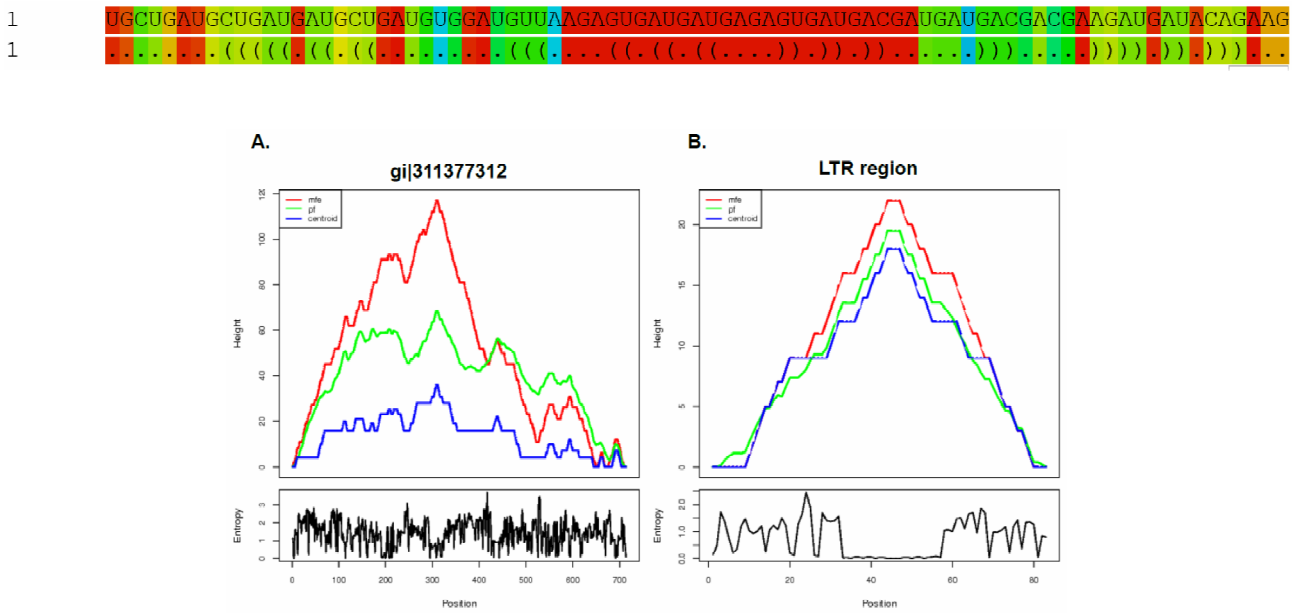


Figure 2: Mountain plots of MFE secondary structure of whole gene and LTR region. Color by positional entropy: (figure 2A). The free energy of the thermodynamic ensemble is -208.31 kcal/mol. The frequency of the MFE structure in the ensemble is 0.00 %. The ensemble diversity is 246.66.

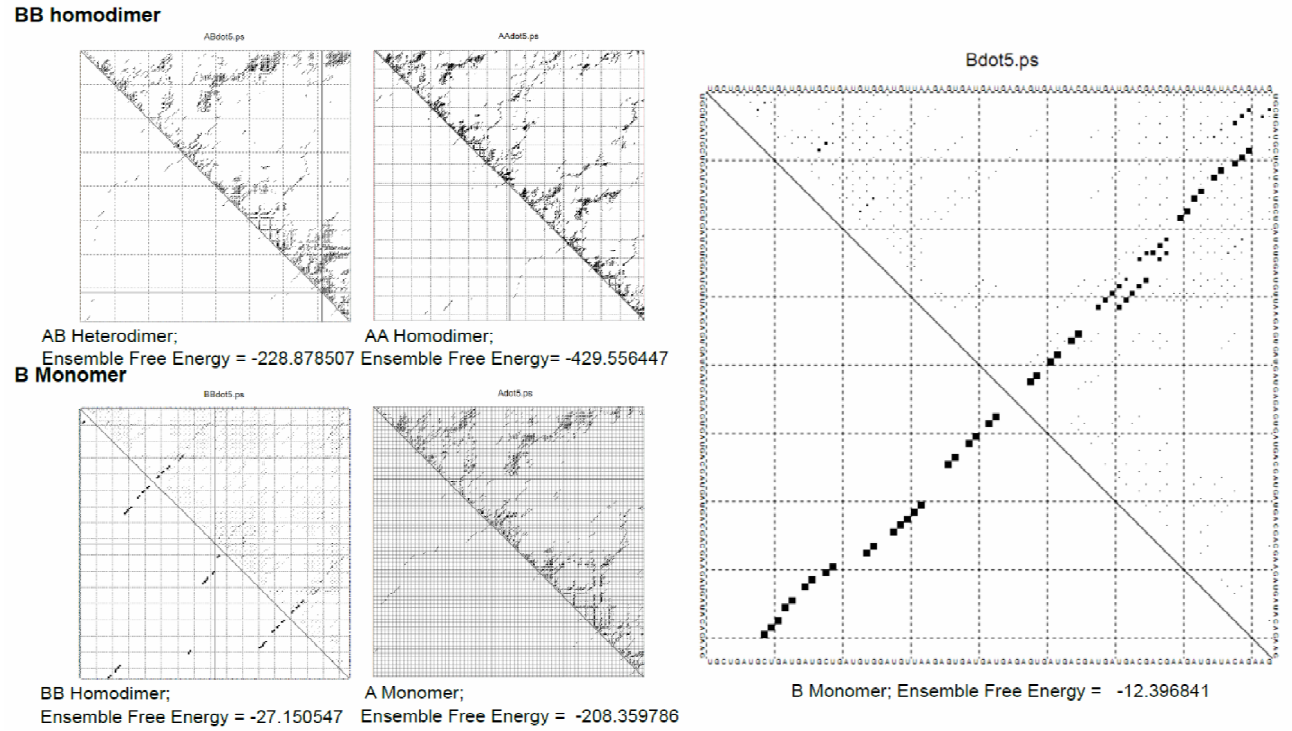


Figure 3: Dot plot showing the formation of homodimer and monomer with the minimum ensemble energy

The centroid secondary structure in dot-bracket notation with a minimum free energy of -89.29 kcal/mol is given below.

For LTR region:

Figure 1C and 1D showing secondary structure of RNA on the basis of their positional entropy (range 0 to 2.5) and MFE (range 0 to 1) structure drawing encoding base-pair probabilities for LTR region found in the same gene.

The free energy of the thermodynamic ensemble is -

12.40 kcal/mol. (Figure 3)

The frequency of the MFE structure in the ensemble is **6.35** %.

The ensemble diversity is **15.53.**

The centroid secondary structure in dot-bracket notation with a minimum free energy of **-6.54** kcal/mol.

***Color by positional entropy for LTR region*:**

Mountain plot representation of the minimum free energy structure. Thermodynamic ensemble of RNA structures and the centroid structure in gene (Figure 2A) and LTR region (Figure 2B). The lower panel of both A and B exhibit the positional entropy for each position.

The MFE structure of an RNA sequence is the secondary structure that contributes a minimum of free energy. This structure is predicted using a loop-based energy model and the dynamic programming algorithm introduced by Zuker *et al.*, 1981. RNA secondary structure can be uniquely decomposed into loops and external bases.

RNA cofold Result: The optimal secondary structure of the hetero-dimer in dot-bracket notation with a minimum free energy of -213.79 kcal/mol is given below "&" separates the two sequences and the two structures (Figure 1).

***Thermodynamic ensemble prediction*:** The free energy of the thermodynamic ensemble is -228.88 kcal/mol. The frequency of the MFE structure in the ensemble is 2.33227e-11 %. "G for heterodimer binding is -8.12 kcal/mol. "G (Gibb's free energy) is also the chemical potential that is minimized when a system reaches equilibrium at constant pressure and temperature, it shows heterodimer stability (Figure 3).

Conclusion

The importance of accurate predictions of secondary structures has increased due to the recent finding of functional non-coding RNAs whose functions are closely related to their secondary structure. The cross match output lines defined the LTR region in 366 abiotic stress genes of Solanaceae family (the gene accession HI543260.1 and the gene Id is gi|311377312 of *C. annuum)*. Secondary structure of the LTR region having very less thermodynamic energy *i.e.* -12.40 kcal/mol and for the whole gene sequence it is -208.31 kcal/mol. A large number of advances in the prediction of RNA secondary structure by free energy minimization have been reported (Mathew and Turner 2006). Recent experiments continue to reveal the sequence dependence of RNA loop stability (Wuchty 1999) indicating stability of the secondary structure in nature, its role in RNA processing (Katz and Burge 2003), and also it helps in gene delivery process, genome evolution, TE regulation, transposition mechanism, transgenesis, gene therapy (Janicki et al. 2011) and to the understanding of its biological structure and function as shown in Figure 4.

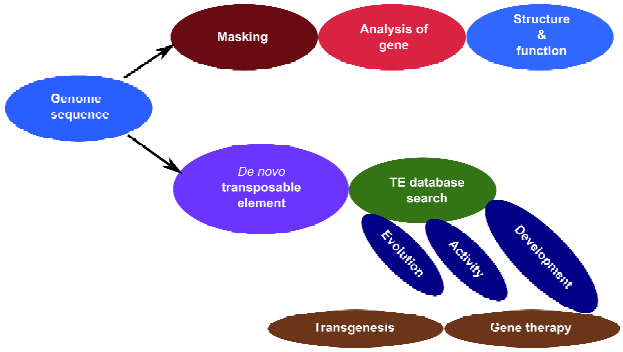


Figure 4: Flow-chart of a simplified field map of TE research in context of genomic research (Janicki et al. 2011).

Acknowledgement

The authors acknowledge the facilities provided by the Director, ICAR-Indian Institute of Vegetable Research, Varanasi for providing facilities to conduct the study.

सारांश

लांग टर्मिनल रिपिट, रिट्रोट्रॉस्पोसॉन का एक प्रकार है, जिसमें लम्बे अन्तःस्थापित तत्वों के सभी सामान्य लक्षण जैसे आर.एन.ए. पॉलिमरेस–III प्रमोटर, पॉली ए टेल और फ्लांकिंग रिपिट पायी जाती है। यहां सोलेनेसी कुल के अजैविक प्रतिबल के जीन में एक एल. टी.आर. तत्व की पहचान और लक्षण के वर्णन के साथ उसकी माध्यमिक संरचना से उसके कार्यात्मक महत्व की व्याख्या करने का प्रयास किया गया है। एनसीबीआई न्यूक्लियोटाइड डाटाबेस के सोलेनेसी कुल के पाँच प्रजातियों जैसे *सोलेनम लाइकोपर्सिकम, एस. मेलांनजेना एस. ट्यूबरोसम, निकोटीयाना टैबैकम* और *कैप्सिकम एनम* की 366 अजैविक प्रतिबल जीन को पुनः प्राप्त किया। *सी. एनम* के प्रभेद एचआई 543260.1 अभिलेख के जीन आईडी जीआई 311377312 में एलटीआर क्षेत्र उपस्थित है। जीन की कथित तौर पर प्रतिबल सहिश्णुता और उच्च उत्पादन में भूमिका है। जीन मास्किंग में, 16 सिम्पल रिपिट्स का पता लगाया गया, जो 308042 बीपी में से 569 बीपी को सबसे अच्छी प्रकार मास्क करते हैं, जबकि प्रोटीन मास्किंग के दौरान जेबी 308042 बीपी में से 1200 बीपी को मास्क किया गया, तब 83 बीपी का 1 एलटीआर (टीवाई1/कोपिया), 603 बीपी के 17 सरल दोहरता क्षेत्र 514 बीपी के 15 अल्प जटिलता के क्षेत्रों का पता चला। प्राप्त एलटीआर क्षेत्र की माध्यमिक संरचना में कम ताप विद्युत ऊर्जा (–12.40 किलो कैलोरी/मोल) है, और पूरे जीन अनुक्रम के लिए यह –208.31 किलो कैलोरी प्रति मोल है, जो स्थिरता, जीन थेरेपी और आरएनए प्रसंस्करण में इसकी भूमिका को दर्शाता है। जीन डिलीवरी प्रक्रिया, जीनोम विकास, टीइ विनियमन, पारस्परिक तंत्र, ट्रांसजेनेसिस, जीन थेरेपी और इसके जैविक संरचना और कार्यों की समझ हेतु में निष्कर्ष उपयोगी होंगे।

Reference

Anson DS (2004) The use of retroviral vectors for gene therapy-what are the risks? A review of retroviral pathogenesis and its relevance to retroviral vector-mediated gene delivery. Genet Vaccines Ther 2: 9.

Alzohairy AM, Sabir JSM, Gyulai G, Younis RAA, Jansen RK and Bahieldin A (2014) Environmental stress activation of plant long-terminal repeat retrotransposons. *Fu*nctional Plant Biology 41: 557-567.

Baranov PV, Hammer A W, Zhou J, Gesteland R F and Atkins J F (2005) Transcriptional slippage in bacteria: distribution in sequenced genomes and utilization in IS element gene expression. Genome Biol 6: R25.

Boeke JD and Stoye JP (1997) Retrotransposons. endogenous retroviruses, and the evolution of retroelements. Cold Spring Harbor: Cold Spring Harbor Laboratory Press, 343-436.

Bureau TE and Wessler SR (1992) Tourist: a large family of small inverted repeat elements frequently associated with maize genes. The Plant cell 4: 1283-1294.

Casacuberta E and González J (2013) The impact of transposable elements in environmental adaptation. Molecular Ecology 22: 1503–1517.

Chénais B, Caruso A, Hiard S and Casse N. (2016) The impact of transposable elements on eukaryotic genomes: From genome size increase to genetic adaptation to stressful environments. Gene 509: 7–15.

Ding Y, Chan CY and Lawrence CE (2004) Sfold web server for statistical folding and rational designing of nucleic acid. Nucleic Acids Res 135-141.

Doty P, Boedtker H, Fresco J R, Haselkorn R and Litt M (1959) Secondary Structure in Ribonucleic Acids. Proc Natl Acad Sci USA 45: 482-499.

Engelen S and Tahi F (2007) Predicting RNA secondary structure by the comparative approach: How to select the homologous sequences. BMC Bioinformatics 8: 464.

Grandbastien MA, Audeon C, Bonnivard E, Casacuberta JM, Chalhoub B, Costa APP, Le QH, Melayah D, Petit M, Poncet C, Tam SM, van Sluys MA and Mhiri C (2005) Stress activation and genomic impact of Tnt1 retrotransposons in Solanaceae. Cytogenet Genome Res 110: 229–241.

Griffiths-Jones S, Moxon S, Marshall M, Khanna A, Eddy SR and Bateman A (2005) Rfam: annotating non-coding RNAs in complete genomes. Nucleic Acids Res 121-124.

Hamada M, Sato K and Asai K (2010) Prediction of RNA secondary structure by maximizing pseudo-expected accuracy. BMC Bioinformatics 11: 586.

Hofacker IL (2003) Vienna RNA secondary structure server. Nucleic Acids Res, 31, 3429-3431.

Hou Y, Rajagopal J, Irwin PA and Voytas DF (2010) Retrotransposon vectors for gene delivery in plants. Mob DNA 1: 19.

Janicki M, Rooke R and Yang G (2011) Bioinformatics and genomic analysis of transposable elements in eukaryotic genomes. Chromosome Res 19: 787-808.

Katz L and Burge CB (2003) Wide spread selection for local RNA secondary structure in coding regions of bacterial genes. Genome Res 13: 2042-2051.

Kidwell MG and Lisch DR (2001) Perspective: transposable elements, parasitic DNA, and genome evolution. Evolution 55: 1-24.

Kriedt RA, Cruz GMQ, Bonatto SL and Freitas LB (2014) Novel Transposable Elements in Solanaceae: Evolutionary Relationships among Tnt1-related Sequences in Wild *Petunia* Species. Plant Molecular Biology Reporter 32: 142–152.

Kumar R, Solanky SS and Singh M (2012) Breeding for drought tolerance in vegetables. Vegetable Science 39: 1-15

Kumari A, Kumar A, Wany A, Prajapati GK and Pandey DM (2013) Identification and annotation of abiotic stress responsive candidate genes in peanut ESTs. Bioinformation 8: 1211-1219.

Lescoute A, Leontis NB, Massire C and Westhof E (2005) Recurrent structural RNA motifs, Isostericity Matrices and sequence alignments. Nucleic Acids Res 33: 2395-2409.

Mathews DH, Sabina J, Zuker M and Turner DH (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J Mol Biol 288: 911-940.

Mathews DH and Turner DH (2006) Prediction of RNA secondary structure by free energy minimization. Curr Opin Struct Biol 16: 270-278.

McCaskill JS (1990) The equilibrium partition function and base pair binding probabilities for RNA secondary structure. Biopolymer 29: 1105-1119.

SantaLucia J Jr. (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. Proc Natl Acad Sci USA 95: 1460-1465.

Tramontano A, Donath A, Bernhart SH, Reiche K, Bohmdorfer G, Stadler PF and Bachmair A (2011) Deletion analysis of the 3' long terminal repeat sequence of plant retrotransposon Tto1 identifies 125 base pairs redundancy as sufficient for first strand transfer. Virology 412: 75-82.

Wuchty S, Fontana W, Hofacker IL and Schuster P (1999) Complete suboptimal folding of RNA and the stability of secondary structures. Biopolymers 49: 145-165.

Zuker M and Stiegler P (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res 9: 133-148.