PtRNASS: Prediction of tRNA Secondary Structure from Nucleotide Sequences

Li-Yeh Chuang, Yu-Da Lin, and Cheng-Hong Yang, *Member, IAENG*

Abstract—**tRNA is an important small molecule that was preserved throughout evolution. It plays a central role in the molecular translation process. All tRNAs have a characteristic structure which resembles cloverleaves and lengths within 63-200 bases. The same anticodon of tRNAs from orthologous species are usually folded into a highly similar secondary structure. Hence, tRNAs have been extensively discussed in research of molecule evolution. Many reports indicate the important possibility that the structure of tRNAs lead biologists to understand the evolution process. In recent research, it shows that the method is to predict the secondary structure of tRNA. In this method, the covariance model (CM) helps to bring the advantage of its accuracy. Many current available functions are used in the prediction of secondary structure. However, it is not yet satisfied the most of biologists. In some cases, the predictions of some tRNA genes are impossible to perform with the available methods. We propose a novel method to predict tRNA secondary structures. This helps to achieve a detection sensitivity of 99.77% from the Sprinzl database within the species of Archaea, Bacteria, and Eukaryota. Therefore, as the result, it is the best prediction for the secondary structure of tRNA.**

Index Terms—**tRNA, secondary structure, evolutionary.**

I. Introduction

The studies of non-coding RNAs are very important to search the function or roles in cells. In order to understand the function, we must find the secondary structure. The family of tRNAs is a type of RNA molecules, it has the special function to translate amino acids into protein-building machinery. In addition, the amino acids concatenate simultaneously through ribosome to form protein. Each tRNA molecule is able to recognize the codons triplet from mRNA, and then tRNA carries out the respective amino acid to the protein-building machinery. In order to add amino acid successfully, the tRNA has to read the coded segment accurately from mRNA. Hence, the prediction of anticodon of tRNA becomes an important subject for research. Furthermore, both of the characteristics of central roles are played by tRNA to sustain every vital task in a cell. tRNA's short sequence length makes it a popular tool in the field of research. In recent reports, some suggested that the conserved structure in tRNA involves the evolutional origin.

A standard secondary structure of tRNA molecule takes the form of a cloverleaf to comprise four stacked pairs (stem structure), four hairpin loops, one multi-loop and three spacer bases. The determination of secondary structure is folded by the stable structure and the stable structure must be the one that contains the pair amount of hydrogen bonds (i.e., G-U, A=U and G≡C).

There are many tools providing the method of prediction of tRNA secondary structure, e.g. tRNAscan-SE [1], ARAGORN [2] and tRNAfinder [3]. The tRNAscan-SE tool features the most sensitive prediction result, which is composed by three algorithms: (I) tRNAscan 1.4 conservatively calls ambiguous nucleotides as always forming base and the highest scoring choice in consensus promoter matrices rules. (II) EufindtRNA, searches the four tRNA features that are the nucleotide composition of the A box, the nucleotide composition of the B box, the nucleotide distance between the boxes of A and B, and the distance between the B boxes and RNA polymerase III termination signals to identify tRNA location. (III) Covariance models are probabilistic representation of a typical tRNA secondary structure and primary sequence consensus. This method provides the reliable sensitivity and selectivity of the prediction. The ARAGORN algorithm provides faster tRNA gene detection through utilizing consensus sequence as the search model and it offers the capability for prediction of tRNA secondary structure. The consensus sequences are built by their tools; most depend on segments within boxes A and B. Although many tRNAs have highly conserved consensus sequences, the use of this model causes the failure in predicting unusual tRNA genes. However, most of them cannot satisfy the users' demands. According to the above cases, we adjusted the prediction of tRNA secondary structure method. We considered four major factors, they are the structure of hydrogen bonds, the GC%, characteristic, and the existence of introns to provide a simple way of performance for the search of tRNA gene and predict the secondary structure.

L.Y. Chuang is with the Department of Chemical Engineering, I-Shou University , Kaohsiung 84001, Taiwan (E-mail: chuang@isu.edu.tw)
Y.D. Lin is with the Department of Electronic Engineering, National Kaohsiung University of Applied Sciences, Kaohsiung 80778, Taiwan (E-mail: e0955767257@yahoo.com.tw)
C. H. Yang is with the Department of Electronic Engineering, National Kaohsiung University of Applied Sciences, Kaohsiung 80778, Taiwan and Department of Network Systems, Toko University, Chiayi 61363, Taiwan (corresponding author to provide phone: +886-7-3814526ext5639; fax: +886-7-3836844; e-mail: chyang@cc.kuas.edu.tw).

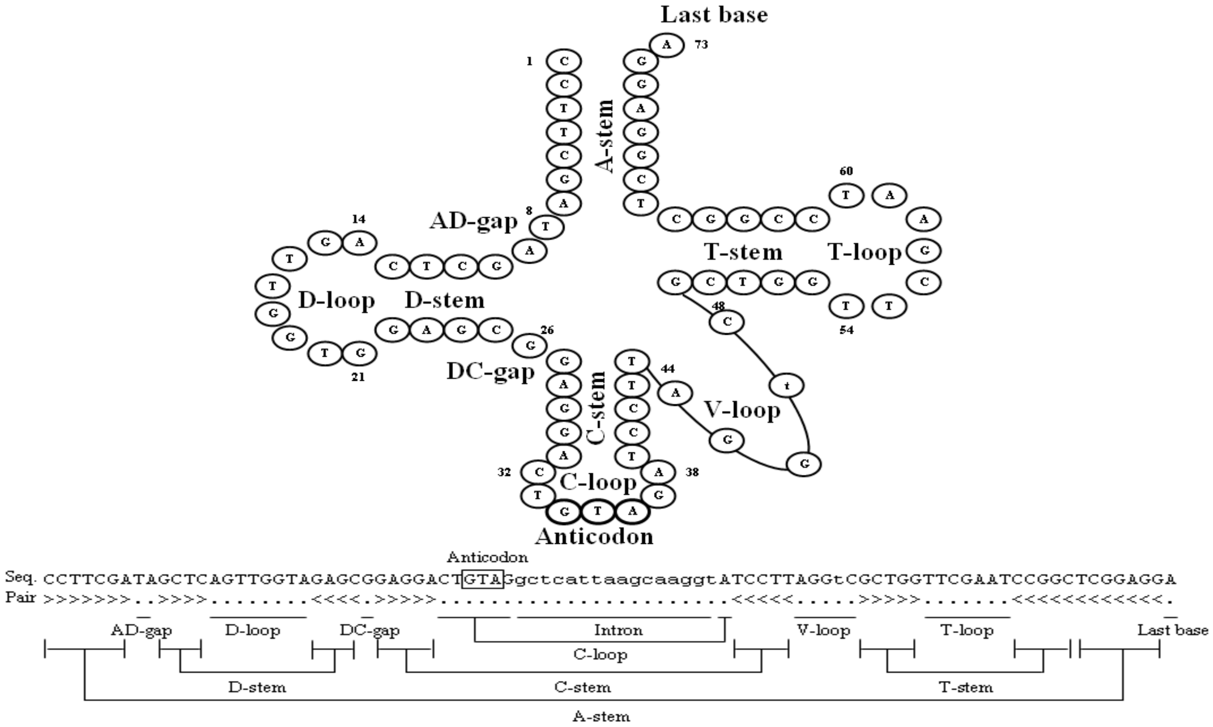
Last base
73
1
8
14
21
26
32
38
44
48
54
60
A-stem
AD-gap
D-loop
D-stem
DC-gap
T-stem
T-loop
C-stem
V-loop
C-loop
Anticodon
Anticodon
Seq. CCTTCGATAGCTCAGTTGGTAGAGCGGAGGACTGTAGgctcattaagcaaggtATCCTTAGGtCGCTGGTTCGAATCCGGCTCGGAGGA
Pair >>>>>>>..>>>>........<<<<.>>>>>.......................<<<<<.....>>>>>.......<<<<<<<<<<<<.
AD-gap
D-loop
DC-gap
Intron
V-loop
T-loop
Last base
C-loop
D-stem
C-stem
T-stem
A-stem

Fig.1 tRNA cloverleaf diagram

II. Method

2.1 Definition tRNA secondary structure model

The notions used in this paper are shown in Fig. 1. The information describes the tRNA secondary structure. The first line is a predicted tRNA sequence. From the first line, the introns and extra base of non-numbering system [4] are printed in lower-case letters. In the block, the GTA is represented as anticodon. The second line contains the folding of the tRNA of the prediction secondary structure with the nested > and < symbols to represent the based pairings. The four stacked pairs , they are acceptor stem (A-stem) of 7 bp long, dihydrouridine stem (D-stem) of 4 bp long, anticodon stem (C-stem) of 5 bp long, TΨC stem (T-stem) of 5 bp long. The four hairpin loops, they are TΨC loop (T-loop) of 7 bases long, variable loop (V-loop) of 5 bases long, anticodon loop of bases long, and dihydrouridine loop (D-loop) of 8 bases long. The intron hides in C-stem between the sequence positions of 37 and 38 at sometime. Our parameters of structures come from our experimental runs through tRNAscan-SE, ARAGORN and tRNAfinder. The observation for multiple databases such as GtRNAdb [5] (http://rna.wustl.edu/GtRDB/), tRNADB-CE [6] (http://trna.nagahama-i-bio.ac.jp/cgi-bin/trnadb/index.cgi), tRNAdb [7] (http://trnadb.bioinf.uni-leipzig.de/), and literatures [8, 9, 10, 11, 12] gave us an insight to approach the secondary structure model. The selection of parameters and adjustment of their values are optimized after reducing the incorrect predictions. The observations of characteristics from the irregular tRNA structures are shown in the constraint A of Table 1.

2.2 Prediction of the tRNA gene and tRNA secondary structure method

The studies of tRNA, two important predictions are the search of tRNA and prediction of the secondary structure. In this section, function of our proposed algorithm is categorized into two parts:

(1) Searching tRNA gene from genomic sequence:

The first part is from the input of the complete genomic sequence to find tRNA gene. The PtRNASS is designed by tRNA secondary structure. The known tRNA patterns are applied to recognize tRNA gene from complete genomic. A suitable recognizing is calculated by 1290 known tRNAs from Sprinzl database. The ability of global search for tRNA gene depends on the validity of tRNA secondary structure prediction. The consideration of the predicting process is shown in details as below.

(2) Prediction of the optimal tRNA secondary structure:

Secondly, the folding of the optimal tRNA secondary structure from tRNA sequence is to distinguish the anticodon. There are multiple choices to construct tRNA secondary structure. From one tRNA gene sequence, it is possible to folds up various tRNA secondary structures, however, it is also possible to predict the anticodon and unfortunately the found anticodon from secondary structure may be a false. The false of anticodon leads to one of the loss of amino acid.

Several requirements are considered, they seem to have impacts on prediction. There are numbers of hydrogen bonds inside each portion of the structure, the total number of the sequence length, the loop length, the intron length, GC% and the significant patterns. Table 1 is the lists of the critical constraints of the work. Fig. 3 shows the prediction flowchart. From Table 1, its constraints and steps are described in details as below.

Table 1. Constraints and parameters use in the search of tRNA secondary structure.

Constraint A: substructure length										
	A-stem	AD-gap	D-stem	D-loop	DC-gap	C-stem	C-loop	V-loop	T-stem	T-loop
Minimum length	6	2	4	4	1	5	7	4	5	4
Maximum length	7	2	4	11	1	5	7	21	5	7

Constraint B: intron and tRNA length			
	Archaea	Bacteria	Eukarya
Intron length	6 to 121	0	1 to 60
tRNA length	63 to 217	63 to 95	63 to 155

Constraint C: Numbers of GU pairings and mm allowed									
	A-stem		D-stem		C-stem		T-stem		All stems
	GU	mm	GU	mm	GU	mm	GU	mm	GU + mm
Maximum Number	3	3	2	3	2	3	3	3	7

Constraint D: Minimal GC percent
18%

In the constraint A, the minimum size of D-loop for the four important positions at 14, 15, 18 and 19 are playing as the determinative roles in folding into tertiary structure. They support the critical L-shaped structure of tRNA molecule. In the constraint B, the introns are found in the C-loop between sequence positions of 37 and 38; lengths of intron for various species are referred from [11]. According to the above description, lengths of a complete tRNA for various species are obtained. The constraint C is mainly a constraint to throw out the unfeasible sequence. We adopted the constraint C from literature [11]. The constraint D is the minimum GC% in sequence. According to our experiments on tRNA sequence of *Saccharomyces cerevisiae* mitochondrion tRNA-Arg (accession number NC_001224 from the range of 69289 to 69362) are discovered from GtRNAdb database.

In our method, we uses GC% as a basis to sift out the unfeasible sequence fragments. This preprocess will speed up the computation in prediction tRNA gene. The percentage of nucleotides G and C in sequence S is denoted as $GC_{ratio}(S)$. The notions of $GC_{ratio}(S)$ and GC%(S) are defined as follows:

$$GC_{ratio}(S) = \frac{S_{total}(G) + S_{total}(C)}{|S|} \qquad (1)$$

$$GC\%(S) = \begin{cases} \text{discard, if } GC_{ratio}(S) < 18\% \\ \text{retain, if } GC_{ratio}(S) \geq 18\% \end{cases} \qquad (2)$$

From genomic sequence, there are sequences that satisfy GC% requirement. An optimal structure will be constructed based upon the characteristics of hydrogen bonds. The next step, the constraints C is used to delete many of unfeasible structures. In the folding, if any of the stem is not satisfied by the minimum base paring, then the corresponding loop will attempt to adjust the size until it reaches the maximum, in order to find the best stem. Whenever the cloverleaf is built, the score is given by the calculated amount of hydrogen bonds through Eq. 3, i.e. one, two and three hydrogen bonds for AU, GC and GU pairs are given as below.

$$\text{hydrogen bond score} = \begin{cases} 1 & \text{, if GU base pair} \\ 2 & \text{, if AU base pair} \\ 3 & \text{, if GC base pair} \end{cases} \qquad (3)$$

Although the most stable structure for some sequences can be calculated, however, the problem of non-tRNA sequences are falsely predicted to fold into cloverleaf structure awaiting to be solved. To overcome this difficulty, we provide the score that depends on a graphical pattern by Marck [11]. A punished score is computed by Eq. 4, it modifies the prediction score that makes a boundary to distinguish between the true and false of all candidates (cut-off = 50). Finally, the integral score of all substructures are calculated by Eq.5. It is to search the potential tRNAs.

$$\textbf{punished score} = \begin{cases} \textbf{-4} & \textbf{, if base pair conform to fig.4} \\ \textbf{0} & \textbf{, if base pair not conform to fig.4} \end{cases} \qquad (4)$$

$$\text{Score} = \sum \text{hydrogen bonds score} + \sum \text{punished score} \qquad (5)$$

The next step, when each sequence fragment is predicted, the other strand will be transformed into Minus strand to check whether any of tRNAs are existed in this region. The input of sequence called Plus strand and Minus strand are illustrated in Fig. 2.

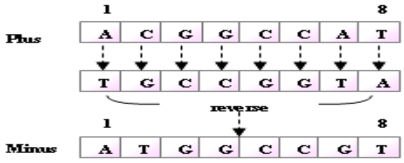

Fig.2 Plus strand transforms into Minus strand diagram

Table 2. tRNA detection rates for tRNAscan-SE, ARAGORN and PtRNASS

Sequence soure	No. of	No. of tRNAs detected			Detection rate (%)		
	tRNAs	tRNAscan-SE [1]	ARAGORN [2]	PtRNASS	tRNAscan-SE [1]	ARAGORN [2]	PtRNASS
Archaea	161	160	**161**	**161**	99.38	**100.00**	**100.00**
Bacteria	686	682	684	**686**	99.42	99.71	**100.00**
Eukaryota	443	437	435	**440**	98.65	98.19	**99.32**
total	1290	1279	1280	**1287**	99.12	99.22	**99.77**

A correct position termination site is identified, and its individual score value is added to the punished score to obtain a total score value. It needs to satisfy the training value (cut-off = 50). When any candidates that are overlapping with each other, according to their scores, they will be searched, as the result, the one with the highest score will be selected as the algorithm result.

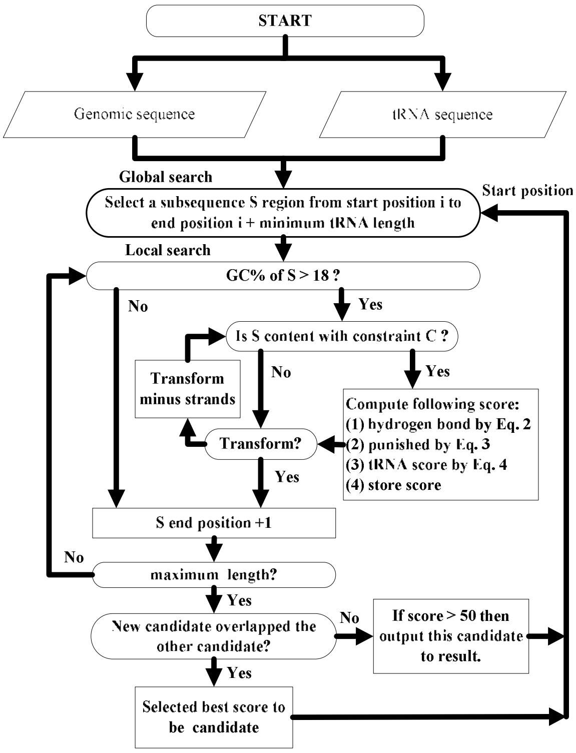
START
Genomic sequence
tRNA sequence
Global search
Select a subsequence S region from start position i to end position i + minimum tRNA length
Start position i +1
Local search
GC% of S > 18 ?
No
Yes
Is S content with constraint C ?
Transform minus strands
No
Yes
Compute following score:
(1) hydrogen bond by Eq. 2
(2) punished by Eq. 3
(3) tRNA score by Eq. 4
(4) store score
Transform?
Yes
S end position +1
No
maximum length?
Yes
New candidate overlapped the other candidate?
No
If score > 50 then output this candidate to result.
Yes
Selected best score to be candidate

Fig.3 tRNA gene search and secondary structure prediction flowchart

III. Result and Discussion

The test for our tRNA secondary structure prediction is based on tRNA gene sequences that are obtained from the database Sprinzl, it was updated in 2007. (http://www.old.uni-bayreuth.de/departments/biochemie/sprinzl/trna/) The tRNAscan-SE and ARAGORN test their own algorithm in 1995 and 1999 versions. The Sprinzl database provides a set of reliable true tRNA for testing the sensitivity of prediction. It contains the most comprehensive tRNA from wide variety of organisms, and are divided into three different sets of tRNA genes, from Archaea (161 sequences), Bacteria (686 sequences) and Eukaryota (443 sequences). In addition, the three complete chromosome genomes obtain from species (NC_*): NC_000909 [*M.jannaschii*], (NC_*): NC_002695 *E.coli* O157:H7], and (NC_*): NC_001133 to NC_001148 [*S.cerevisiae*] are used to testify the tRNA gene search method.

According to Sprinzl database, our test results reveal the prediction sensitivity for species Archaea and Bacteria are both 100% accurate. The species of Eukarya is 99.3% (Table 2). There are three incorrect predictions for Eukarya. The two incorrect predictions are Sprinzl ID DQ8510 and DA9360, they are missed by PtRNASS. For the third one, the anticodon is falsely predicted. The prediction of tRNA gene from chromosome are presented as follow: NC_000909 found 37 genes are correctly predicted in Table 3, NC_002695 found 103 genes are correctly predicted in Table 4, NC_001133 to NC_001148 found 275 genes are correctly predicted in Table 5. Thus, the results demonstrated our suggested method outperforming other methods in various areas.

During the process, we noticed an identical sequence showed in several of different configurations have the same anticodon. This unexpected finding brought our attention to the length of a secondary structure. It does not affect the stability of stem structure, as long as the following rules are satisfied.

In tRNA cloverleaf, many loops of their sizes can be adjusted to fulfill the construction of stems. If many of the non-pairing bases are appeared at the stem. This situation often occurs at only one stem. We applied the non-pairing occurrences in stem structures process; therefore, the flexibility is given to each stem with a better prediction. In addition, some restrictions are made to maintain its integrity of overall structure, e.g. if DS appears to have only one base pair, then the other stems will not have multiple mismatches. Although structures are from different compositions which have the same anticodon, however, the deciding factor is whether the prediction can be folded into a tertiary structure or not. Thus, we limit the number of prediction as one of the restriction in tertiary structure. The common features in predicting tRNA secondary structure are nucleotides appearing at the fixed positions. These features were analyzed by Marck [11] with more than 4,000 sequences. This characteristic in our prediction result show that there is no nucleotide is fixed at one position in secondary structure. According to the above factor, ARAGORN system may result in failure. The reason of this failure is due to the irregular tRNA gene, it can not be applied in T loop from its motif “TRGYNAA”.

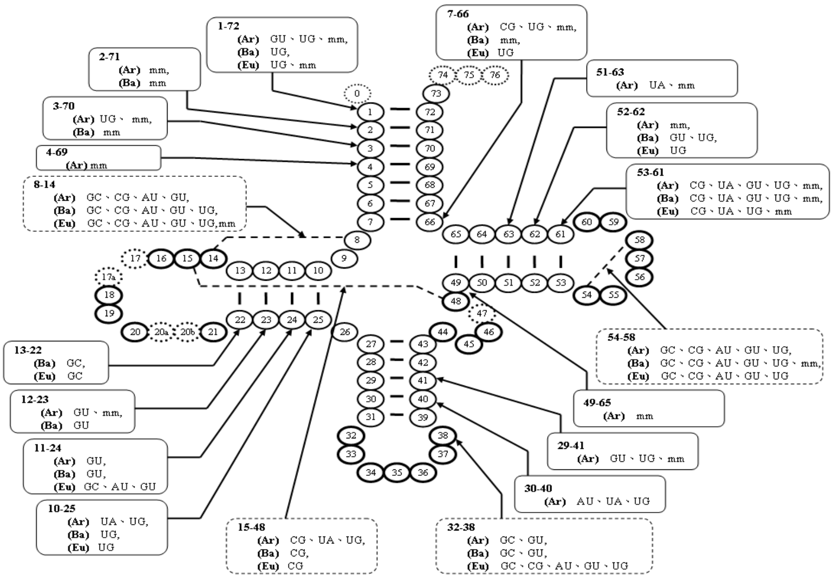

Fig.4 tRNA patterns

A cloverleaf structure and all patterns show that the never found base pairs within over 4000 tRNA genes are from 50 fully sequenced genomes [11]. In illustration, each position for never found base pair is shown in block. 3D base pairs are indicated by dotted lines with the base numbers. The other base pairs are indicated by real lines. If any stem is predicted by these base pairs, then the punished score can be modified by prediction score. This makes a boundary to distinguish true and false from the candidates.

The using of this pattern is effective for searching tRNA genes. We applied a training score (cut-off = 50) to provide a flexible structure prediction from a modified system. The given score allows the predicted structure to have unusual characteristics, i.e., loops, stems or patterns. If their unusual proportions are over the limited restriction then the penalties will decline the total value of tRNA. We decided to set up the threshold as 50 for the optimal result after multiple tests. In order to process the protein synthesis, tRNAs matured in the cytoplasm needs to have 3'CCA terminus at the positions of 74, 75 and 76. However, tRNAs in Eukarya lacks of 3'CCA characteristic. From many of tRNAs, Bacteria and Archaea do not have 3'CCA characteristics, so we abandoned tRNA gene searching feature.

When we compared with the other popular tools, tRNAscan-SE and ARAGORN, in conclusion, there is no absolute winner. According to the previous perspective, if computation search time is the fastest, then ARAGORN will become the lead than other tools. The reason of being the lead is when ARAGORN as the basis for using consensus sequence. In contrast, we used a combination of GC% and acceptor stem to search the most fitting segment. In order to evaluate the quality of prediction, we need to compare the sensitivity value with others. As the result, our method received the highest sensitivity value because we used the extraordinary structure model in secondary structure prediction. The tRNA secondary structure prediction is the most important contribution of our method.

IV. Conclusion

Our method provides the prediction of tRNA gene and the secondary structure. Users can use either a complete chromosome or sequence fragments to predict the locations of tRNA gene and tRNA secondary structure.

We chose the three chromosome genomes from species *M.jannaschii*, *E.coli* O157:H7 and *S.cerevisiae* as the testing sets to find the complete tRNA genes. We adopted a score from training values, they are considered as the unique tRNA secondary structure to predict the tRNA locations. We constructed possible structures and selected the most stable structure. As the result, it not only demonstrates the exact tRNA location but also predicts the best structure. Its tRNA anticodon prediction also matches the literature. This paper is from IMECS 2010 conference [14].

Table 3. Information of the counted anticodons for searching NC_000909.

2nd base

1st base	A		G		C		U		3rd base
A	Phe	0	Ser	0	Cys	0	Tyr	0	A
G		1		1		1		1	
C	Leu	0	Ser	0	Trp	1	Stop	0	
U		1		1	Stop	1		0	
A	Leu	0	Pro	0	Arg	0	His	0	G
G		1		1		1		1	
C	Leu	0	Pro	0	Arg	0	Gln	0	
U		1		1		1		1	
A	Val	0	Ala	0	Gly	0	Asp	0	C
G		1		1		1		1	
C	Val	1	Ala	0	Gly	0	Glu	0	
U		1		2		1		2	
A	Ile	0	Thr	0	Ser	0	Asn	0	U
G		1		1		1		1	
C	Met	3	Thr	0	Arg	0	Lys	0	
U	Ile	0		1		1		1	

Table 4. Information of count anticodons for searched NC_002695.

2nd base

1st base	A		G		C		U		3rd base
A	Phe	0	Ser	0	Cys	0	Tyr	0	A
G		2		2		1		3	
C	Leu	1	Ser	1	Trp	1	Stop	0	
U		1		1	Stop	1		0	
A	Leu	0	Pro	0	Arg	4	His	0	G
G		1		1		0		1	
C	Leu	3	Pro	1	Arg	1	Gln	2	
U		1		2		4		2	
A	Val	0	Ala	0	Gly	0	Asp	0	C
G		2		2		4		3	
C	Val	0	Ala	0	Gly	1	Glu	0	
U		5		3		1		4	
A	Ile	0	Thr	0	Ser	0	Asn	0	U
G		3		2		1		4	
C	Met	15	Thr	1	Arg	1	Lys	0	
U	Ile	0		1		8		5	

Table 5. Information of count anticodons for searched NC_001133 to NC_001148

2nd base

1st base	A		G		C		U		3rd base
A	Phe	0	Ser	11	Cys	0	Tyr	0	A
G		10		0		4		8	
C	Leu	10	Ser	1	Trp	6	Stop	0	
U		7		3	Stop	0		0	
A	Leu	0	Pro	2	Arg	6	His	0	G
G		1		0		0		7	
C	Leu	0	Pro	0	Arg	1	Gln	1	
U		3		10		0		9	
A	Val	14	Ala	11	Gly	0	Asp	0	C
G		0		0		16		16	
C	Val	2	Ala	0	Gly	2	Glu	2	
U		2		5		3		14	
A	Ile	13	Thr	11	Ser	0	Asn	0	U
G		0		0		2		10	
C	Met	10	Thr	1	Arg	1	Lys	14	
U	Ile	4		4		11		7	

REFERENCES

[1] Lowe, T.M. and Eddy, S.R. "tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence," *Nucleic Acids Research*, Vol.25, 1997, pp.955-964.

[2] Laslett, D. and Canback, B., "ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequence," *Nucleic Acids Research*, Vol. 32, 2004, pp.11-16.

[3] Kinouchi, M. and Kurokawa, K., "tRNAfinder: A Software System To Find All tRNA Genes in the DNA Sequence Based on the Cloverleaf Secondary Structure," *Journal of Computer Aided Chemistry*, Vol.7, 2006, pp.116-124.

[4] Sprinzl, M. and Vassilenko, K.S., "Compilation of tRNA sequences and sequences of tRNA genes," *Nucleic Acids Res*, **33**, 2005, pp.139-140.

[5] Chan, P.P. and Lowe, T.M., "GtRNAdb: a database of transfer RNA genes detected in genomic sequence," *Nucleic Acids Research*, Vol.37, 2009, pp.93-97.

[6] Abe, T., Ikemura, T., Ohara, Y., Uehara, H., Kinouchi, M., Kanaya, S., Yamada, Y., Muto, A. and Inokuchi, H., "tRNADB-CE: tRNA gene database curated manually by experts," *Nucleic Acids Research*, Vol.37, 2009, pp.163-168.

[7] Juhling, F., Morl, M., Hartmann, R.K., Sprinzl, M., Stadler, P.F. and Putz, J., "tRNAdb 2009: compilation of tRNA sequences and tRNA genes," *Nucleic Acids Research*, Vol. 37, 2009, pp. D159-D162.

[8] Dirheimer, G., Keith, G., Dumas, P. and Westhop, E., "Primary secondary and tertiary structures of tRNAs," *ln: Söll D, RajBhandary U, eds. tRNA: Structure, biosynthesis, and function. Washington DC*: ASM Press, 1995, pp.93-126.

[9] Macey, J.R., Schulte, J.A., 2nd, Larson, A., Tuniyev, B.S., Orlov, N. and Papenfuss, T.J. "Molecular Phylogenetics, tRNA Evolution, and Historical Biogeography in Anguid Lizards and Related Taxonomic Families," *Molecular Phylogenetics and Evolution*, Vol.12, No.3, 1999, pp. 250-272.

[10] Kinouchi, M., Kanaya, S. and Kudo, Y., "Detection of Transfer RNA Based on the Cloverleaf Secondary Structure," *Journal of Computer Aided Chemistry*, Vol.1, 2000, pp.76-81.

[11] Marck, C. and Grosjean, H. "tRNomics: Analysis of tRNA genes from 50 genomes of Eukarya, Archaea, and Bacteria reveals anticodon-sparing strategies and domain-specific features," *Bioinformatics*, Vol.8, 2002, pp.1189-1232.

[12] Tsui, V., Macke, T. and Case, D.A., "A novel method for finding tRNA genes," *RNA*, Vol.9, 2003, pp.507-517.

[13] Galindo, J., Bermudez, C.I and Daza E.E., "tRNA structure from a graph and quantum theoretical perspective," *Journal of Theoretical Biology,* 2006, pp. 574-582.

[14] Chuang, L.Y., Lin, Y.D., and Yang, C.H., "A Method of Predicting tRNA Secondary Structure from Nucleotide Sequences," Lecture Notes in Engineering and Computer Science: Proceedings of The International MultiConference of Engineers and Computer Scientists 2010, IMECS 2010, 17-19 March, 2010, Hong Kong, pp223-227