

## CIDA: A Proposed Data Base for Foot-and-Mouth Disease Virus (FMDV)

Alaa Eldin Abdallah Yassin<sup>\*1</sup>, Ahmed Farouk Al-Sadek<sup>1</sup>, Amr Badr<sup>2</sup>

<sup>1</sup> Central Lab for Agricultural Experts Systems, Ministry of Agriculture and Land Reclamation  
Giza, Egypt

<sup>2</sup> Department of Computer Science, Faculty of Computers and Information, Cairo University  
Cairo, Egypt  
[aboelmnzer@gmail.com](mailto:aboelmnzer@gmail.com)

### Abstract

*The lack of vaccines and drugs for viruses becomes a big challenge in the recent years, especially with the big variety of families and serotypes for each virus, and this problem appear clearly with RNA viruses; because the big mutation and instability nature of this type of viruses, which prevent from the long life affect of the vaccines, for that reason a lot of researches have been conducted to overcome this problem whether at life cycle, genome sequencing, second and tertiary structure level. one of the viruses which affects directly on the animal health and its life at all, and also affects on the economics of the world is foot-and-mouth disease virus (FMDV) which is belonging to picornafamily. The objective of this paper is : introduce a data base for FMDV to help researchers in this point of research. The proposed data base consisting of four level of data: Complete genome, Internal ribosome entry site (IRES), Domain 3 of IRES, and Apical region of domain 3 with total number of records around 2000 records at the different levels of data . The data base named CIDA (the first letter of each level of data) and can accessed and downloaded free at : [www.claes.sci.eg/fmdv.CIDA](http://www.claes.sci.eg/fmdv.CIDA). The data are available in FASTA format.*

**Keywords:** FMDV data bases; IRES; Apical region.

## 1. Introduction

The continuously growing of list of discovering novel serotypes of FMDV , and the huge amount of IRES sequences which still unknown, adding to that the importance of FMDV virus and untranslated region IRES, explain the need to establishing a special databases for this virus . Till now there is no any database specialized for FMDV with its different regions and domains, but because IRES region considered the key role in replication process of the virus [1,2,3] ; specialized database is created to this region named "IRESite database" [4] , where "IRESite" stored around 600 records, but small number of these records are addressed to FMDV. IRES in FMDV consists of five domains, and domain 3 is the most important domain, also this domain is consists of two regions, apical and basal region, Apical region is the functional region which includes the conserved motifs and junctions which affects on the IRES shape in tertiary level [ 2, 5], and affects on IRES function at all, but also there is no any data base created to this important domains and region, and researchers in this field collect the needed data from general databases like GenBanks[6], and RAG[7]. This paper proposed a special data base for FMDV in the important four different levels, Complete genome, Internal ribosome entry site (IRES), Domain 3 of IRES, and Apical region of domain 3. We named the data

set “CIDA”. This paper is organized into the following sections; the first section will introduce FMDV, its importance and challenges, the second section will summarize the importance domains and regions in FMDV, the third section is the most important section in the paper which addresses to talk about our proposed database called CIDA, then the evaluation section which evaluate the CIDA data set, finally the conclusion of the paper by summarize the added value for the proposed data base.

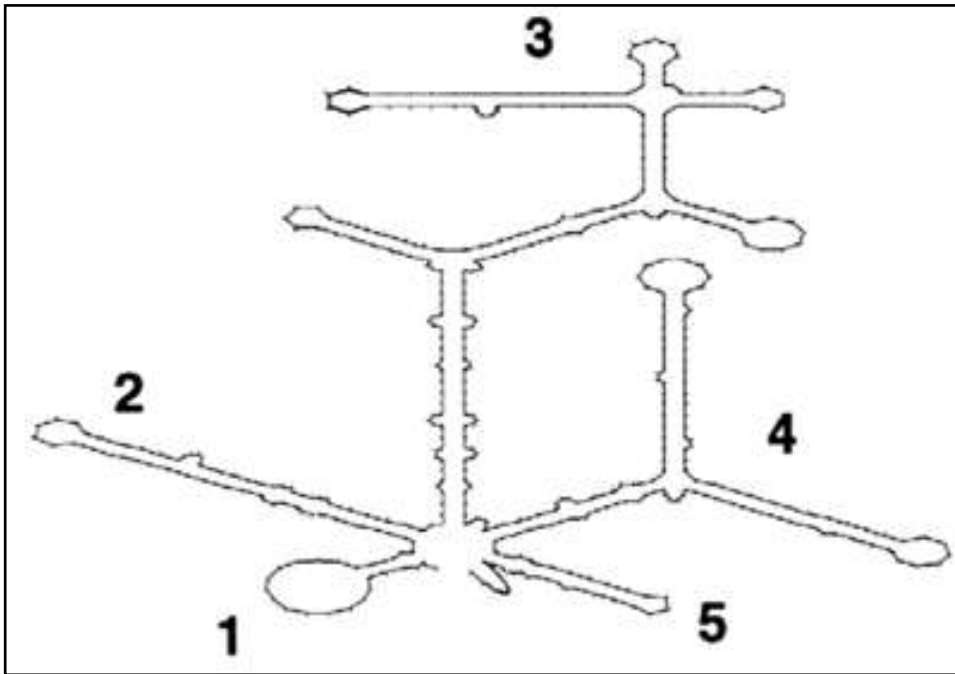
## 2. FMDV Importance and Challenges

Foot-and-mouth disease (FMD) is belonging to picornavirus family [8,9,10] and is an acute, systemic disease of domestic and wild cloven-hoofed animals including cattle, pigs, sheep, goats, buffalo and various wildlife species, where FMD is caused by Foot-and-mouth disease virus (FMDV)[8,11]. Foot-and-mouth disease virus (FMDV) is responsible for the most important viral disease of cattle and other cloven-hoofed animals which affect directly on the economic side where FMD is a major constraint to international trade in livestock and livestock products [12], because FMDV has high contagious nature, which associated productivity losses and make it to be considered the most important livestock disease in the world [8]. The seriousness of this disease lies in the virus spread quickly once it appears where it is a contagious disease, adding to the seriousness of the disease, that in some animals as cattle, accompanied by other symptoms of illness, such as fever, lameness and loss of appetite are appear, which makes control of the disease more difficult [12], also young calves may die without showing any clinical signs. The infection of this disease causes mastitis in dairy cattle and economically, which affects a direct impact on the production of milk as well as meat production, also the big losses as a result to the mortality which occur in young animals [12]. FMDV has the ability to contaminate the environment, where some symptoms of the virus like blisters rupture, releasing vast numbers of infectious virus particles [12], and the infection happened through wind and water , where the wind are sufficient to infect over 250 km. And which increases the seriousness of the disease and the difficulty of dealing with him, that when sheep or goats become infected with FMD, the disease may not be diagnosed for a considerable time because the clinical signs may be very mild, and also one pig sheds 400,000,000 virus particles per day and only takes 10-12 virus particles to infect one cow. Foot and mouth disease virus can be transmitted via: Milk, Semen, Saliva, Feces, Urine, Breath Vectors and Contact with contaminated objects (equipment, footwear, and clothing) [13]. The big challenge facing the veterinarian's diagnosis of FMD is the convergence of the symptoms of Vesicular stomatitis virus (VS) with FMD virus that makes the veterinarian's was unable to diagnose the disease once it sees the symptoms that accompany the disease [13], and we also have to look after the economic side and the extent of affected because of this serious disease, take into account the ethical side, because the infected animal is suffering a lot and suffer because of injury, in addition to the symptoms that show it from the blisters in his mouth prevented him from eating in a natural way.

## 3. The Importance Domains and Regions in FMDV

FMDV like any other virus consists of two main regions; translated region or “Coded region” and untranslated region “”, the coded region is responsible for producing the proteins of the virus, and the untranslated region UTR named “IRES” although it's don't replicated to be a protein but it's very important region and considered the first step in replication process [14], because IRES placed before the start codon“ AUG” and contains some conserved motifs affect in its activity to start its functions in initiation and replication process to the virus cycle at all. The biologists through laboratory tests discovered that, the composition of IRES region differs from virus family to another, for an example the hepatitis C virus which is belonging to Flavivirus family, internal

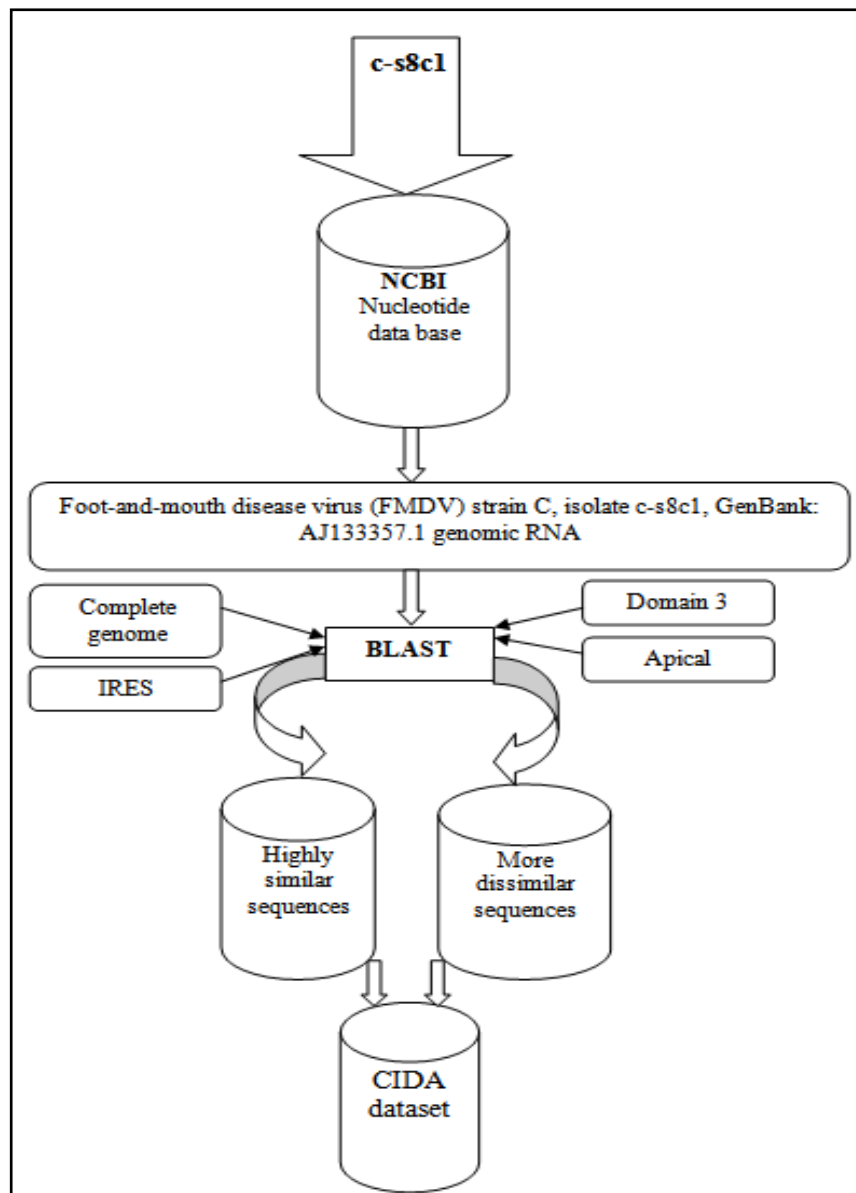
ribosome entry site (IRES) element contains four domains, but in picornavirus family which FMDV belonging to it, IRES consists of ~ 450 nt and can fold in multiple stem-loops organized in five domains as presented in figure 1 [14]. Shortly domain 3 is the most important domain in IRES, and its consists of two parts or region; basal and apical [16]. Apical region is the functional region which includes the conserved motifs of IRES region [14]; so for the short notes that we mentioned the proposed data base "CIDA" consists of 4 level of data according to the importance of those parts in FMDV virus, the different 4 level will discussed in the later sections.



**Figure 1. Simple Figure to the Five Domains of IRES Region in FMDV, Explain the Differ in Size and Shape between Domains [15]**

#### **4. "CIDA" Data collection phases**

The similarity is the key of solving a lot of biology problems because the previous researches proved that : if the sequence is similar, this mean that they often from the same ancestral sequence, which means that they probably have the same ancestor, share the same structure and function [17], and BLAST is the most popular similarity tool, so we use BLAST tool to collect our dataset . First : We took "Foot-and-mouth disease virus (FMDV) strain C, isolate c-s8c1, genomic RNA" from NCBI portal data base and use it as an input to BLAST tool, and we make similarity process twice , first select the most similar sequences " the top 100 sequence " , and in the second process we make the similarity process done with " 500 sequence" , and we done those two level of sequencing 4 time; with the number of dataset levels (complete genome, IRES, Domain 3 and Apical region). The data set collection phases are shown in figure 2, and later figures represented some print screens from the collection process.



**Figure 2. CIDA Data Set Collection Phases**

NCBI Resources How To

Nucleotide Nucleotide C-SB Search

Advanced

Display Settings: GenBank Send: Change region shown

NCBI Resources How To

Nucleotide Nucleotide Search

Advanced

Display Settings: GenBank Send: Change region shown

Customize view

Foot-and-mouth disease virus - type C polypeptide (P81) gene, partial cds

GenBank: M11027.1

FASTA Graphics

Go to: Analyze this sequence Run BLAST

Figure 3. Typing the Reference Sequences in NCBI Portal

GenBank: M11027.1

FASTA Graphics

Go to: Click here Analyze this sequence Run BLAST Pick Primers Highlight Sequence Features Find in this Sequence LinkOut to external resource

LOCUS ABHP61 1685 bp ss-RNA linear VRL 03-ADG-2006

DEFINITION Foot-and-mouth disease virus - type C polypeptide (P81) gene, partial cds.

ACCESSION M11027

VERSION M11027.1 GI:210476

KEYWORDS polyprotein; protein P18; protein P61; replicase.

SOURCE Foot-and-mouth disease virus - type C (FMDV-C)

ORIGIN Foot-and-mouth disease virus - type C

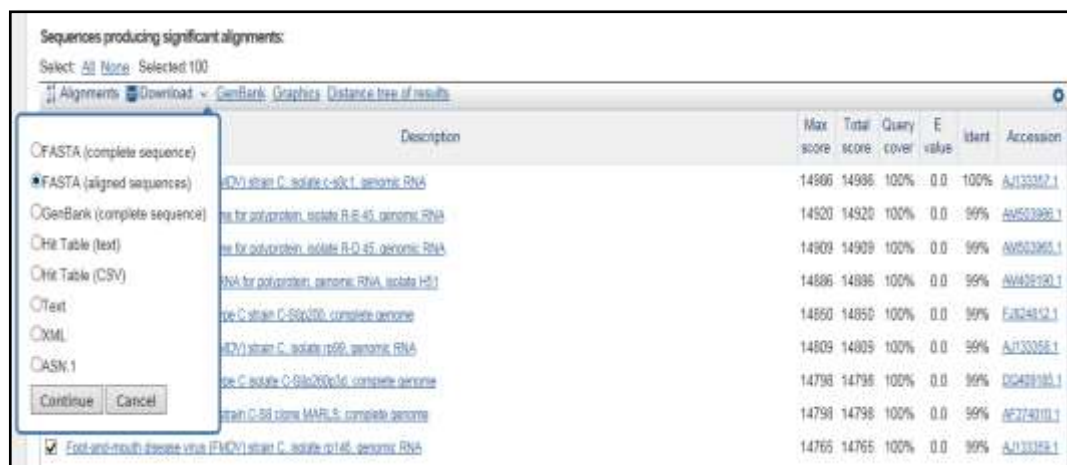
Sequences producing significant alignments:

Select All None Selected:0

Alignments Download GenBank Graphics Distance from results

Description	Max score	Total score	Query cover	E value	Ident	Accession
Foot-and-mouth disease virus - type C polypeptide (P81) gene, partial cds	3112	3112	100%	0.0	100%	M11027.1
Foot-and-mouth disease virus (FMDV) strain C isolate c-slk1 genomic RNA	2940	2940	100%	0.0	98%	AF133357.1
Foot-and-mouth disease virus mRNA for polyprotein, genomic RNA isolate H51	2835	2935	100%	0.0	98%	AM409190.1
Foot-and-mouth disease virus - type C strain C-SB200, complete genome	2929	2929	100%	0.0	98%	FJ824812.1
Foot-and-mouth disease virus gene for polyprotein, isolate R-B 45, genomic RNA	2929	2929	100%	0.0	98%	AM503908.1
Foot-and-mouth disease virus - type C isolate C-SB260000, complete genome	2929	2929	100%	0.0	98%	DQ409184.1

Figure 4. Similarity Process Using BLAT, and its Results



**Figure 5. Downloading the Result in FAST Format**

## 5. CIDA Dataset Structure

CIDA dataset is organized or divided to 4 levels of FMDV sequences :

- 1- Complete genome.
- 2- IRES
- 3- Domain 3.
- 4- Apical region.

CIDA dataset contains 2089 records as a result from the similarity process eight times, four times with the high similarity sequence; top 100 sequences, and four times with dissimilar sequences; 500 sequences. The collected data are formatted in FASTA format.

### Highly similar sequences:

- 100 records for complete genome.
- 100 records for IRES.
- 100 records for Domain 3 of IRES.
- 100 records for Apical regions.

### More dissimilar sequences :

- 500 records for complete genome.
- 500 records for IRES.
- 402 records for Domain 3 of IRES.
- 287 records for Apical regions.

CIDA contains a data for four serotypes of FMDV; serotype C,A,O and serotype Asia, so CIDA dataset structured to 4 folders each one named with the type of data collected; “Complete Genome” folder, “IRES”, “Domain 3” and “Apical” folder. Each folder from those four folders contains another 4 folder named with the 4 different serotypes ; “serotype C”, “serotype A”, “serotype O” and serotype “Asia 1”, thus CIDA have 16 folder in high similarity level of sequences (top 100) , and the dissimilar sequences (500) CIDA have 7 folders the first 3 for “Complete Genome”, “IRES”, and “Domain 3”, regardless serotypes and 4 folders for “Apical “ region with each serotype individually, added to CIDA dataset another repeated folders named ”unknown” which include the sequences that don’t belonging to any of the surveyed serotypes. The number of records in each folder mentioned below and in figure 6 and figure 7. The full size of “CIDA” data set after compressed is around 3 MB , and 22 MB after extracted.

***Complete Genome : in High similar sequences (100)***

- serotype A : 40 records.
- serotype C : 24 records.
- serotype O : 27 records.
- serotype Asia : 4 records.
- unknown serotype : 5 records.

***Complete Genome : in little similar (dissimilar)sequences (500)***

- all serotype : 500 records

***IRES : in High similar sequences (100)***

- serotype A : 40 records.
- serotype C : 24 records.
- serotype O : 27 records.
- serotype Asia : 4 records.
- unknown serotype : 5 records.

***IRES : in little similar (dissimilar)sequences (500)***

- all serotype : 500 records

***Domain 3 : in High similar sequences(100)***

- serotype A : 49 records.
- serotype C :22 records.
- serotype O : 20 records.
- serotype Asia : 4 records.
- unknown serotype : 5 records.

***Domain 3 : in little similar (dissimilar)sequences (402)***

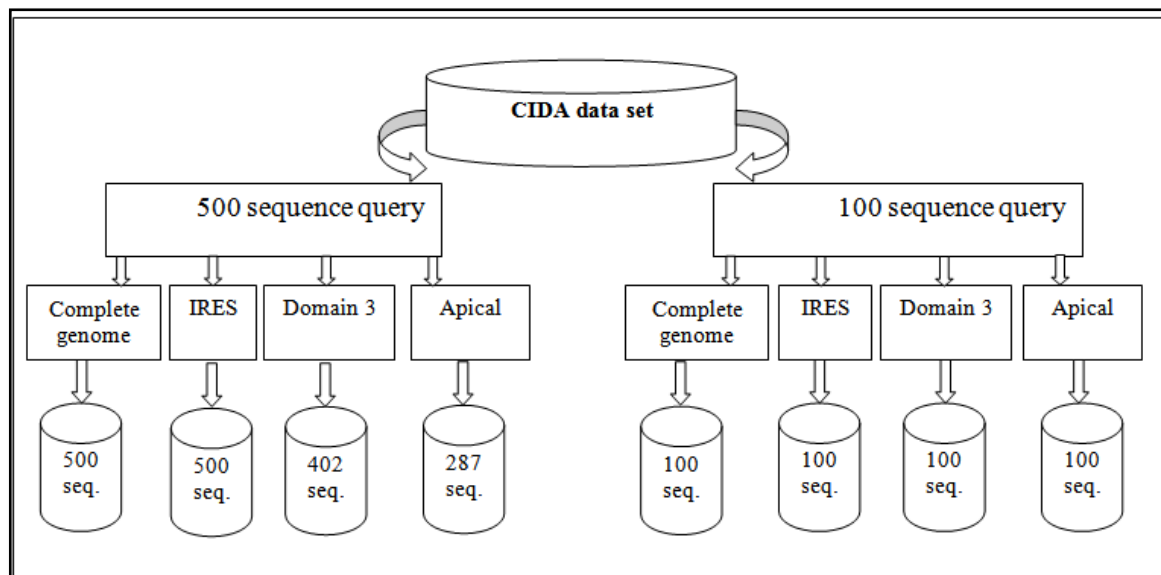
- all serotype :402 records.

***Apical : in High similar sequences (402)***

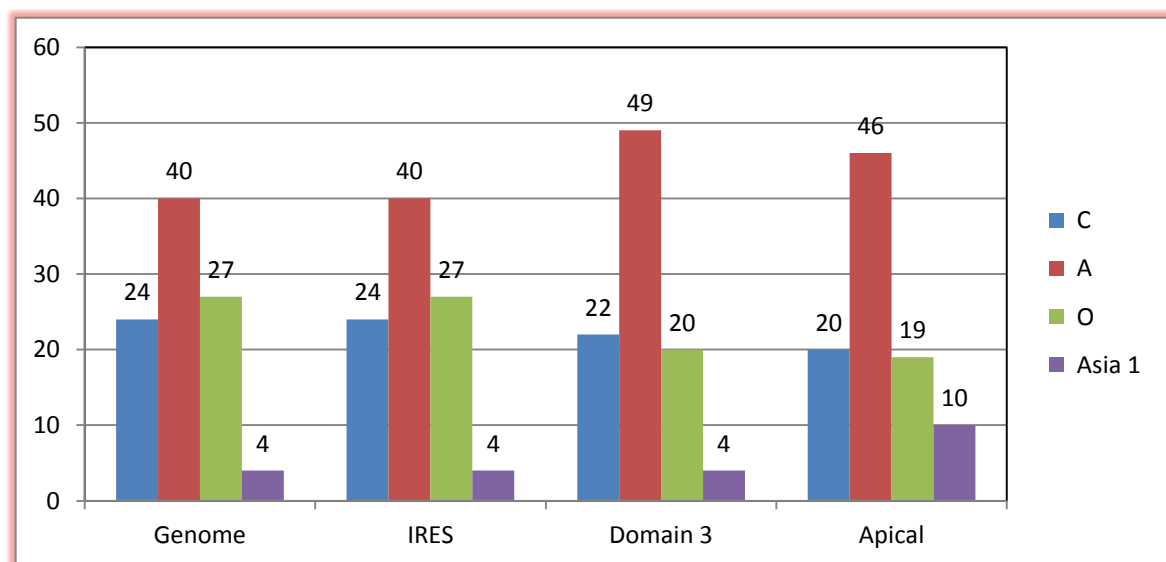
- serotype A : 46 records.
- serotype C :20 records.
- serotype O : 19 records.
- serotype Asia : 10 records.
- unknown serotype : 5 records.

***Apical: in little similar (dissimilar)sequences (287)***

- serotype A : 75 records.
- serotype C :22 records.
- serotype O : 165 records.
- serotype Asia : 20 records.
- unknown serotype : 5 records.



**Figure 6. "CIDA" Data Set Structure**



**Figure 7. "CIDA" Data Set Structure with High Similarity Query According to Serotypes**

## 6. Contribution

The proposed dataset produced a specialized dataset for FMDV which may help researchers in this field in their researches; like studying the effectiveness of mutated nucleotides in each serotypes and also comparative analysis to this collected data and so on.

## References

- [1] G. J. Belsham, "Divergent picornavirus IRES elements" , Elsevier: Virus Research 139, **(2009)** 183–192.
- [2] S. Jung and T. Schlick, "Candidate RNA structures for domain 3 of the foot-and-mouth-disease virus internal ribosome entry site", Nucleic Acids Research, **(2013)**, vol. 41, no. 3 1483–1495.
- [3] S. Lopez de Quinto, Elafuente and E. Martenz-Salas, *et al.*, "IRES interaction with translation initiation factors: Functional characterization of novel RNA contacts with eIF3, eIF4B, and eIF4GII", RNA **(2001)**, 7:1213-1226, DOI: 10+1017+S1355838201010433.
- [4] S. Bonnal, C. Boutonnet, L. Prado-Lourenc\_o and S.Vagner, "IRESdb: the Internal Ribosome Entry Site database", Nucleic Acids Research, **(2003)**, vol. 31, no. 1 427–428.
- [5] A. Badr, A. ElSadek and A.Yassin, "Computational Based Analysis for Internal Ribosome Entry Site (IRES) and Viral Replication in FMDV", International Journal of Computer Science Issues (IJCSI), vol. 12, no. 4, **(2015)**.
- [6] "<http://www.ncbi.nlm.nih.gov/genbank/>", accessed at 6/2016
- [7] "<http://www.biomath.nyu.edu/rag/home>", accessed at 6/2016
- [8] C. Carrillo, E. R. Tulman, G. Delhon, Z. Lu, A. Carreno, Vagnozzi, G. F. Kutish and D. L. Rock, "Comparative genomics of foot and mouth diseases viruses", J. Virol, vol. 79, no. 10, **(2005)**, pp. 6487.
- [9] E.V.Pilipenko, V.M.Blinov, B.K.Chemov, T.M.Dmitrieva and V.I.Agol, "Conservation of the secondary structure elements of the 5'-untranslated region of cardio and aphthovirus RNAs", Nucleic Acids Research, vol. 17, no.14, **(1989)**.
- [10] K. Sandberg, S. E. Mulroney "RNA Binding Proteins: New Concepts in Gene Regulation", chapter: Sung Key Jang, Eckard Wimmer, "The Role of RNA-Binding Proteins in IRES-Dependent Translation", **(2002)**, pp.1-33.
- [11] Institute for Animal Health, Pirbright Laboratory, "Foot-and-mouth disease: scientific problems and recent progress", 1st annual report prepared for DEFRA, Science Directorate, **(2003)**.
- [12] P. Lawrence, M. LaRocco, B. Baxt and E. Rieder, "Examination of soluble integrin resistant mutants of foot-and-mouth disease virus", V. Journal, vol. 10, no. 2, **(2013)**.
- [13] J. Musser and S. Burnham, "Foot-and-Mouth Disease", Texas A&M University College of Veterinary Medicine, presentation (120 slide).
- [14] A. Badr, A. ElSadek and Alaa.Yassin: "Computational Based Analysis for Internal Ribosome Entry Site (IRES) and Viral Replication in FMDV", International Journal of Computer Science Issues (IJCSI), vol. 12, no. 4, **(2015)**.
- [15] R. Ramos and E. Martínez-Salas, "Long-range interaction between structural domains of the aphthovirus internal ribosome entry site (IRES)", RNA, **(1999)**, 5: 1374-1383.
- [16] O. Fernandez -Miragall and E. Maryinez-Salas, "Structural organization of viral IRES depends on the integrity of the GNRA motif", rna journal, RNA, **(2003)**, 9:1333–1344.
- [17] J.-M. Claverie, and C. Notredame, "Bioinformatics for Dummies" – 2nd Edition, Wiley Publishing, Inc., **(2007)**.

