Journal of Animal and Veterinary Advances 11 (15): 2630-2640, 2012
ISSN: 1680-5593

Geometrical Study on FMDV Genome Based on Z-Curve

Hua Tang, Xin-Sheng Liu, Yu-Zhen Fang, Li Pan, Zong-Wang Zhang, Peng Zhou,
Jian-Liang Lv, Shou-Tian Jiang, Yong-Lu Wang and Yong-Guang Zhang
State Key Laboratory of Veterinary Etiological Biology,
National Foot and Mouth Disease Reference Laboratory, Lanzhou Veterinary Research Institute,
Chinese Academy of Agricultural Sciences, 730046 Lanzhou, China

Abstract: The Z-curve is a geometrical tool for visualizing and comparing genomes. Since, the curve contains the information carried by the given sequence, DNA sequences could be analyzed systematically. In this study, the Foot and Mouth Disease Virus (FMDV) was analyzed by the Z-curve Method. This research concludes that all serotypes FMDV have a close consanguinity but each has its own characteristic by contrast, the three South African serotypes have a much closer affinity. FMDV have three A + T-rich regions and two G + C-rich regions. The most variable regions of FMDV lie in protein coding areas. And the non-coding region is very conservative. O serotype FMDV is most variable compared with A and Asia1 serotype. Those O serotype FMDV in buffalo and goat have a mutation trend. Compared with host, geographic location have a grater influence on FMDV mutation. The A serotype FMDV strains occurred in India after 2002 are more stable and have a closer phylogenetic relationships compared with the Asia1 strains which were prevalent before 2002. In general, Z-curve Method shows a widely application prospect in the areas of life sciences. Such as gene sequence, molecular epidemiology, homology, evolution, phylogenetic relationships and genetic diversity analysis.

Key words: Z-curve, geometrical study, FMDV, mutation, homology analysis, China

INTRODUCTION

The increasing genome sequences urge us to exploit some methods to study these sequences. Currently, geometric approach is widely used to analysis DNA sequences and amino acid sequences. The merits of graphical research of these sequences are: intuitive, visual and vivid. Graphical models have been frequently used to describe complex interaction patterns and dependent structures among genes and other phenotypes (Chu *et al*., 2009). Li *et al.* (2010) introduced a cluster-free algorithm based on a graph-regularized version of Partial Least Squares (PLS) regression to learn sequence patterns represented by graphs of k-mers or graph-mers that predict gene expression trajectories. Giacomo *et al.* (2007) applied a graph-based approach to identify cancer signaling pathways from published gene expression signatures.

Z-curve is a 3-dimensional graph which is unique to the given the DNA sequence. That is to say, this graph contains the information carried by the given sequence. Since it was first proposed, the Z-curve has been extensively applied to life science research such as sequence segmentation (Zhang and Zhang, 2003a; Wen and Zhang, 2003), horizontal gene transfer detection (Zhang and Zhang, 2003b), isochoric domain inference (Zhang and Zhang, 2003a, b; Wen and Zhang, 2003) and sequence analysis, for instance, nucleotide distribution analysis (Ou *et al*., 2003), replication origins Identification (Zhang and Zhang, 2005), protein coding genes recognition (Zhang and Wang, 2000), separate base usages of genes revelation (Guo and Yu, 2007) and so on. People have compared Z-curve Version 1.0 and Glimmer version 2.02 which was developed by the American Institute of Genomic Research. The results indicated that Z-curve version 1.0 has a much lower false positive rate than Glimmer 2.02, especially for high G + C content bacterial and archaealgenomes. What's more Z-curve 1.0 has more accurate gene start prediction, lower additional prediction rate and higher accuracy for the prediction of horizontally transferred genes (Guo *et al*., 2003). Because of its outstanding performance, Z-curve version 1.0 attracted widely attention in the filed of international bioinformatics and genomics. Foot and Mouth Disease (FMD) is an acute, highly contagious

Corresponding Author: Yong-Lu Wang, State Key Laboratory of Veterinary Etiological Biology,
National Foot and Mouth Disease Reference Laboratory, Lanzhou Veterinary Research Institute,
Chinese Academy of Agricultural Sciences, 730046 Lanzhou, China

disease of cloven-hoofed animals and is endemic in many regions in Africa, Asia and South America (Alexandersen and Mowat, 2005; Davies, 2002; Thompson *et al*., 2002; Knowles and Samuel, 2003; Cottam *et al*., 2008; Schumann *et al*., 2008). It is caused by the Foot and Mouth Disease Virus (FMDV), a member of the genus *Aphthovirus* of the family Picornaviridae which is highly transmissible and can causes high morbidity outbreaks with moderate to low mortality in most cases. There are seven serotypes of FMDV (A, O, C, Asia1, SAT1, SAT2 and SAT3) with multiple subtypes within each serotype (Paprocka, 2006). The FMDV is a positive sense, single-stranded RNA virus. The genome is about 8,000 nucleotides (nt) in length. The harmfulness of FMD is conspicuous (Sumption *et al*., 2008; Gibbs, 2003; Sobrino *et al*., 2001) and there is no effective way to control this disaster (Mort *et al*., 2005). FMD is perhaps the most important animal disease limiting trade of animals and animal products (Rufael *et al*., 2008; Barasa *et al*., 2008; Perry and Rich, 2007). No effective drugs are currently available to cure this disease. It is unlikely that a single vaccine approach will solve the many short comings of current vaccines (Rodriguez and Grubman, 2009). Gaining insight into the information carried by the genomes of FMDV would be helpful to discover drugs and develop vaccines.

MATERIALS AND METHODS

The FMDV genomes and annotated information were downloaded at the web (site http://www.ncbi.nlm.nih.gov/). The Z-curve coordinates compute software (Zplotter) was downloaded at the web site (http://tubic.tju.edu.cn/zcurve/). Users are recommended to use Z-curve plotter Java applet version which can be available at the website (http://tubic.tju.edu.cn/zcurve/). However, for those who experience difficulty in using Java applet or want to compute Z-curve coordinates locally, this local version of Zplotter can be used.

The Z-curve: The Z-curve is a 3-dimensional space curve which consists of a series of nodes Pn where n = 0, 1, 2, ..., N whose coordinates are calculated by Eq. 1 (Damien and Fares, 2006):

$$\begin{cases} x_n = (An + Gn) - (Cn + Tn) \equiv Rn - Yn \\ y_n = (An + Cn) - (Gn + Tn) \equiv Mn - Kn \\ z_n = (An + Tn) - (Cn + Gn) \equiv Wn - Sn \end{cases} \quad (1)$$

$$n = 0, 1, 2, ..., N,\ x_n, y_n, z_n \in [N, N]$$

where, $A_0 = C_0 = G_0 = T_0 = 0$ and $x_0 = y_0 = z_0 = 0$. In the equation, R, Y, M, K, W and S represent the bases of purine, pyrimidine, amino, keto, weak hydrogen bonds and strong hydrogen bonds, respectively. Usually for an AT-rich genome, z_n is approximately a monotonously increasing linear function of n whereas for a GC-rich genome, z_n is approximately a monotonously decreasing linear function of n. In both cases, it is convenient to fit the curve of z_n~n which is defined in Eq. 2 (Zhang and Zhang, 2003a, b):

$$z = kn \quad (2)$$

However in practice, the resolution ratio of z = kn curve is very low, so in order to view the relevant graph more intuitively, the designer also introduce a z_n'~n curve (simply as Z-curve), it is defined in Eq. 3 (Zhang and Zhang, 2003a, b):

$$z_n' = z_n - kn \quad (3)$$

where, let $\overline{G+C}$ denote the average G + C content within a region Δn in a sequence, researchers deduce Eq. 4 (Zhang and Zhang, 2003a, b):

$$\overline{G+C} = \frac{1}{2}(1 - k - \frac{\Delta z_n'}{\Delta n}) \equiv \frac{1}{2}(1 - k - k') \quad (4)$$

RESULTS AND DISCUSSION

Study on the seven FMDV serotypes: In order to study the whole serotypes FMDV, corresponding FMDV (Table 1) complete genome sequences are analyzed by Z-curve. The results indicate that all serotypes FMDV have a close consanguinity but each has its own characteristic, by contrast, the three South Africa FMDV have a much closer affinity (Fig. 1a, b). The most variable regions of FMDV lie in protein coding areas and the

Table 1: The mark and name and the NCBI accession number as well as the global ACGT base distribution of the seven serotypes FMDV strains mapped in Fig. 1a, b

Mark	Strain annotated in NCBI	Accession no.	The global distribution of ACGT base
a\13	a5allier iso45	AY593780.1	A:2043 C:2315 G:2118 T:1723 GC content:54.60%
o\13	o5india iso34	AY593828.1	A:2007 C:2301 G:2110 T:1739 GC content:54.08%
c\13	C1 Oberbayern c1ober iso88	AY593805.1	A:2031 C:2313 G:2073 T:1725 GC content:53.87%
Asia1\10	YNBS/58	AY390432.1	A:2013 C:2263 G:2105 T:1782 GC content:53.51%
Sat1\2	SAT1/5sa/61 iso13	AY593842.1	A:2072 C:2366 G:2012 T:1726 GC content:53.55%
Sat2\8	SAT2-1rhod/48	AY593847.1	A:2056 C:2321 G:2017 T:1725 GC content:53.43%
Sat3\2	SAT3-3kenya 11/60	AY593852.1	A:2046 C:2316 G:2035 T:1767 GC content:53.22%

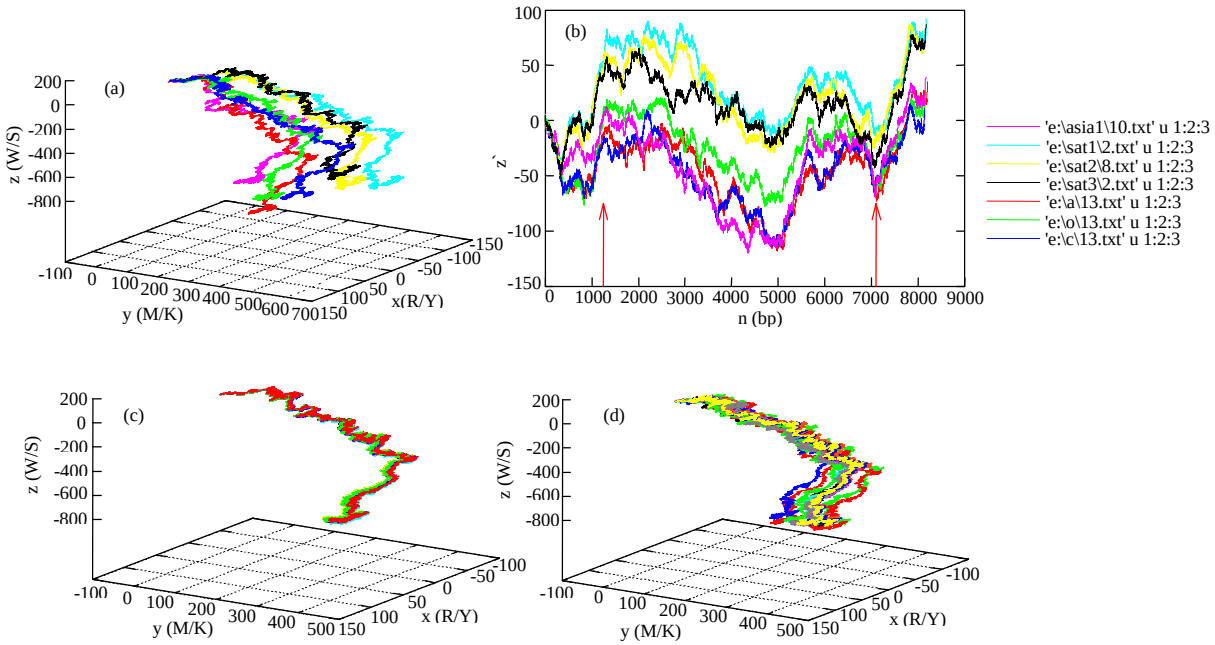

Fig. 1: a-d) Z-curves and the 3-dimensional curves of 7 serotypes FMDV strains

non-coding region is very conservative (Fig. 1b) and FMDV have three A + T-rich and two G + C-rich regions. And the G + C-rich regions are immediately followed by A + T-rich regions (Fig. 1b).

The mutation research of O serotype FMDV based on host and geographic location: Here, 12 FMDV strains isolated in United Kingdom which have different hosts and another 12 strains isolated in different regions which have an identical host (Table 2 and 3) are analyzed to study the mutation of FMDV based on host and geographic location. And the former show a much closer consanguinity than the latter. Results also show that the impact of geographic location on FMDV mutation is greater than that of host (Fig. 1c, d). An up jump in the Z-curve indicates a decrease of G + C content whereas a drop indicates an increase of G + C content. Any minimum point in the Z-curve indicates an abrupt change from G + C-rich region to G + C-poor region. The most different region (between the two arrows) means a most variable region of FMDV. The 3-dimensional curves of which are more similar and overlap better have closer consanguinity. Figure 1a, b indicates that the three South Africa serotypes have a much closer affinity and the most variable regions of FMDV lie in protein coding areas; Fig. 1c shows the 3-dimensional curves of 12 strains in United Kingdom which have different hosts and Fig. 1d shows the 3-dimensional curves of 12 strains in Asia and Europe which have the same host cattle. Figure 1a, b shows that compared with host, geographic location have a grater influence on FMDV mutation.

The mutation research of O serotype FMDV based on host: Researchers have studied the mutation of O serotype FMDV based on their hosts. The O serotype FMDV (Table 2 and 3) which have host annotated and complete genome available in NCBI are analyzed by Z-curve. The results indicate that O serotype FMDV is most stable in sheep and most variable in swine (Fig. 2a-c). Israel 07-6378 (goat) is very similar to PAK/45/2008 (buffalo) while they are obviously different from others in the foreside. So, researchers speculate that O serotype FMDV has a mutation trend in goat and buffalo. And researchers guess that there is a certain relationship between their occurrence when taking the collection time and isolation region into consideration (Fig. 2b-d). Figure 2b shows the 3-dimensional curves of O serotype FMDV isolated in sheep. Fig. 2c is 3-dimensional curve of O serotype FMDV isolated in bovine and Fig. 2 shows the 3-dimensional curves of Israel 07-6378 (goat) and PAK/45/2008 (buffalo). These illustrate that the O serotype is most stable in sheep and most variable in swine. In addition, the O serotype has a mutation trend in goat and buffalo. And may be, there is a certain relationship between the occurrence of Israel 07-6378 (goat) and PAK/45/2008 (buffalo).

The mutability research of Asia1, A, O, serotypes: Corresponding Asia1, A, O, serotypes (Table 2, 4 and 5), complete genomes have been conducted Z-curve analyses, arming to study the mutability of them. According to the results, the O serotype FMDV is most variable compared with Asia1 and A serotype (Fig. 3a-c).

Table 2: The Z-curve plot marks, names, accession numbers, hosts as well as isolation regions and collection dates of Asia1 serotype FMDV shown in Fig. 3a

Mark	Strain annotated in NCBI	Accession no.	Host annotated in NCBI	Region	Collection date
Asia1\1	Vaccine IND 63/72	AY304994.1		India	2003
Asia1\2	IND 334-00	DQ989304.1	Cattle	India	2000
Asia1\3	IND 151-94	DQ989303.1	Cattle	India	1993
Asia1\4	IND 81-86	DQ989306.1	Cattle	India	1986
Asia1\5	Asia1-2	AY593796.1		Israel	1963
Asia1\6	IND 397-97	DQ989308.1	Cattle	India	1997
Asia1\7	Asia1/YS/CHA/05	GU931682.1	Bovine	China	2005
Asia1\8	ZB/CHA/58(att)	DQ533483.1	Cattle	China	1958
Asia1\9	Asia1/MOG/05	EF614458.1	Cattle	Mongolia	2005
Asia1\10	YNBS/58	AY390432.1	Cattle	China	1958
Asia1\11	Asia1/Jiangsu/China/2005	EF149009.1	Bovine	China	2005
Asia1\12	Asia1/Jiangsu/China/2005	EF149009.1	Bovine	China	2005
Asia1\13	Asia1leb4 iso4	AY593799.1		Lebanon	1983
Asia1\14	Asia1leb-89 iso89	AY593798.1		Lebanon	1983
Asia1\15	Asia1leb83 iso28	AY593800.1		Lebanon	1983
Asia1\16	Asia1/WHN/CHA/06	FJ906802.1	Pig	China	2002
Asia1\17	IND 116-90	DQ989305.1	Cattle	India	1990
Asia1\18	asia1-1pak iso3	AY593795.1	Cattle	Pakistan	1954
Asia1\19	Asia1/VN/QT03/2007	GU125645.1	Cattle	Viet Nam	2007
Asia1\20	IND 82-96	DQ989309.1	Cattle	India	1996
Asia1\21	IND 247-92	DQ989307.1	Cattle	India	1992
Asia1\22	IND 491/97; WBN 117/85	AY687344.1		India	1985
Asia1\23	IND 52-87	DQ989313.1	Buffalo	India	1986
Asia1\24	IND 13-91	DQ989312.1	Sheep	India	1990
Asia1\25	IND 148-01	DQ989317.1	Buffalo	India	2000
Asia1\26	IND 47-93	DQ989315.1	Cattle	India	1993
Asia1\27	Asia1/HNK/CHA/05	EF149010.1	Bovine	China	2005
Asia1\28	IND 423-01	DQ989319.1	Cattle	India	2001
Asia1\29	IND 321/01	AY687333.1	Cattle	India	2001
Asia1\30	IND 97-03	DQ989323.1	Cattle	India	2002
Asia1\31	IND 37-02	DQ989311.1	Buffalo	India	2002
Asia1\32	IND 139-02	DQ989322.1	Cattle	India	2002
Asia1\34	IND 354-01	DQ989314.1	Cattle	India	2001
Asia1\35	IND 438-01	DQ989321.1	Cattle	India	2001
Asia1\36	IND 182-02	DQ989320.1	Cattle	India	2002
Asia1\37	IND 101-99	DQ989310.1	Cattle	India	1999
Asia1\38	MAY/9/99	HQ632774.1	Cattle	Malaysia	1999
Asia1\41	IND 61-02	DQ989318.1	Cattle	India	2002

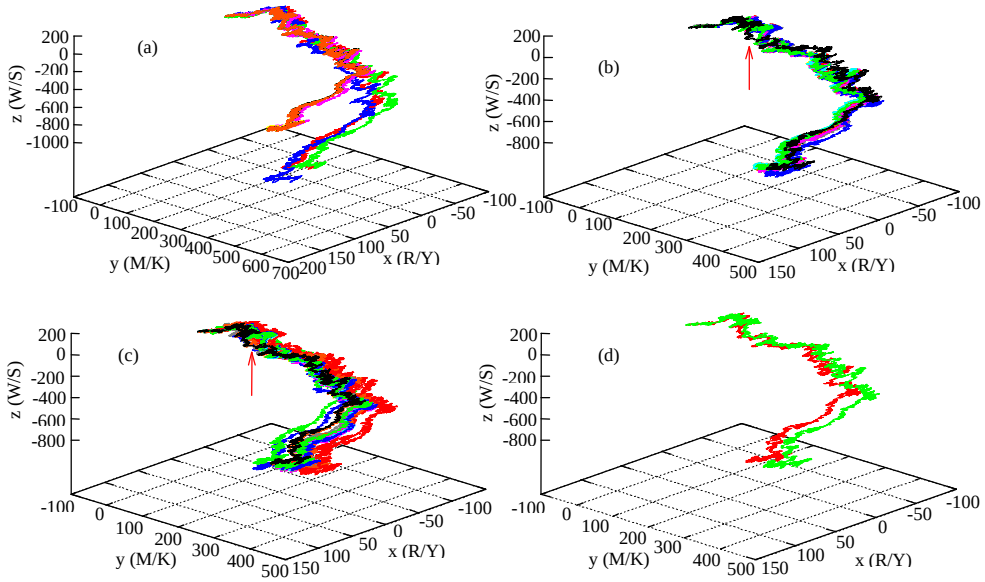

Fig. 2: a-d) The 3-dimensional curves of O serotype FMDV isolated in swine

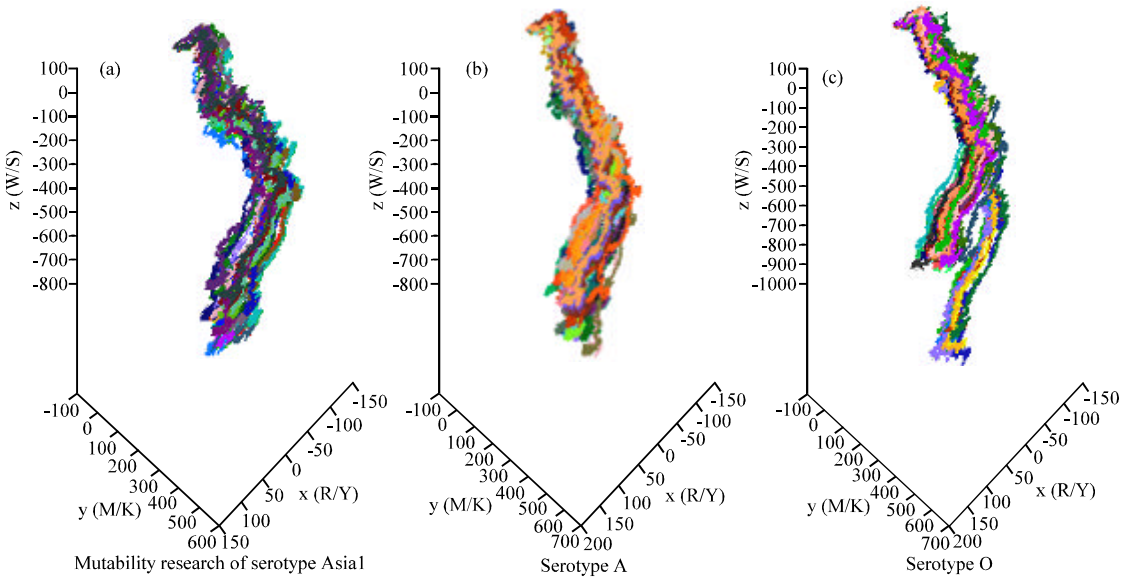

Fig. 3: a-c) The 3-dimensional curves of Asia1, A and O serotypes FMDV. The O serotype FMDV is most variable

Table 3: The Z-curve plot marks, names, accession numbers, hosts as well as isolation regions and collection dates of A serotype FMDV shown in Fig. 3b

Mark	Strain annotated in NCBI	Accession no.	Host annotated in NCBI	Region	Collection date
a\1	a10holland iso82	AY593751.1		Netherlands	1942
a\2	A4 West Germany/72	AY593779.1		Germany	1972
a\3	a4 W Germany iso42	AY593777.1		Germany	2004
a\4	a2spain iso7	AY593774.1		Spain	1969
a\6	a1bayern iso41	AY593759.1		Germany	1971
a\7	a3mecklenburg iso81	AY593776.1		Germany	1968
a\8	a12valle 119 iso20	AY593752.1		United Kingdom	1932
a\10	a24cruzeiro iso71	AY593768.1		Brazil	1955
a\11	a5westerwald iso73	AY593781.1		Germany	1951
a\12	A8 Parma aparma iso55	AY593792.1		Italy	1962
a\13	a5allier iso45	AY593780.1		France	1960
a\14	a20ussr iso10	AY593760.1		Russia	1964
a\15	a14 spain iso39	AY593754.1		Spain	1959
a\16	a4spain iso62	AY593778.1		Spain	2005
a\17	acanefa iso48	AY593789.1		Argentina	1961
a\18	a25 argentina iso38	AY593769.1		Argentina	1959
a\19	a17 Aguarulbos iso83	AY593757.1		Brazil	1967
a\20	a bage iso63	AY593787.1		Brazil	1977
a\21	a venceslau iso70	AY593803.1		Brazil	1979
a\22	a18zulia iso40	AY593758.1		Venezuela	1967
a\23	a13brazil iso75	AY593753.1		Brazil	1958
a\24	a24 argentina iso9	AY593767.1		Argentina	1965
a\25	a27columbia iso78	AY593771.1		Colombia	1967
a\26	a16belem iso80	AY593756.1		Brazil	1959
a\27	a29peru iso37	AY593773.1		Peru	1969
a\28	a32ven iso36	AY593775.1		Venezuela	1970
a\29	IND17/82	HM854024.1	Cattle	India	1982
a\30	a sabana iso68	AY593794.1		Colombia	1985
a\31	a26arg iso74	AY593770.1		Argentina	1966
a\32	a brazil iso67	AY593788.1		Brazil	1958
a\33	aphilippines iso50	AY593793.1		Philippines	1975
a\34	aargp64 iso100	AY593785.1		Argentina	2001
a\35	aargp55 iso99	AY593784.1		Argentina	2001
a\36	a general lopez iso102	AY593790.1		Argentina	2001
a\37	aarg2001 iso93	AY593783.1		Argentina	2001
a\38	aarg Trenquelauquen iso103	AY593786.1		Argentina	2001
a\39	A uruguay 2001 iso98	AY593802.1		Uruguay	2001

Table 3: Continue

Mark	Strain annotated in NCBI	Accession no.	Host annotated in NCBI	Region	Collection date
a\40	A30 Uruguay/68 iso90	AY593801.1		Uruguay	1968
a\41	IND17/77	HM854022.1	Cattle	India	1977
a\42	a28 Turkey iso44	AY593772.1		Turkey	1972
a\43	IND 109/2006	HQ832589.1	Cattle	India	2006
a\44	a argentina 2000 iso104	AY593782.1		Argentina	2000
a\45	IND 21/1990	HQ832576.1		India	1990
a\46	a23kenya iso8	AY593766.1		Kenya	1965
a\47	a21kenya iso77	AY593761.1		Kenya	1964
a\48	IND 110/1999	HQ832577.1		India	1999
a\49	IND258/99	HM854023.1	Cattle	India	1999
a\50	IND 818/2003	HQ832580.1	Cattle	India	2003
a\51	IND 281/2003	HQ832579.1	Buffalo	India	2003
a\52	IND 43/2006	HQ832586.1		India	2006
a\53	IND 245/2007	HQ832590.1	Cattle	India	2007
a\54	IND40/00	HM854025.1	Bovine	India	1999
a\55	IND 447/2005	HQ832583.1		India	2005
a\57	IND 161/2003	HQ832578.1		India	2003
a\58	c1ober iso88	AY593805.1		Germany	1960
a\59	IND 26/2006	HQ832585.1		India	2005
a\60	IND81/00	HM854021.1	Buffalo	India	2000
a\61	a22iraq-95 iso95	AY593762.1		Iraq	1964
a\62	a22iraq64 iso86	AY593763.1		Iraq	1964
a\63	IND 50/2006	HQ832587.1	Bullock	India	2005
a\64	a22iraq70 iso92	AY593764.1		Iraq	1970
a\65	IND 88/2006	HQ832588.1	Cattle	India	2005
a\66	IND 64/2004	HQ832581.1	Cattle	India	2004
a\68	IND 249/2004	HQ832582.1	Buffalo	India	2004
a\69	a22turkey iso66	AY593765.1		Turkey	1965
a\72	IND 437/2008	HQ832591.1	Cattle	India	2008
a\73	IND 17/2009	HQ832592.1	Cattle	India	2009
a\74	a15thailand iso43	AY593755.1		Thailand	1960
a\76	IND 22/2006	HQ832584.1	Cattle	India	2005
a\77	airan iso105	AY593791.1		Iran	1998
a\79	A/VN/03/2009	GQ406249.1	Cattle	Viet Nam	2009
a\82	MAY/3/2007	HQ632773.1	Cattle	Malaysia	2007
a\84	A/VN/16/2009	GQ406251.1	Pig	Viet Nam	2009
a\85	A/VN/02/2009	GQ406248.1	Cattle	Viet Nam	2009
a\86	A/VN/11/2009	GQ406250.1	Cattle	Viet Nam	2009
a\88	A/VN/20/2009	GQ406252.1	Buffalo	Viet Nam	2009
a\91	A/VN/09/2009	GQ406247.1	Cattle	Viet Nam	2009

Table 4: The Z-curve plot marks, names, accession numbers, hosts as well as isolation regions and collection dates of O serotype FMDV shown in Fig. 3c

Mark	Strain annotated in NCBI	Accession no.	Host annotated in NCBI	Region	Collection date
o\1	Serotype="O"	AF308157.1	Swine	Taiwan	1997
o\2	Otaiwan97 iso106/112	AY593835.1		Taiwan	1997
o\3	Openghu iso108	AY593833.1		Taiwan	1999
o\6	O/YM/YN/2000	HQ412603.1	Sus scrofa	China	2000
o\7	HKN/2002	AY317098.1		China	2002
o\8	WFL	EF175732.1		China	
o\9	lz	DQ248888.1		China	
o\10	O/ES/2001	AY686687.1		China	2001
o\11	MAY/8/2005	HQ632771.1	Pig	Malaysia	2005
o\12	O/HK/2001	EU400597.1		China	2001
o\13	o5india iso34	AY593828.1		India	1962
o\14	Akesu/58	AF511039.1	Cattle epithelial blister	China	1958
o\15	OGBF15	DQ478936.1	Cattle	China	
o\16	OMIII artificially attenuated from strain Akesu/58	AY359854.1		China	1958
o\17	OGBF15 derivative	DQ478937.1		China	
o\18	o3venezuela iso15	AY593827.1		Venezuela	1971
o\19	o1valle iso64	AY593825.1		Argentina	1939
o\20	o2brescia iso17	AY593826.1		Italy	1947
o\21	Tibet/CHA/99	AJ539138.1		China:Tibet	1999
o\22	o1manisa iso87	AY593823.1		Turkey	1969
o\23	o1m11 iso57	AY593822.1			
o\24	China/1/99(Tibet)	AF506822.2		China:Tibet	

Table 4: Continue

Mark	Strain annotated in NCBI	Accession no.	Host annotated in NCBI	Region	Collection date
o\25	TAW/2/99 BOV	AJ539137.1		Taiwan	
o\26	TAW/2/99 TC	AJ539136.1		Taiwan	
o\27	SAR/19/2000	AJ539140.1		South Africa	
o\28	UKG/4141/2001	EF552689.1	Sheep	United Kingdom	2001
o\30	MAY/1/2004	HQ632770.1	Cattle	Malaysia	2004
o\31	UKG/1734/2001	FJ542368.1	Cattle	United Kingdom	2001
o\32	UKG/126/2001	DQ404179.1	Porcine	United Kingdom	2001
o\33	UKG/127/2001	DQ404178.1	Porcine	United Kingdom	2001
o\34	UKG/128/2001	DQ404177.1	Porcine	United Kingdom	2001
o\35	UKG/150/2001	DQ404176.1	Bovine	United Kingdom	2001
o\36	FRA/1/2001	AJ633821.1	Bovine	France:Mayenne	2001
o\37	UKG/35/2001	AJ539141.1		United Kingdom	2001
o\38	UKG/7299/2001	EF552692.1	Sheep	United Kingdom	2001
o\39	UKG/11/2001	DQ404180.1	Porcine	United Kingdom	2001
o\40	PAK/45/2008	GU384683.1	Buffalo	Pakistan	2008
o\41	PAK/44/2008	GU384682.1	Cattle	Pakistan	2008
o\42	UKG/2085/2001	FJ542370.1	Cattle	United Kingdom	2001
o\43	UKG/1558/2001	FJ542367.1	Cattle	United Kingdom	2001
o\44	UKG/7039/2001	EF552690.1	Cattle	United Kingdom	2001
o\45	UKG/3952/2001	EF552688.1	Sheep	United Kingdom	2001
o\46	UKG/9327/2001	DQ404167.1	Ovine	United Kingdom	2001
o\47	UKG/9788/2001	DQ404166.1	Bovine	United Kingdom	2001
o\48	UKG/9964/2001	DQ404165.1	Bovine	United Kingdom	2001
o\50	UKG/8098/2001	EU214601.1	Sheep	United Kingdom	2001
o\51	UKG/4014/2001	EF552693.1	Cattle	United Kingdom	2001
o\52	UKG/220/2001	DQ404173.1	Bovine	United Kingdom	2001
o\53	UKG/7038/2001	DQ404169.1	Ovine	United Kingdom	2001
o\54	UKG/9011/2001	DQ404168.1	Bovine	United Kingdom	2001
o\55	MAY/3/2000	HQ632768.1	Pig	Malaysia	2000
o\56	UKG/1450/2001	FJ542366.1	Sheep	United Kingdom	2001
o\57	UKG/417/2001	FJ542365.1	Cattle	United Kingdom	2001
o\58	UKG/5470/2001	EF552696.1	Cattle	United Kingdom	2001
o\59	UKG/4998/2001	EF552694.1	Cattle	United Kingdom	2001
o\60	UKG/173/2001	DQ404175.1	Bovine	United Kingdom	2001
o\61	UKG/438/2001	DQ404174.1	Ovine	Ireland	2001
o\62	UKG/4569/2001	DQ404171.1	Ovine	United Kingdom	2001
o\63	UKG/7675/2001	DQ404170.1	Ovine	United Kingdom	2001
o\64	UKG/11676/2001	DQ404164.1	Bovine	United Kingdom	2001
o\65	O UK2001-FB	AY593832.1		United Kingdom	2002
o\66	O UK2001-ED	AY593831.1		United Kingdom	2002
o\67	UKG/2526/2001	FJ542371.1	Cattle	United Kingdom	2001
o\68	UKG/2000/2001	FJ542369.1	Sheep	United Kingdom	2001
o\69	UKG/621/2001	DQ404172.1	Bovine	United Kingdom	2001
o\70	UKG/14476/2001	DQ404162.1	Ovine	United Kingdom	2001
o\71	UKG/14524/2001	DQ404160.1	Bovine	United Kingdom	2001
o\72	O PanAsia uk2001	AY593836.1	Cattle	United Kingdom	2001
o\73	UKG/2640/2001	FJ542372.1		United Kingdom	2001
o\74	UKG/5681/2001	EF552697.1	Sheep	United Kingdom	2001
o\75	UKG/9443/2001	EF552695.1	Sheep	United Kingdom	2001
o\76	UKG/9161/2001	EF552691.1	Cattle	United Kingdom	2001
o\77	UKG/14391/2001	DQ404161.1	Bovine	United Kingdom	2001
o\78	Israel 07-6380	FJ175662.1	Cattle	Israel	2007
o\79	o1skr iso85	AY593824.1		South Korea	2000
o\80	Israel 07-6389	FJ175664.1	Cattle	Israel	2007
o\81	Israel 07-6378	FJ175661.1	Goat	Israel	2007
o\82	UKG/14603/2001	DQ404159.1	Bovine	United Kingdom	2001
o\83	UKG/14339/2001	DQ404163.1	Bovine	United Kingdom	2001
o\84	SKR/2000	AJ539139.1		South Korea	
o\90	UKG/15101/2001	DQ404158.1	Ovine	United Kingdom	2001
o\91	Israel 07-6382	FJ175663.1	Cattle	Israel	2005
o\92	Israel 07-6391	FJ175665.1	Cattle	Israel	2007
o\93	orey-iran iso53	AY593834.1		Iran	1966
o\97	o10phil76 iso76	AY593812.1		Philippines	1958
o\98	o10phil54 iso54	AY593811.1		Philippines	1958
o\100	"O/SKR/14/02"	EF614457.1	Cattle	South Korea	2002

Table 5: The Z-curve marks of those FMDV plotted in Fig. 1c, d, Fig. 2a-c, Fig. 4a and c. The strain names, accession numbers, hosts as well as isolation regions and collection dates are shown in Table 2-4

Fig. 1c	Fig. 1d	Fig. 2a	Fig. 2b	Fig. 2c	Fig. 4a	Fig. 4c
o\32	o\30	o\11	o\28	o\14	Asia1\2	a\29
o\33	o\41	o\6	o\45	o\15	Asia1\3	a\41
o\34	o\14	o\11	o\50	o\30	Asia1\4	a\43
o\39	o\15	o\55	o\56	o\31	Asia1\6	a\49
o\53	o\78	o\32	o\68	o\40	Asia1\17	a\50
o\61	o\80	o\33	o\38	o\41	Asia1\20	a\51
o\62	o\100	o\34	o\46	o\42	Asia1\23	a\53
o\63	o\31	o\39	o\53	o\43	Asia1\24	a\54
o\47	o\42		o\61	o\44	Asia1\25	a\60
o\48	o\43		o\62	o\47	Asia1\26	a\61
o\52	o\89		o\63	o\48	Asia1\28	a\63
o\53	o\91		o\70	o\51	Asia1\30	a\65
			o\90	o\52	Asia1\31	a\66
			o\81	o\53	Asia1\32	a\72
				o\57	Asia1\34	a\73
				o\58	Asia1\35	a\76
				o\60	Asia1\36	
				o\78	Asia1\37	
				o\80	Asia1\41	
				o\89		
				o\100		

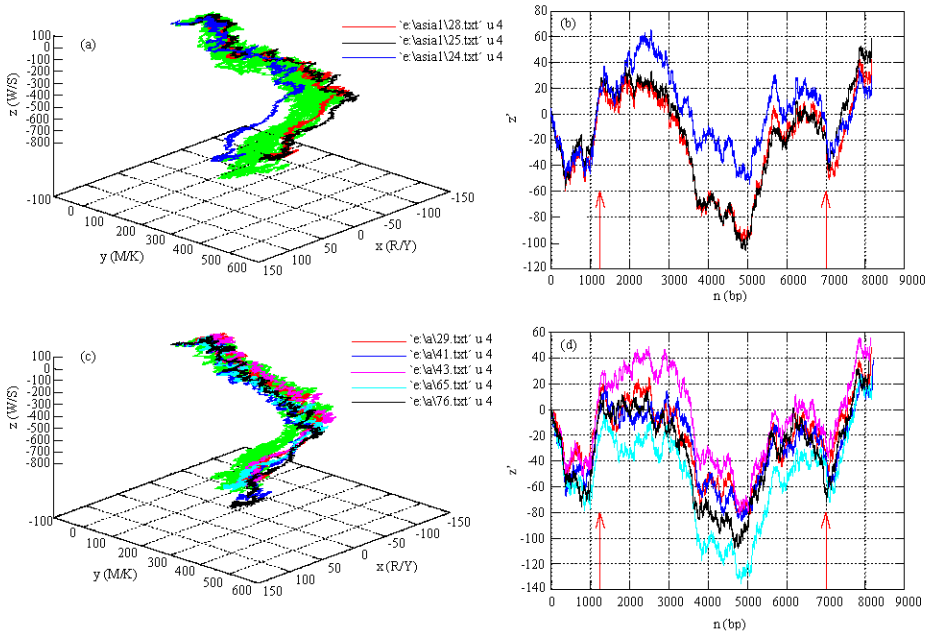

Fig. 4: a-d) The 3-dimensional curve of Asia1 serotype FMDV occurred in India before 2002

What's more, the O serotype FMDV are distinctly divided into two clusters according to the 3-dimensional curves (Fig. 3c).

Study on the Asia1 and A serotype FMDV which occurred in India: Statistics in Table 4 and 5 show that the mainly epidemic FMDV in India before 2002 were Asia1 serotype whereas it changed to A serotype after 2002. They are extracted out and conducted Z-curve analysis (Table 3-5). The results illustrate that the Asia1 serotype FMDV strains which were epidemic in India before 2002 varied too much (Fig. 4a, b). The A serotype FMDV strains occurred in India after 2002 are more stable and have a closer phylogenetic relationships compared with the Asia1 strains which were prevalent before 2002 (Fig. 4c, d). The variation regions exactly locate in the protein coding areas and the non-coding region is very conservative (Fig. 4b, d). Figure 4c shows the 3-dimensional curves of A serotype FMDV occurred in India after 2002. As can be observed, these curves can be divided into a similarity cluster and a dispersity cluster. Researchers plot the Z-curves of these dispersity clusters,

respectively in Fig. 4b, d. Results indicate that the A serotype strains occurred in India after 2002 are more stable and have a closer phylogenetic relationships compared with the Asia1 strains which were prevalent before 2002. What's more, the variation regions exactly locate in the protein coding areas.

FMDV has been considered as the most economically severe animal virus with a remarkable degree of antigenic diversity (Damien and Fares, 2006). It has a high mutation rate and spontaneous mutants which may be readily isolated in the laboratory (McCahon *et al*., 1977). The biological significance of the rapid evolution of RNA viruses have become increasingly obvious in systems such as infections by influenza virus, poliovirus and vesicular stomatitis virus (Holland *et al*., 1982; Holland, 1984). As is known to all, the genetic diversity of FMDV is a consequence of the high mutation rate due to the error-prone RNA poly-merase lacking proof reading activity. It has been recognized that FMDV is undergoing a rapid evolution in the field situation (Kumar *et al*., 2004). FMDV evolution is strongly influenced by high mutation rates and a quasispecies dynamics (Domingo *et al*., 2003). As argued in this study, geographic location plays a key role in the evolution of FMDV. The variability in FMDV (RNA) caused by high errors rates during their replication seems to be crucial for the survival of virus populations when the environmental conditions are altered. Damien and Fares (2006) suggested that serotypes with wide geographical distribution have accumulated compensatory mutations as a strategy to ameliorate the effect of slightly deleterious mutations fixed by genetic drift. Also, we find that host play an important role in the variation of FMDV. Researchers have proved that selection exerted by host antibody also plays a major role in the rapid evolution of FMDV Asia1 as observed in other serotypes. What's more, researchers find that compared with host, geographic location have a grater influence on FMDV mutation.

In this study, researchers find that the most variable regions of FMDV lie in protein coding areas and the non-coding region is very conservative. Insights into viral RNA sequence conservation and variability in nature will likely impact the understanding of FMDV infections, host range and transmission (Carrillo *et al*., 2005). As can be observed in Z-curve, researchers know that Z-curve can also analyze the structural proteins and non-structural proteins coding genes. Researchers know that VP1 is a structural protein of FMDV. It has been long recognized that the *VP1* gene differs in about 30-50% between serotypes (Knowles and Samuel, 2003). It is recognized that non-structural proteins, crucial roles for viral propagation are more conserved than the structural proteins. Most mutations or deletions in the non-structural proteins could be detrimental to viral replication and protein processing. Converged mutations in structural proteins may help FMDV to evade immune response mounted by the host while maintaining the functional capsid. Consequently, the gene sequences of the structural proteins could mutate with a higher frequency for retaining evolutional advantages (Feng *et al*., 2004).

In addition, researchers find Z-curve is also helpful to study the molecular epidemiology of FMDV. The use of molecular epidemiology is an important tool in understanding and consequently controlling FMDV (Joern, 2009). As mentioned before, Israel 07-6378 (goat) and PAK/45/2008 (buffalo) are very similar while they are different from others. So, it indicates that there is a variation trend in buffalo and goat of O serotype FMDV. If we take the collection time and isolation region of them into consideration, we should ponder whether or not there is a relationship between their occurrence. If there is, how does the virus transmit from Israel to Pakistan, how does it conquer the species diversity from goat to buffalo, etc. Researchers also analyzed the mainly FMDV prevalent in India, researchers find that the A serotype FMDV strains which occurred in India after 2002 are more stable and have a closer phylogenetic relationship than the Asia1 serotype FMDV strains which were epidemic before 2002. All these enlighten us that Z-curve is helpful to study the molecular epidemiology of FMDV. An increased understanding of how FMDV strains move between geographic regions will play a pivotal role in the development of future disease control strategies (Knowles and Samuel, 2003).

In this study, researchers can see that Z-curve also show the ability of homology analysis. According to Fig. 3c, the O serotype FMDV are distinctly divided into two clusters. What's more, researchers find that all serotypes FMDV have a close consanguinity but each has its own characteristic by contrast, the three South Africa FMDV have a much closer affinity. Researchers can distinguish the kinship from the overlap extent and the similarity of 3-dimensional Z-curves and Z-curves. However, this kind of means is only appropriate for analyzing a small quantity of data. We should explore more about Z-curve if we want to analyze large amount of data. Such as measuring the differences among the 3-dimensional Z-curves.

Zheng *et al*. (2005) use the geometric center and the eigenvectors to describe evolutionarily of FMDV. They constructed a phylogenetic tree of the analyzed FMDV and found that it is consistent with those of previous analyses (Zheng *et al*., 2005).

CONCLUSION

In general, Z-curve Method shows a widely application prospect in the areas of life sciences. Such as gene sequence, molecular epidemiology, homology, evolution, phylogenetic relationships and genetic diversity analysis. In addition, the Z-curve also can be used to recognize protein coding genes. Z-curve_E, a new self-training system derived from Z-curve has been used to recognize protein coding genes (Guo and Zhang, 2006). Currently, the Z-curve Method is still in its premature stage. Since, the more information is extracted from the Z-curve, the more accurate result can be gained, researchers are expected to explore more about the Z-curve. Researchers look forward to exploit some novel and more effective algorithms to extract information contained in the Z-curves.

ACKNOWLEDGEMENTS

This study was financially supported by the National High Technology Research and Development Program of China (863 Program) (No. 2011AA10A211) and Ministry of Agriculture' Special Funds for Scientific Research on Public Causes (No. 201203039).

REFERENCES

Alexandersen, S. and N. Mowat, 2005. Foot-and-mouth disease: Host range and pathogenesis. Curr. Top. Microbiol. Immunol., 288: 9-42.

Barasa, M., A. Catley, D. Machuchu, H. Laqua, E. Puot, D.T. Kot and D. Ikiror, 2008. Foot-and-mouth disease vaccination in South Sudan: Bene?t-cost analysis and livelihoods impact. Transbound. Emerg. Dis., 55: 339-351.

Carrillo, C., E.R. Tulman, G. Delhon, Z. Lu and A. Carreno *et al*., 2005. Comparative genomics of foot-and-mouth disease virus. J. Virol., 79: 6487-6504.

Chu, J.H., S.T. Weiss, V.J. Carey and B.A. Raby, 2009. A graphical model approach for inferring large-scale networks integrating gene expression and genetic polymorphism. BMC. Syst. Biol. Vol. 33. 10.1186/1752-0509-3-55

Cottam, E.M., J. Wadsworth, A.E. Shaw, R.J. Rowlands and L. Goatley *et al*., 2008. Transmission pathways of foot-and-mouth disease virus in the United Kingdom in 2007. PLoS. Pathog. Vol. 4.

Damien, C.T. and M.A. Fares, 2006. Unravelling selection shifts among foot-and-mouth disease virus (FMDV) serotypes. Evol. Bioinform. Online, 2: 211-225.

Davies, G., 2002. Foot and mouth disease. Res. Vet. Sci., 73: 195-199.

Domingo, E., C. Escarmis, E. Baranowski, C.M. Ruiz-Jarabo, E. Carrillo, I.J. Nunez and F. Sobrino, 2003. Evolution of foot-and-mouth disease virus. Virus Res., 91: 47-63.

Feng, Q., H. Yu, Y. Liu, C. He and J. Hu *et al*., 2004. Genome comparison of a novel foot-and-mouth disease virus with other FMDV strains. Biochem. Biophys. Res. Commun., 323: 254-263.

Giacomo, F., M.M. Francesco, C. Davide and H. Muller, 2007. Graph-based identification of cancer signaling pathways from published gene expression signatures using PubLiME. Nucleic Acids Res., 35: 2343-2355.

Gibbs, P., 2003. The foot-and-mouth disease epidemic of 2001 in the UK: implications for the USA and the J. Vet. Med. Educ., 30: 121-132.

Guo, F.B. and C.T. Zhang, 2006. ZCURVE_V: A new self-training system for recognizing protein-coding genes in viral and phage genomes. BMC Bioinformatics, Vol. 7.

Guo, F.B. and X.J. Yu, 2007. Separate base usages of genes located on the leading and lagging strands in *Chlamydia muridarum* revealed by the Z curve method. BMC Genomics, 8: 1471-2164.

Guo, F.B., H.Y. Ou and C.T. Zhang, 2003. ZCURVE: A new system for recognizing protein-coding genes in bacterial and archaeal genomes. Nucleic Acids Res., 31: 1780-1789.

Holland, J.J., 1984. Continuum of Change in RNA Virus Genomes. In: Concepts in Viral Pathogenesis, Notkins, A.L. and M.B.A. Oldstones (Eds.). Spinger Wien, New York, pp: 137-143.

Holland, J.J., J.C. De La Torre and D.A. Steinhauer, 1982. RNA virus populations as quasispecies. Curr. Top. Microbiol. Immunol., 176: 1-20.

Joern, K., 2009. Understanding the molecular epidemiology of foot-and-mouth-disease virus. Infect. Genet. Evol., 9: 153-161.

Knowles, N.J. and A.R. Samuel, 2003. Molecular epidemiology of foot and mouth disease virus. Virus. Res., 91: 65-80.

Kumar, R.M., A. Sanyal, D. Hemadri, C. Tosh, J.K. Mohapatra and S.K. Bandyopadhyay, 2004. Characterization of foot-and-mouth disease serotype asia1 viruses grown in the presence of polyclonal antisera in serology and nucleotide sequence analysis. Arch. Virol., 149: 1801-1814.

Li, X., C. Panea., C.H. Wiggins, V. Reinke and C. Leslie, 2010. Learning graph-mer motifs that predict gene expression trajectories in development. PLoS Comput. Biol. Vol. 6. 10.1371/journal.pcbi.1000761

McCahon, D., W.R. Slade, R.A.J. Priston and J.R. Lake, 1977. An extended genetic recombination map of foot-and-mouth disease virus. J. Gen. Virol., 35: 555-565.

Mort, M., I. Convery, J. Baxter and C. Bailey, 2005. Psychosocial effects of the 2001 UK foot-and-mouth disease epidemic in a rural population: Qualitative diary based study. BMJ Vol. 331.

Ou, H.Y., F.B. Guo and C.T. Zhang, 2003. Analysis of nucleotide distribution in the genome of streptomyces coelicolor A3(2) using the Z curve method. FEEBS. Lett., 540: 188-194.

Paprocka, G., 2006. Foot-and-mouth disease virus and its molecular structure. Med. Weter., 62: 753-756.

Perry, B.D. and K.M. Rich, 2007. Poverty impacts of foot-and-mouth disease and the poverty reduction implications of its control. Vet. Rec., 160: 238-241.

Rodriguez, L.L. and M.J. Grubman, 2009. Foot-and-mouth disease virus vaccines. Vaccine, 27: 90-94.

Rufael, T., A. Catley, A. Bogale, M. Sahle and Y. Shiferaw, 2008. Foot-and-mouth disease in the Borana pastoral system, southern Ethiopia and implications for livelihoods and international trade. Trop. Anim. Health. Pro., 40: 29-38.

Schumann, K.R., N.J. Knowles, P.R. Davies, R.J. Midgley and J.F. Valarcher *et al*., 2008. Genetic characterization and molecular epidemiology of foot-and-mouth disease viruses isolated from Afghanistan in 2003-2005. Virus Genes, 36: 401-413.

Sobrino, F., M. Saiz, M.A. Jimenez-Clavero, J.I. Nunez, M.F. Rosas, E. Baranowski and V. Ley, 2001. Foot-and-mouth disease virus: A long known virus, but a current threat. Vet. Res., 32: 1-30.

Sumption, K., M. Rweyemamu and W. Wint, 2008. Incidence and distribution of foot-and-mouth disease in Asia, Africa and South America, combining expert opinion, official disease information and livestock populations to assist risk assessment. Transbound. Emerg. Dis., 55: 5-13.

Thompson, D., P. Muriel, D. Russell, P. Osborne, A. Bromley, M. Rowland, T.S. Creigh, and C. Brown, 2002. Economic costs of the foot- and-mouth disease outbreak in the United Kingdom in 2001. Rev. Sci. Tech., 21: 675-687.

Wen, S.Y. and C.T. Zhang, 2003. Identification of isochore boundaries in the human genome using the technique of wavelet multiresolution analysis. Biochem. Biophys. Res. Commun., 311: 215-222.

Zhang, C.T. and F. Zhang, 2003a. An isochore map of the human genome based on the Z curve method. Gene., 317: 127-135.

Zhang, C.T. and J. Wang, 2000. Recognition of protein coding genes in the yeast genome at better than 95% accuracy based on the Z curve. Nucleic Acids Res., 28: 2804-2814.

Zhang, R. and C.T. Zhang, 2003b. Identification of genomic islands in the genome of Bacillus cereus by comparative analysis with *Bacillus anthracis*. Physiol. Genomics., 16: 19-23.

Zhang, R. and C.T. Zhang, 2005. Identification of replication origins in archaeal genomes based on the Z-curve method. Archaea, 1: 335-346.

Zheng, W.X., L.L. Chen, H.Y. Ou, F. Gao and C.T. Zhang, 2005. Coronavirus phylogeny based on a geometric approach. Mol. Phylogenet. Evol., 36: 224-232.