GENE ANALYSIS OF A NEWLY ISOLATED LASSA VIRUS STRAIN

[1]LE OKOROR*, [2]TH ALAIYA, [3]FI ESUMEH, AND R PAUL[4]

1,3,4 Department of Microbiology, Faculty of Natural Sciences, Ambrose Alli University, PMB 14, Ekpoma, Nigeria.
2, Department of Microbiology, Ambrose Alli University, Ekpoma Nigeria.

*Correspondence Email: Larison86@yahoo.com

Background

Lassa virus is the cause of lassa fever which has been a leading cause of morbidity and mortality in many parts of West Africa. It usually occurs in epidemic form where the spread could be more than HIV/AIDS in seasons of epidemics. Recently, two new strains of the virus were isolated from Ekpoma Nigeria. Gene analysis was carried out with strain "Nig04-02" and the function prediction of the corresponding protein. There have been no reason adduced for these yearly outbreak of lassa virus infection despite high level of circulating antibodies in the population. And still no vaccine in sight.

Method
Sequence alignment of the gene predicted significant hits of between 85% to 45% with matrix set at PAM50, PAM100 and blosum 62 using blast and wu-blast tools. FGENESH, GENEID, GENSCAN was used to predict gene function Global alignment was done with wu-blast. While homology and multiple sequence alignment was carried out using clustal W using both high scoring sequences and low scoring sequences. These were from mopeia and filovirus. The best score was predicted using bestscor at biology workbench. Transcription factors was predicted with TFsearch. And the likely gene product predicted with amigo. Protein sequence alignment was carried out with blastp.

Result
Sequence alignment (global) gave significant hits at PAM50, PAM100 and BLOSUM62 matrices which were between 85% to 45% homology. FGENESH gave no relevant prediction, while GENSCAN predicted 43.48% C+G and Isochore 2 (43-51 C+G%) and GENEID predicted a CDS1, CDSf, TSS, CDSi, PolA regions. All predictions were done at probability of >95%. There were 40 transcription factors with a score of 100% and six open reading frames. The gene ontology amigo predicted the gene product to be that of low molecular weight cystein rich protein with high score to *Arabinose thalia.* Blastp was able to predict high scoring proteins at PAM50 of which were most glycoprotein precursors of various viruses but at PAM100 a bacteria protein was hit. Results from clustalW reveal a high conservation with total alignment score of 16055 and highly conserved areas.

Discussion

Despite the presence of circulating antibodies against lassa virus in the population referred, there is still yearly out break of lassa virus epidemics. Hence it has become imperative to study new strains of the virus to actually know what would have been responsible for the yearly out break. This we did by carrying out a complete analysis of the Ng04-02 to determine if the difference in the new strain could be responsible for the recurrent out break of epidemics in most part of Nigeria. From the foregoing the virus showed great homology with very high scores both at the global alignment and the multiple sequence alignment. However there were area of differences with results from TF search when compared to another recently isolated lassa virus gene, though they both have 40 high scoring sites. We will then conclude that these differences in the transcription factors of the new strains is sufficient enough to make the new strain evade circulating antibodies as they may have alternative pathway for synthesis of gene products. This study also suggests that the gene products would have been responsible for disease causation and not the virus itself. The is because the virus would have naturally been wiped out by circulating antibodies who may not easily recognize the new strain and by the time of recognition, the viral gene product would have been produced through the unknown pathway. The new strain would have evolve in the course of another season. Hence we conclude here that recurrent out break of lassa fever is caused by the emergence of new strains unrecognized by circulating antibodies and also alternative pathway of production of gene products and/or new gene products since transcription factors differ, though all of similar or the same virus.

```
-------------------------AGGATTGCGCTTTTAGAGATCTTTG
lassa_1                              -------------------------
AGGATTGCGCTTTTAGAGATCTTTG
embl_DD323072_DD323072_Devel
GGGTCTTTTCTGCAGTCACCGTCGTCGACACGTGTGATCAGATATCGCGG
embl_AY772170_AY772170_Mopei         ----------------CGCCCTTGTGGATCC-
TAGGCTTTTTGGTTGC-G
                                                                  *
*      *    *

my_lassa2                            TGTGAGTGGGCCTCATCAAGCCATGGGACAAATC-
TAACATTTTTCCAGG
lassa_1
TGTGAGTGGGCTTCATCAAGCCATGGGGCAAATCATAACATTTTTCCAGG
embl_DD323072_DD323072_Devel
CCGCTCTAGAGATATCGCCGCCATGGGCCAGATCGTGACCTTCTTCCAGG
embl_AY772170_AY772170_Mopei         CATTTCTAGAG-
CATCTCGGAGATGGGGCAGATAGTCACCTTCTTTCAAG
                                               * *          *  ***** ** **  *
** ** ** ** *

my_lassa2
AAGTACCTCATGTCATAGAGGAAGTGATGAACATTGTCCTGATTGCTCTC
lassa_1
AAGTACCTCATGTCATAGAGGAAGTGATGAACATTGTCCTGATTGCTCTC
embl_DD323072_DD323072_Devel
AGGTGCCCCATGTGATCGAGGAGGTGATGAACATCGTGCTGATCGCCCTG
embl_AY772170_AY772170_Mopei
AGGTGCCACACATCCTTGAAGAAGTGATGAACATTGTGCTGATGACCCTC
```

```
                                        * ** ** **  *  * ** ** ********** **
*****  * **

my_lassa2
AGCCTTCTGGCAATACTGAAAGGGATCTACAATGTTGCTACCTGTGGCCT
lassa_1
AGCCTCCTGGCAATACTGAAGGGAATCTACAATGTTGCCACCTGTGGCCT
embl_DD323072_DD323072_Devel
AGCGTGCTGGCCGTGCTGAAGGGCCTGTACAACTTCGCCACCTGCGGCCT
embl_AY772170_AY772170_Mopei
TCAATCTTGGCCATCCTAAAGGGCATCTACAATGTGATGACCTGTGGAAT
                                          *  ****  * ** ** **  * *****  *
***** **  *

my_lassa2
CTTTGGATTGGTGTCCTTTCTCCTCTTATGTGGAAGATCATGT-TCAACA
lassa_1
CTTTGGATTGGTGTCCTTTCTCCTCTTATGTGGAAGATCATGC-TCAACA
embl_DD323072_DD323072_Devel
GGTGGGCCTGGTGACCTTCCTGCTGCTGTGCGGCAGGAGCTGCACCACCA
embl_AY772170_AY772170_Mopei
CATCGGTTTGATAACATTTTTGTTCTTGTGTGGGAGATCATGCTCAAGCA
                                      * **  ** *  * **  *  *  * ** ** **
**    * **

my_lassa2                           ACTT--
ACAAGGGTGTTTATGAGCTACAAACTCTAGAGCTAGACATGGCA
lassa_1                             ACTT--
ACAAGGGTGTTTATGAGCTACAAACTCTAGAGCTAGACATGGCA
embl_DD323072_DD323072_Devel
GCCTGTACAAGGGCGTGTACGAGCTGCAGACCCTGGAGCTGAACATGGAG
embl_AY772170_AY772170_Mopei        TCT---
ATAAGGACAACTATGAGTTCTTCTCTTTCGACCTCGACATGTCT
                                     *    * ****     ** *** *     *  * **
**  *****

my_lassa2                           AGTCTAAACATGACAATGCCCTTGTCT----------
-------------
lassa_1                             AACCTTAATATGACAATGCCCTTATCT----------
-------------
embl_DD323072_DD323072_Devel
ACCCTGAACATGACCATGCCCCTGAGCTGCACCAAGAACAACAGCCACCA
embl_AY772170_AY772170_Mopei
TCACTGAATGCAACGATGCCTCTCTCCTGCTCAAAGAACAACTCCCATCA
                                       ** **    ** *****  *
```

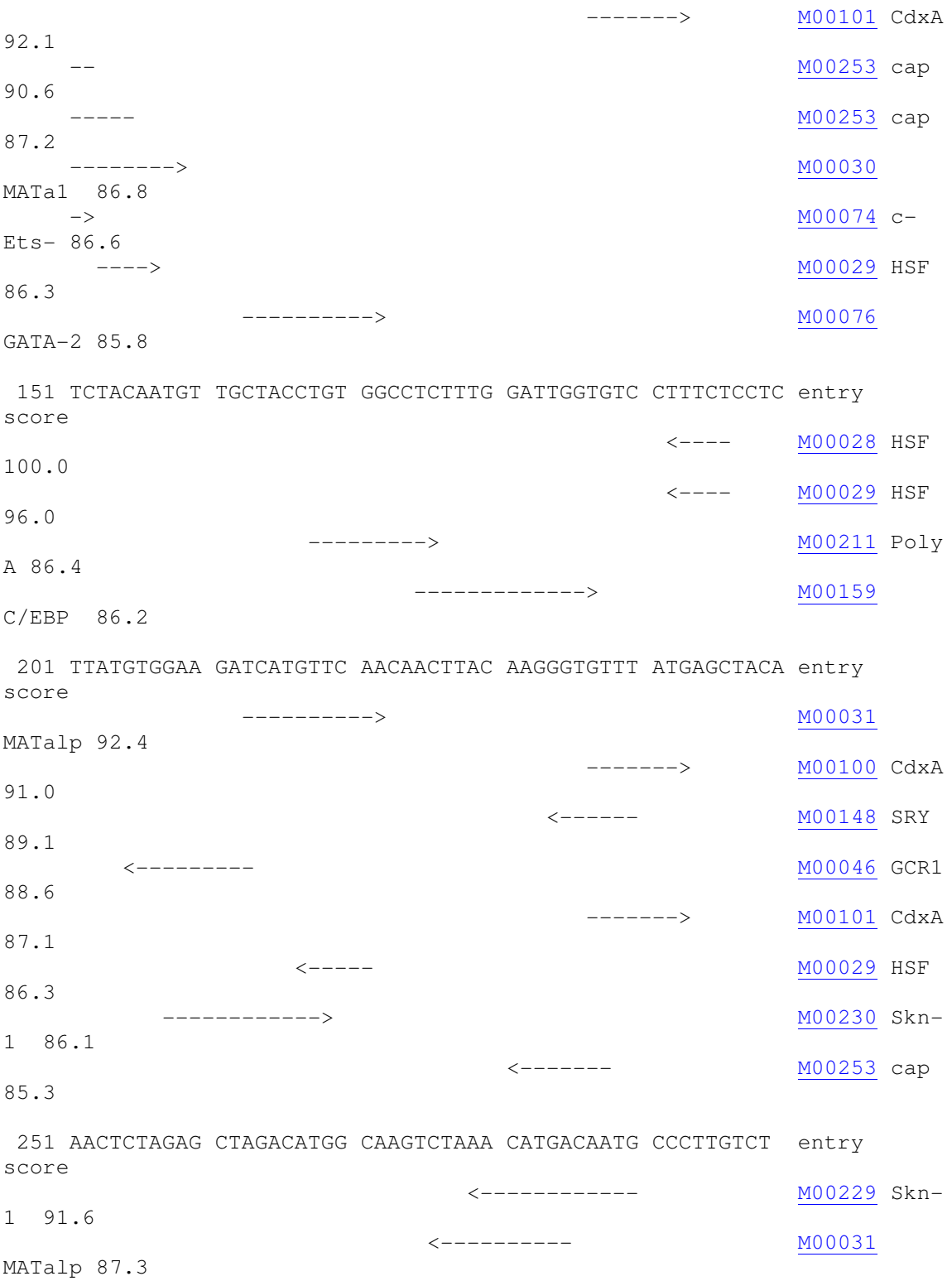
Fig I: Multiple sequence alignment using clustal W, areas in blue colouration are single conserved areas.

Fig II: TFMATRIX entries with High-scoring:

```
   1 AGGATTGCGC TTTTAGAGAT CTTTGTGTGA GTGGGCCTCA TCAAGCCATG entry
score
             <----------                                    M00216 TATA
90.4
                          ------>                           M00100 CdxA
87.2
                                            <----------     M00075
GATA-1 86.5

  51 GGACAAATCA TAACATTTTT CCAGGAAGTA CCTCATGTCA TAGAGGAAGT entry
score
                                          ---------->       M00031
MATalp 96.2
                       <---------                           M00223
STATx  94.2
                       --------->                           M00223
STATx  91.3
                          --------->                        M00032 c-
Ets- 91.2
                                                  <-----    M00253 cap
90.6
                                    <------                 M00271 AML-
1a 88.7
                                                    <--     M00253 cap
87.2
                     <-----                                 M00029 HSF
86.9
                                                      -     M00030
MATa1  86.8
                                             ------------   M00074 c-
Ets- 86.6
                                          ------------>     M00229 Skn-
1  86.5
          ------->                                          M00101 CdxA
86.4
                     <-----                                 M00028 HSF
85.9
                 ---------------------->                    M00259 STAT
85.5
                  <------                                   M00101 CdxA
85.0

 101 GATGAACATT GTCCTGATTG CTCTCAGCCT TCTGGCAATA CTGAAAGGGA entry
score
                               -------->                    M00253 cap
96.7
                                    <-----                  M00028 HSF
95.3
                                              <--------     M00253 cap
95.3
                                    <-----                  M00029 HSF
93.7
                 ---------->                                M00075
GATA-1 92.7
```



Total 40 high-scoring sites found.

Higgins, D.G., Bleasby, A.J. and Fuchs, R. (1992) CLUSTAL V: improved software for multiple sequence alignment. Computer Applications in the Biosciences (CABIOS), 8(2):189-191.

Thompson J.D., Higgins D.G., Gibson T.J. "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice." Nucleic Acids Res. 22:4673-4680(1994).

Felsenstein, J. 1989. PHYLIP -- Phylogeny Inference Package (Version 3.2). Cladistics 5: 164-166.