Translated Blast of L Polymerase as a Hit for Novel Arenaviruses Species

Hasan Alrashedi[1,3], Ian Jones[1], Geraldine Mulley[1], Benjamin W Neuman[1,2]

1) [1]School of Biological Sciences, University of Reading, Whiteknights, RG6 6AH, UK
2) [2] Biology Faculty, Texas A&M University-Texarkana, 7101 University Ave, Texarkana, TX 75503, USA
3) [3]College of Science, Department of Biology, Misan University, Misan, Iraq

Contact information: Benjamin W. Neuman: bneuman@tamut.edu or Hasan Alrashedi: h.s.h.alrashedi@student.reading.ac.uk

Abstract

Many pathogenic viruses can transmit between human and animals as zoonotic viruses and cause dangerous diseases with obvious clinical signs globally. However, the world deals seriously with these viruses when the viruses infected either human or animals especially if the infection were confirmed that classified as zoonotic (Lal et al. 2005). There are many viruses distribution in many countries around the world including Ebolavirus, Marburgvirus, SARS and MERS coronaviruses, Hendra, Nipah and arenavirus haemorrhagic fever viruses were categorized as zoonotic RNA viruses that cause epidemic in some regions such as African countries (Fichet-Calvet & Rogers 2009) (Ehichioya et al. 2010). Consequently, structural bioinformatics of virus protein like L polymerase of arenaviruses was used for monitor the future outbreak that could be happens by new species of viruses. At this research, significant similarities with hemorrhagic fever viruses including arenaviruses were found on GenBank database. Translated blast (tBLASTn) available on https://blast.ncbi.nlm.nih.gov/Blast.cgi was used for searching translated nucleotide databases using a protein query of arenavirus L polymerase (McGinnis & Madden 2004). At this research, the new and archival metazoan transcriptome sequence data of the new TSA species that available on NCBI was used for identification with arenaviruses genes. Therefore, structure bioinformatics was utilized for better understanding and predication the evolution and natural history of the pools of

uncharacterized virus on Genbank database that have led to emerging haemorrhagic fever in near future around the world.

Introduction

The family of *Arenaviridae* contains 27 within two genuses, *Mammarenavirus* and *Reptarenavirus* according to the most recent International Committee on Taxonomy of Viruses (ICTV) report (Adams et al. 2016). These viruses are distributing worldwide and cause febrile diseases for both humans and animals (Buchmeier et al. 2007). Mammalian arenaviruses of the genus *Mammarenavirus* are traditionally divided in two serogroups, the Old World arenaviruses, which contains the Lassa virus, and lymphocytic choriomeningitis virus (LCMV) serogroups and New World arenaviruses that consist of the *Tacaribe virus* serogroup (Clegg 2002). *Reptarenaviruses* including some arenaviruses that are pending formally to classify as *Reptarenaviruses* such as *Califoronia Academy of Science Viruses* (CASV) and *Golden Gate virus* (GGV). Arenaviruses have a segmented ambisense single strand RNA genome, which is typically grouped as negative sense RNA viruses based on polymerase homology and replication strategy (Buchmeier et al. 2007) (Torre 2009). The large segment named as L and the short segment named as S (Bodewes et al. 2013). The L segment encoded viral RNA dependent RNA polymerase (RdRp) or L polymerase and small zinc finger motif protein (Z), whereas the S segment encodes nucleoprotein (NP) and glycoprotein precursor (GPC) (Salvato et al. 1989). Viral RNA dependent RNA polymerase (L) is a largest protein in arenaviruses that reaches around 7.2 kb (Zapata & Salvato 2013) and contains endonuclease domain that paly significant role in mRNA cap-snatching during viral transcription (Morin et al. 2010).

At this research, structural bioinformatics online tools and some software were used for finding gene sequences of the living species that have homology with some arenaviruses species in GenBank database (Zhang et al. 2005) (Altman & Dugan 2005) (Clote & Backofen 2000) (Pevzner 2000) (Liljas, A. et al. 2001) (Gopakumar 2012) including *Tacaribe virus* (TCRV), *lymphocytic choriomeningitis virus* (LCMV), *Golden Gate virus* (GGV) and *California Academy of Science virus* (CASV) as shown in table 1. L protein of four arenaviruses include TCRV, LCMV, GGV and CASV were utilized for searching for translated nucleotide, as

transcriptome ShoutGun Assembly (TSA) of translated blast (tBLASTn) on NCBI facility was used for this purpose (McGinnis & Madden 2004). Then, ExPASy translated tools was used for identification open reading frames (ORFs) of translated protein for TSA species (RNA transcript) (Gasteiger et al. 2003) . Subsequently, the HHpred tools kits that available on https://toolkit.tuebingen.mpg.de/#/tools/hhpred was used for homology detection and structure prediction of the target proteins sequences (Söding et al. 2005). As the result, Bayesian tree were designed by MrBayes version 3.2.6 interface for novel/existence arenaviruses with some other viruses amino acids sequences to evaluation evolution history between the viruses (Ronquist et al. 2011) (Hall 2011). At least, FigTree v1.4.3 software available on http://tree.bio.ed.ac.uk/software/figtree/ was used for final presenting of the tree. Therefore, translated blast (tBLASTn) indicates that L protein of some species of arenaviruses has significant similarity with TSA alignment of other species on GenBank database that indicating for novel conserved gene of arenaviruses in archived online database facility.

Material and Methods:

Bioinformatics software and tools were used in this research. Online tools such as translated basic local alignment search tool known as translated blast (tBLASTn) on National Center for Biotechnology Information (NCBI) available online on https://blast.ncbi.nlm.nih.gov/Blast.cgi was used for searching translated nucleotide sequencing of species that available in GeneBank database were Transcriptome ShoutGun Assembly (TSA) was used for reciprocal BLAST (McGinnis & Madden 2004). Also, the open reading frames (ORFs) of TSA species were identified by the ExPASy translation tool (http://web.expasy.org/translate/) (Perry 2002). Additionally, online tool kite including HHpred (https://toolkit.tuebingen.mpg.de/#/tools/hhpred) were utilized for protein homology detection and structure prediction of ORFs of TSA species (Söding et al. 2005). Finally, MrBayes interface version 3.2.6 (Ronquist et al. 2011) was used for designing phylogenetic tree for L protein sequencing data including novel/existence genus of arenaviruses, while the FigTree v1.4.3 software on http://tree.bio.ed.ac.uk/software/figtree/ was used for viewing final phylogenetic tree for all viruses that utilized at this research.

Results

Translated blast (tBLASTn) of Arenaviruses

At this study, new living species that have some gene similarity with L protein of arenaviruses were identified from GenBank database. Translated basic local alignment search tool (blast) known as translated blast (tBLASTn) on National Center for Biotechnology Information (NCBI) that available on https://blast.ncbi.nlm.nih.gov/Blast.cgi was used for searching for translated nucleotide databases using a protein query (McGinnis & Madden 2004). RNA dependent RNA polymerase (RdRp) or (L) proteins sequences of two species of *Mammarenaviruses*, which are *Tacaribe virus* (TCRV) and *lymphocytic choriomeningitis virus* (LCMV) as well as two species of *Reptarenaviruses* that are *Golden Gate virus* (GGV) and *California Academy of Science virus* (CASV) were used in translated blast (tBLASTn) were Transcriptome ShoutGun Assembly (TSA) adjusted at searching set database. At least, six transcribed RNA sequences data of different species of living species including *Channa punctate, Catostomus commersonii, Asymmetron lucayanum, Rhizopus oryzae, Rhopilema esculentum* and *Talitrus saltator* (table 1) were matched in some genes of arenaviruses species that confirm some similarity with arenaviruses L protein were E. Value less than 1 are selected. Consequently, translated blast (tBLASTn) on NCBI was utilized for finding translated nucleotides based on arenaviruses proteins in GenBank database.

Table 1: Translated blast (tBLASTn) of L proteins of arenaviruses**.** Viral RNA dependent RNA polymerase **(**L) of TCRV: *Tacaribe virus* (GenBank accession number NP_694848.1), LCMV: *lymphocytic choriomeningitis virus* (NP_694845.1), GGV: *Golden Gate virus* (YP_006590089.1), CASV: *California Academy of Science virus* (YP_006590093.1) were used for searching translated nucleotide in GenBank database by using translated blast (tBLASTn) facility on National Center for Biotechnology Information (NCBI) were Transcriptome ShoutGun assembly (TSA) allocated at search set of tBLASTn.

Virus	Viral L protein accession	Transcribed RNA of TSA alignment	E-Value	Identity	Accession of TSA
TCRV	NP_694848.1	*Channa punctata* RP_73146	9e-05	41%	GEKU01073124.1
		Catostomus commersonii Contig_131767	0.001	51%	GECX01131681.1
LCMV	NP_694845.1	*Channa punctata* RP_73146	7e-06	43%	GEKU01073124.1
		Asymmetron lucayanum comp27398_c0_seq1	6e-05	26%	GETC01011646.1
		Catostomus commersonii Contig_163647	0.002	36%	GECX01163539.1
		Rhizopus oryzae KNUBEL_RO_010568	0.18	24%	GDUK01010546.1
GGV	YP_006590089.1	*Channa punctata* RP_73146	5e-15	44%	GEKU01073124.1
		Asymmetron lucayanum comp27398_c0_seq1	4e-04	24%	GETC01011646.1
		Catostomus commersonii Contig_131767	0.002	48%	GECX01131681.1
		Rhopilema esculentum c77127_g1_i1	0.068	23%	GEMS01099264.1
CASV	YP_006590093.1	*Channa punctata* RP_73146	8e-13	37%	GEKU01073124.1
		Catostomus commersonii Contig_131767	0.001	35%	GECX01131681.1
		Talitrus saltator comp102738_c0_seq1	0.005	27%	GDUJ01083694.1
		Asymmetron lucayanum comp27398_c0_seq1	0.11	24%	GETC01011646.1

As shown in table 1, viral RNA dependent RNA polymerase (L) protein of *Tacaribe virus* (TCRV), *lymphocytic choriomeningitis virus* (LCMV), *Golden Gate virus* (GGV) and *California Academy of Science virus* (CASV) were used for searching translated nucleotides on GenBank database were translated blast (tBLASTn) on NCBI used for this aim (Belshaw & Katzourakis 2005). Transcriptome ShoutGun assembly (TSA) result shows there are transcribe RNA sequences of six living species that are *Channa punctate, Catostomus commersonii, Asymmetron lucayanum, Rhizopus oryzae, Rhopilema esculentum* and *Talitrus saltator* that have gene similarity with L gene of arenaviruses were E.value less than 1 are preferred. The E. Value indicates the numbers of the hits during database searching that expected by chance when two or more sequences were aligned. The more significant E. Value when the parameter are less than 1 or zero, whilst the E. value equal to one is not consider since it means there is one chance for similarity between the hits (Pearson

1995). As a result, the target transcript RNA sequences (table 1) were used for open reading frames (ORFs) detection and then for protein translation, homology detection and structure prediction with other protein that found online databases. At least, all sequences data including translated protein of TSA species and the sequences of *Mammarenaviruses* (TCRV and LCMV) and *Reptarenaviruses* (GGV and CASV) used for multiple sequence alignment (MSA) and finally for designing Bayesian phylogenetic trees. So, tBLASTn confirmed that there are many living species that have some genes homology with L gene of *Arenaviridea* that could carry and/or infected with arenaviruses that make risk of virus transmission to the human as zoonotic.

Translation protein, homology detection and structure prediction of protein

At consequents, it is significant to use translated protein of TSA species in order to figure out protein homology between virus's genes. Transcript RNA for TSA species (table 1) that are *Channa punctate, Catostomus commersonii, Asymmetron lucayanum, Rhizopus oryzae, Rhopilema esculentum* and *Talitrus saltator* were utilized for obtaining open reading frames (ORFs) for homology detection and protein structure prediction according to other viruses in online database. SIB Bioinformatics Resource Portal such as Basic ExPASy translation tools (Gasteiger et al. 2003) on http://web.expasy.org/translate/ was used for determine (ORFs) of translation protein of TSA species. Then, the large one or two regions of open reading frame (ORFs) were selected and saved. Finally, the sequencing used to find out the homology detection and structure prediction of translated protein with other species of viruses by using online bioinformatics tools kits- HHpred version HHsuite-2.0.16mod (Söding et al. 2005) available on https://toolkit.tuebingen.mpg.de/hhpred. Thus, translated protein of TSA species as a result of translated blast of arenaviruses express some identity between TSA species and some other viruses as a result from HHpred (fig. 1, 2, 3, and 4).

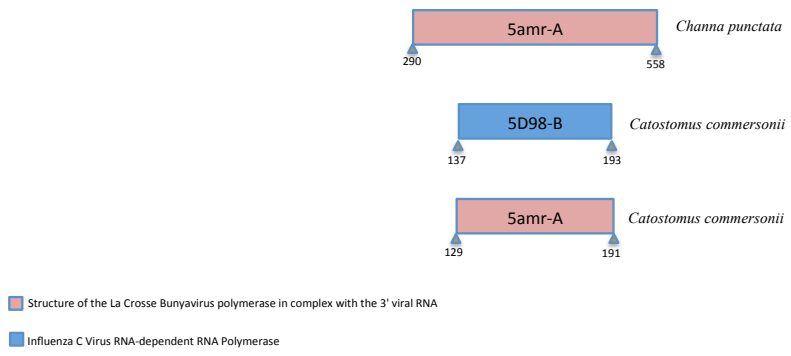


Figure 1: Homology detection and structure prediction of translated L protein of *Tacaribe virus* (TCRV). HHpred bioinformatics tools kit was used for protein analysis, while Microsoft PowerPoint was utilized for drawing the scales. The length of the scale depend on the gene size, the color depend on virus gene homology. The 5amr-A (pink color) refers to structure of the La Crosse Bunyavirus polymerase in complex with the 3' viral RNA, 5D98-B (blue color) refers to Influenza C Virus RNA-dependent RNA Polymerase.

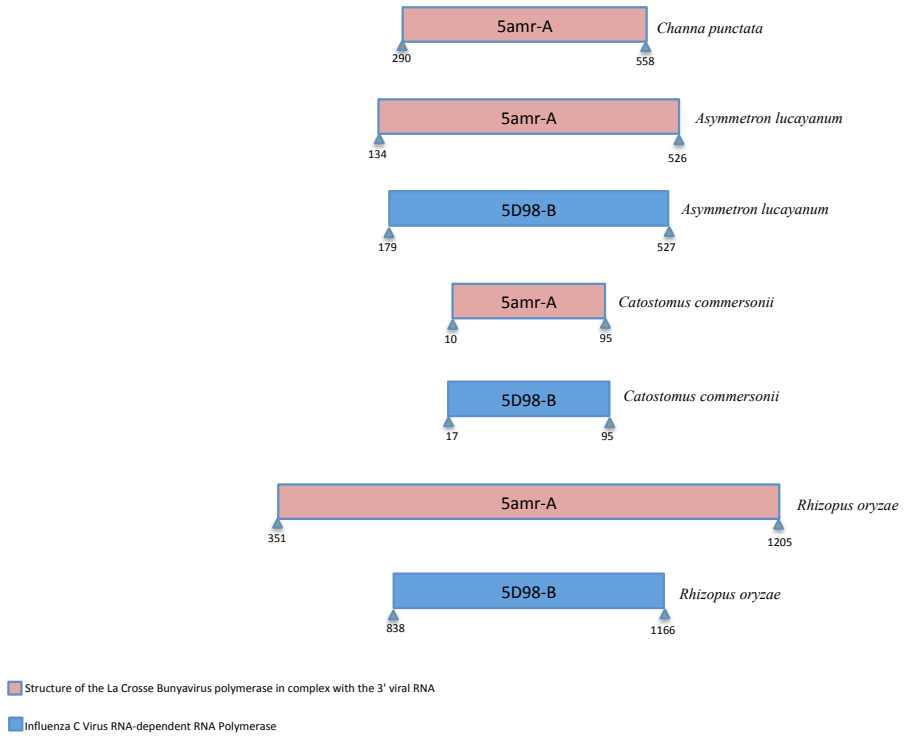


Figure 2: Homology detection and structure prediction of translated L protein of *lymphocytic choriomeningitis virus* (LCMV). HHpred bioinformatics tools kit was used for this purpose were Microsoft PowerPoint was used for drawing the scales. The length and the color of scale depend on gene size and gen homology of virus. The 5amr-A (pink color): structure of the La Crosse Bunyavirus polymerase in complex with the 3' viral RNA, 5D98-B (blue color): refers to Influenza C Virus RNA-dependent RNA Polymerase.

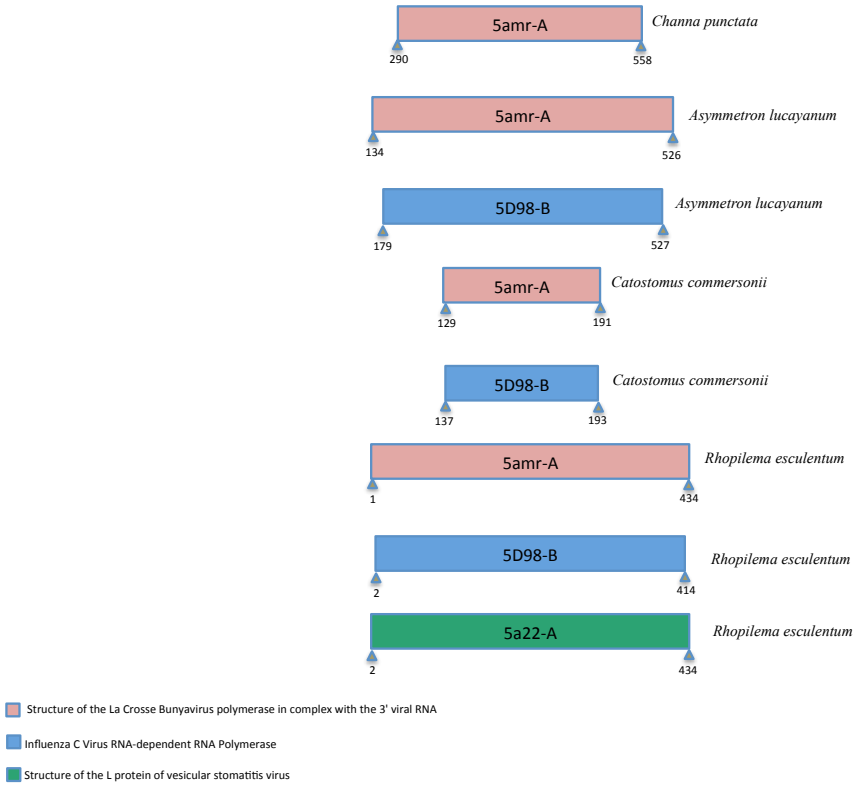


Figure 3: Homology detection and structure prediction of translated L protein of *Golden Gate virus* (GGV). HHpred bioinformatics tools kit was used for protein analysis and Microsoft PowerPoint was used for drawing the scales. The scale in different color and length depends on gene homology and gene size of the viruses, were 5amr-A (pink color) refers to structure of the La Crosse Bunyavirus polymerase in complex with the 3' viral RNA, 5D98-B (blue color) refers to Influenza C Virus RNA-dependent RNA Polymerase, while the 5a22-A (green color) refers to structure of the L protein of vesicular stomatitis virus.

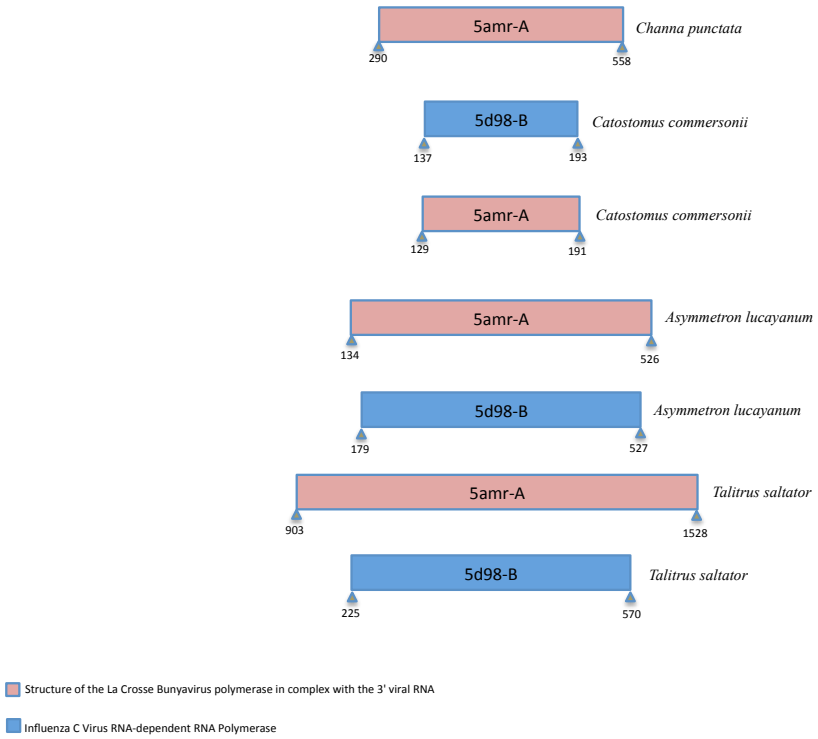


Figure 4: Homology detection and structure prediction of translated L protein of *California Academy of Science virus* (CASV). The translated protein was analysed by using HHpred bioinformatics tools, while the scales were drew by using Microsoft PowerPoint. The scales show in different color and size, the color refers to the identity of the gene, whereas the length refers to gene size of the target virus. The 5amr-A (pink color): structure of the La Crosse Bunyavirus polymerase in complex with the 3' viral RNA, 5D98-B (blue color): Influenza C Virus RNA-dependent RNA Polymerase.

The results reveal that the L protein of *Tacaribe virus* (TCRV), *lymphocytic choriomeningitis virus* (LCMV), *Golden Gate virus* (GGV) and *California Academy of Science virus* (CASV), were used successfully for searching translated nucleotides (tBLASTn) using Transcriptome ShoutGun assembly (TSA) in NCBI online database (McGinnis & Madden 2004). TSA results presented transcribed RNA alignment of six living species that have gene similarity with L protein of (TCRV), (LCMV), (GGV) and (CASV) as were used in this study. Interestingly, HHpred finding demonstrates that all translated protein of *Channa punctate, Catostomus commersonii, Asymmetron lucayanum, Rhizopus oryzae* and *Talitrus saltator* have some gene homology with structure of the La Crosse Bunyavirus polymerase in complex with the 3' viral RNA (5amr-A) in pink color and with Influenza C Virus RNA-dependent RNA Polymerase (5d98-B) in blue color, whereas *Rhopilema esculentum* has another

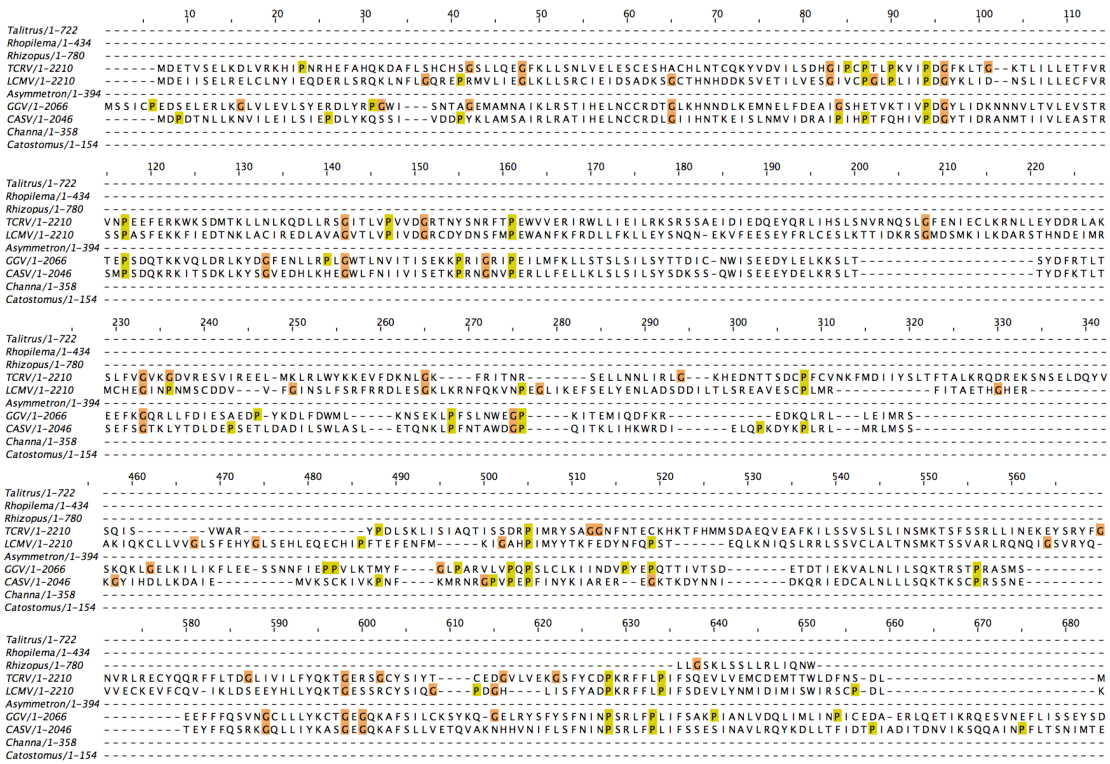
gene homology with the structure of the L protein of vesicular stomatitis virus (5a22-A) in green color (fig. 1, 2, 3, and 4).

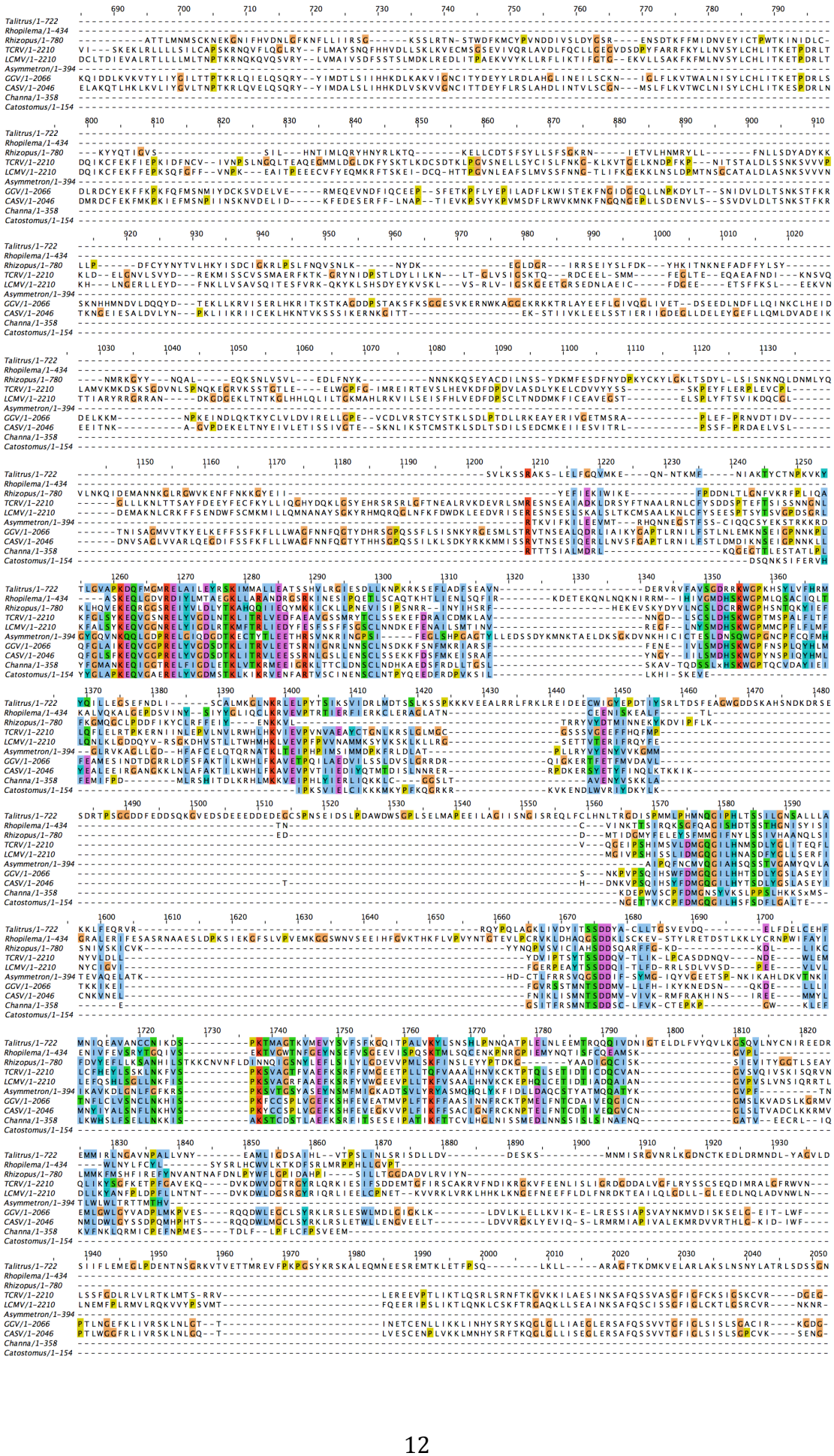
Furthermore, these finding provides evidence that some genes of viral RNA dependent RNA polymerase (RdRp) proteins of arenaviruses is shared with conserved region of open reading frames (ORFs) of translated protein of TSA living species as well as with La Crosse Bunyavirus, Influenza C virus and vesicular stomatitis virus. Also, protein homology shows differences in gene size between translated proteins of *Channa punctate, Catostomus commersonii, Asymmetron lucayanum, Rhizopus oryzae, Rhopilema esculentum* and *Talitrus saltator* with 5amr-A, 5d98-B and 5a22-A. For example, the protein homology of *Rhizopus oryzae* with La Crosse Bunyavirus polymerase in complex with the 3' viral RNA (5amr-A) is the longest gene and reaches about 854 amino acids, while the gene length homology between translated protein of *Catostomus commersonii* with Influenza C Virus RNA-dependent RNA Polymerase (5d98-B) express less gene length and reaches around 56 amino acids. Theses difference in gene size could have influence on protein-virus similarity such as in molecular characteristic and virus virulent, and that might predict during sequence alignments of translated protein. Therefore, ORFs of translated protein of transcribed RNA of TSA species shows homology and structure identification with some other viruses that could assist for prediction virus-host transmission and the alignment sequences that could be used for design phylogenetic tree for all viruses that used in this research.

Multiple Sequence Alignment (MSA) of Arenaviruses and TSA species

To evaluate the homology modeling and conservation of the protein as well as study phylogenetic analysis for expecting evolutionary relationship between proteins of the arenaviruses and TSA living species, multiple sequence alignment (MSA) can be used for this purpose (Rani & Ramyachitra 2016) (Kaya et al. 2014). At this research, open reading frames (ORFs) of translated protein of TSA species (table 1) that including *Channa punctate, Catostomus commersonii, Asymmetron lucayanum, Rhizopus oryzae, Rhopilema esculentum* and *Talitrus saltator* with L protein of

Tacaribe virus (TCRV), *lymphocytic choriomeningitis virus* (LCMV), *Golden Gate virus* (GGV) and *California Academy of Science virus* (CASV) were used for multiple sequence alignments (MSAs). Multiple sequence alignments (MSAs) assists for study the homology, evolutionary and structural relationship between aligned sequences and used for determination how many gaps between them (Rani & Ramyachitra 2016). Also, MSAs allows to identified conserved shared genes that have responsibility for virus virulence and predict possibility for virus transmission and future outbreak causes by viruses that distributed around the world (Blazewicz et al. 2013). To do that, sequences data of HHpred for open reading frames (ORFs) of translated protein of TSA species were aligned with L protein of (TCRV), (LCMV), (GGV) and (CASV) and some other viruses act as phylogeny root inclusive *Orthomyxoviridae*, *Bunyaviridae* and Monoegavirales (table 2) by using Clustal Omega version (1.2.4) available on http://www.ebi.ac.uk/Tools/msa/clustalo/. At least, multiple sequence alignments (MSAs) as shown in Fig.5 was manually presented via Jalview 2.10.1(Waterhouse et al. 2009) So, the multiple sequence alignment shows some indication of protein homology between arenaviruses species and translated protein of TSA species that found at this study.




Talitrus/1-722
Rhopilema/1-434
Rhizopus/1-780
TCRV/1-2210
LCMV/1-2210
Asymmetron/1-394
GGV/1-2066
CASV/1-2046
Channa/1-358
Catostomus/1-154

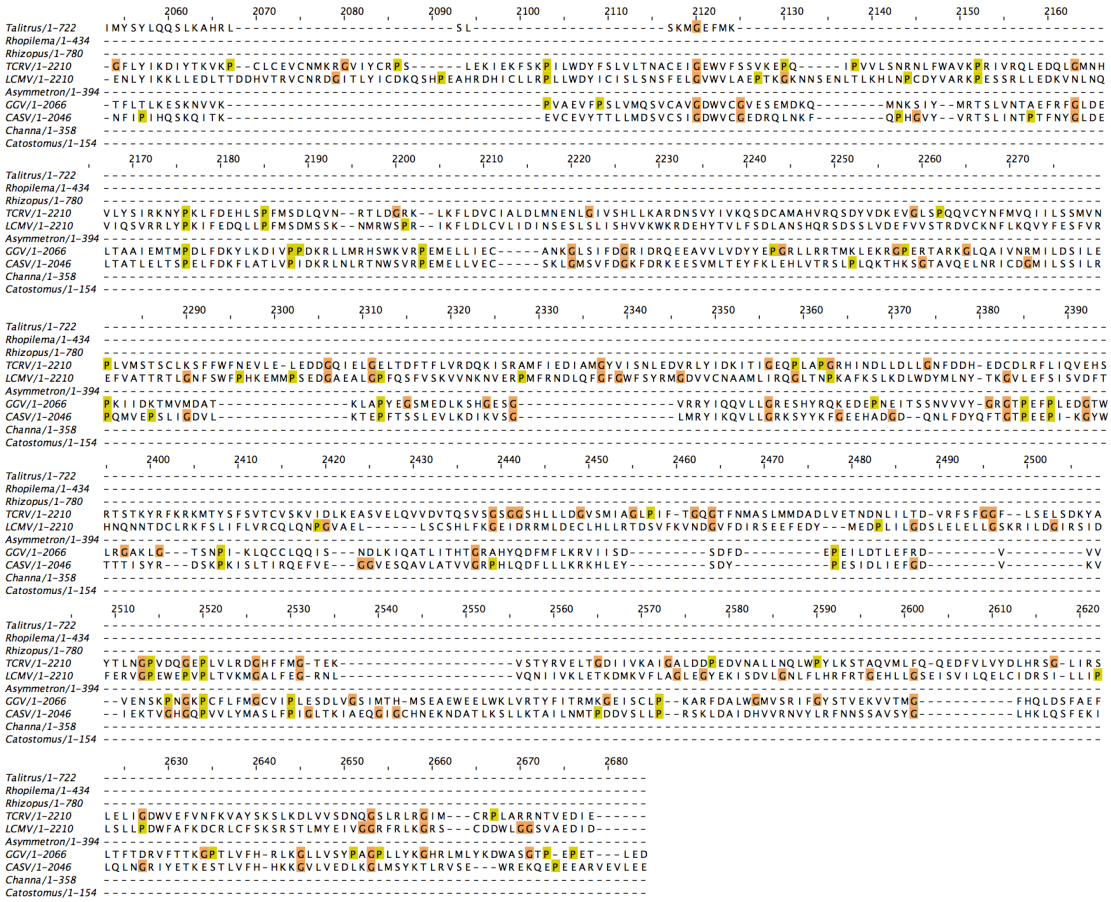


Figure 5: Multiple sequence alignment (MSA) of translated protein of TSA species with L protein of arenaviruses. Translated protein of *Channa punctate, Catostomus commersonii, Asymmetron lucayanum, Rhizopus oryzae, Rhopilema esculentum* and *Talitrus saltator* with L protein of arenaviruses were used for alignment (MSA) purpose. Were TCRV: *Tacaribe virus* (GenBank accession number NC_004292), LCMV: *lymphocytic choriomeningitis virus* (NP_694845.1), GGV: *Golden Gate virus* (YP_006590089.1), CASV: *California Academy of Science virus* (YP_006590093.1)**.** The alignment start where amino acid located from around 1230 to 1870 amino acid, as the sequence of translated protein of TSA species matches with L protein of the arenaviruses species. Clustal Omega (1.2.4) was used for the multiple alignments and Jalview 2.10.1 was used for manual edition.

From the data in Fig. 5, it is apparent that there are some shared regions of protein homology during multiple sequence alignments (MSAs), were translated proteins of *Channa punctate, Catostomus commersonii, Asymmetron lucayanum, Rhizopus oryzae, Rhopilema esculentum* and *Talitrus saltator* aligned with L protein of *Tacaribe virus* (TCRV), *lymphocytic choriomeningitis* virus (LCMV), *Golden Gate virus* (GGV) and *California Academy of Science virus* (CASV). These regions could expect some molecular characterization for proteins homology between TSA species and arenaviruses that could assist for study evolutionary relationship between them

such as when the alignment used for homology detection and 2D and 3D structure prediction of protein or when the data used for designing phylogenetic tree (Orobitg et al. 2015). In Theory, it is possible for design primers form the identical regions that could be able to amplify the cDNA that was isolated from both TSA species as well as from arenaviruses species by reverse transcription PCR (RT-PCR) (Lozano et al. 1997) (Vieth et al. 2007). So, there are significant conserved genes between translating protein of TSA species and L protein of arenaviruses that might assist for awarding some important details about molecular characteristics features between *Arenaviridea* and TSA species.

Bayesian Inference of Phylogeny (MrBayes)

To present clear phylogenetic analysis between the living species of TSA and arenaviruses that use at this study, it significant for designing phylogenetic tree that shows the evaluation history between theses species. MrBayes version 3.2.6 program (Ronquist et al. 2011) was used for creating phylogenetic tree by Bayesian Inference (BI). Bayesian Inference (BI) based on concept of the probabilities that based on prior expectation after analysis some data. MrBayes is command line interface and the outputs rely on what the user writes in mean screen of the program. The input screen opens after launch mb file that comes with MrBayes package as Macintosh (64bit) version was used for this purpose. At first, the L protein of arenaviruses sequencing inclusive TCRV: (GenBank accession number NP_694848.1), LCMV: (NP_694845.1), GGV: (YP_006590089.1), CASV: (YP_006590093.1) were aligned with HHpred date of TSA species by using multiple sequence alignments (MSAs) facility available online at Clustal Omega version (1.2.4) on http://www.ebi.ac.uk/Tools/msa/clustalo/. Some other viruses including viruses of *Orthomyxoviridae*, *Bunyaviridae* and Monoegavirales (table 2) were used as root viruses of phylogeny, then the alignment was saved as .phlip file at a location on PC. Second, the file was converted to .nxs file in order to be readable by MrBayes program. Then, MrBayes was opened were mb file found and then the .nxs file was dragged in after MrBayes prompt (>) in order to input file location, whereas there is a possibility to write file location instead of dragging after MrBayes prompt. Next,

mcmc was input in the MrBayes block, the command mcmc option indicates how to analysis the data via MrBays and the analysis begins after mcmc statement is processed until standard deviation reach 0.01-0.05 or less. When the average standard deviation of split frequencies reach less than 0.01, command line (control-C) was generated by pressing C key were holding ctrl key on keyboards after choosing yes option. Next, the sumt command was wrote after MrBayes block, which specifies the program how to summarize the consensus phylogenetic tree were convergence has been completed. Finally, command sump was input that indicates MrBayes interface how to summarize the values of the parameter as it were saved. At the result, the tree as shown in Fig. 6 were saved or the trees at .troprobs file could visualizes by a phylogenetic tree viewer program like FigTree v1.4.3 software, http://tree.bio.ed.ac.uk/software/figtree/ as was used for this purpose (Ronquist et al. 2011) (Hall 2011). Therefore, phylogenetic tree was successfully creating by using Bayesian Inference (BI) facility when the FigTree v1.4.3 was used for final presenting of the tree.

Table 2: Phylogeny root viruses and it accession number.

Family	Virus	Accession Number
Bunyaviridae	Hantaan virus (HTNV)	NP_941982.1
	Bunyamwera virus (BUNV)	NP_047211.1
	Tomato spotted wilt virus (TSWV)	AIA24440.1
	European mountain ash ringspot-associated virus (EMARAV)	Q6Q305
Orthomyxoviridae	Citrus psorosis virus (CPsV)	Q6DN67
	Rice stripe virus (RSV)	AFM93820.1
	Influenza A virus	P03433
Monoegavirales	Vesicular stomatitis virus (VSV)	AAA48442.1
	Lettuce big-vein associated virus (LBVV)	Q8B0U2
	Midway nyavirus (MIDWV)	C4NFK9
	Borna disease virus (BDV)	CEK41892.1
	Zaire ebolavirus (EBOV)	AAG40171.1
	Human metapneumovirus (hMPV)	AII17600.1

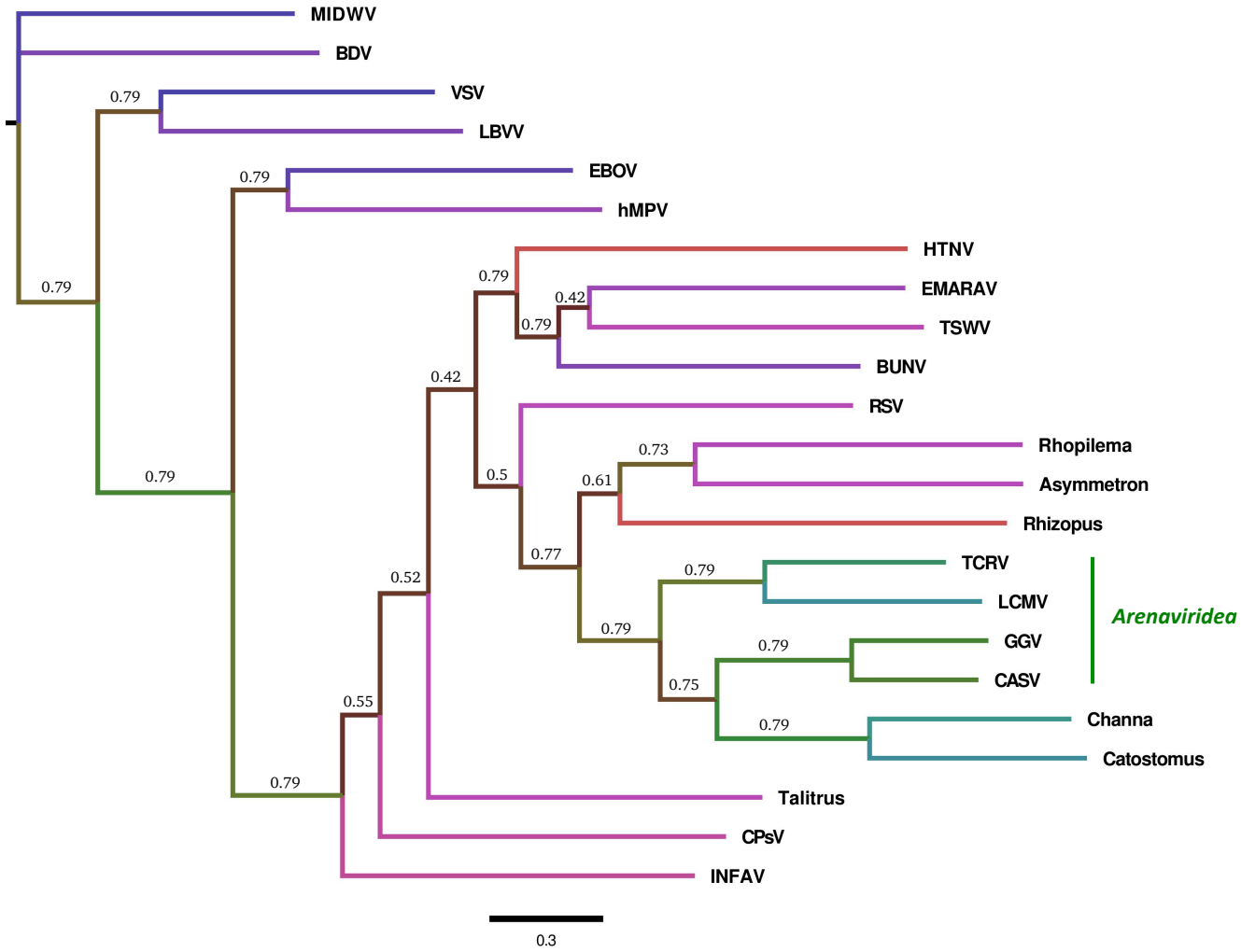


Figure 6: Molecular phylogenetic analysis by Bayesian Inference (BI). The phylogenetic tree was creating by using MrBayes v3.2.6 interface. Average standard deviation of split frequencies was less than 0.01. The analysis involved sequence of 23 amino acid including translated protein sequences of TSA species, L protein sequence of arenaviruses as well as some virus's protein sequencing were used as phylogeny root virus. Were TCRV: Tacaribe virus (GenBank accession number NC_004292), LCMV: lymphocytic choriomeningitis virus (NP_694845.1), GGV: Golden Gate virus (YP_006590089.1), CASV: California Academy of Science virus (YP_006590093.1), HTNV: Hantaan virus (NP_941982.1), BUNV: Bunyamwera virus (NP_047211.1), RSV: Rice stripe virus (AFM93820.1), TSWV: Tomato spotted wilt virus (AIA24440.1), VSV: Vesicular stomatitis virus, CPsV (AAA48442.1): Citrus psorosis virus (Q6DN67), EMARAV: European mountain ash ringspot-associated virus (Q6Q305), LBVV: Lettuce big-vein associated virus (Q8B0U2), INFA: Influenza A virus (P03433), MIDWV: Midway nyavirus (C4NFK9), BDV: Borna disease virus (CEK41892.1), EBOV: Zaire ebolavirus (AAG40171.1) and hMPV: Human metapneumovirus (AII17600.1). The final presenting of three was done by FigTree v1.4.3 and the color depends on branches length.

As demonstrate in Fig. 6, the tree shows some evaluation history between arenaviruses and TSA species. The tree compares between *Mammarenavirus* and *Reptarenavirus* and TSA species including *Channa punctate, Catostomus commersonii, Asymmetron lucayanum, Rhizopus oryzae, Rhopilema esculentum* and *Talitrus saltator* that found during tBLASn of L polymerase of arenaviruses. The online data of protein sequencing of L protein for arenaviruses, open reading frames (ORFs) of translated protein of TSA species were compared during Bayesian Inference analysis**.** The phylogenetic tree was creating using MrBayes v3.2.6 were all sequencing data inclusive some other viruses including viruses of *Orthomyxoviridae*, *Bunyaviridae* and Monoegavirales (table 2) that act as phylogeny tree roots viruses were used for tree building. To get best output from MrBayes v3.2.6 interface, the average standard deviation of split frequencies during MrBayes v3.2.6 processing was around 0.006356 as the manual instruction recommended (Ronquist et al. 2011) after the chain reach about 195000 generation. Finally, the tree was presented by FigTree v1.4.3. available on http://tree.bio.ed.ac.uk/software/figtree/ were the color indicates to the branches length. Therefore, the phylogenetic tree shows that some TSA species have same branches distance with arenaviruses including *Mammarenaviruses* and *Reptarenaviruses* were the color indicated branches length.

Moreover, there are expected virus genera potentially classify within *Arenaviridea* were L polymerase was used for tBLASTn. These genera including two expected species of fishes (Channa and Catostomus*)* that are more close to *Reptarenaviruses* including *Golden Gate virus* (GGV) and *California Academy of Science virus* (CASV) rather than *Mamarenaviruses*. Also, the tree shows that the fishes arenaviruses were located between *Mamarenaviruses* and *Reptarenaviruses*, while *Rhizarenavirus* is located between arenaviruses and *Orthomyxoviridae* (RSV). These viruses are potentially having similar characteristic features with arenaviruses and could transmit to human and animal via many routs such as by bite, sting and consumption respectively (Ter Meulen et al. 1996) (Kernéis et al. 2009). So, there are some novel viruses species that have significant similarity of L polymerase of arenaviruses that found during tBLASTn on GenBank database and potentially transmit some diseases including febrile diseases.

Furthermore, the protein homology and structure prediction of open reading frames (ORFs) from transcribe RNA alignment of TSA species shows similarity with La Crosse Bunyavirus polymerase in complex with the 3' viral RNA (5amr-A), Influenza C Virus RNA-dependent RNA Polymerase (5d98-B) and L protein of vesicular stomatitis virus (5a22-A). These similarity could classify these viruses within phylogenetic tree with same distance that TSA species located and that assist for predicting future outbreak of the diseases cause of all viruses even the hosts and vectors of these viruses are different. So, there are six living species of TSA that have evolution history with *Mammarenavirus* and *Reptarenavirus* were the analysis run out during investigation in online GenBank database and the hits used for design phylogenetic tree by bioinformatics software facilities including MrBayes v3.2.6 interface.

Discussion

At this researches, viral RNA dependent RNA polymerase (RdRp) or L polymerase of arenaviruses was utilized for searching translated nucleotide that archived in GenBank by using Transcriptome ShoutGun assembly (TSA) on translated blast (tBLASTn) (McGinnis & Madden 2004). These arenaviruses including *Tacaribe virus* (TCRV), *lymphocytic choriomeningitis virus* (LCMV), *Golden Gate virus* (GGV) and *California Academy of Science virus* (CASV). The tBLASTn is available online on NCBI for free access facility. The TSA finding reveal that there are six living species which are *Channa punctate, Catostomus commersonii, Asymmetron lucayanum, Rhizopus oryzae, Rhopilema esculentum* and *Talitrus saltator* that have gene homology with L protein of (TCRV), (LCMV), (GGV) and (CASV). Under those circumstances, it is significant to find translated protein of open reading frames (ORFs) form TSA data by using standard translation tool as ExPAy were used for this purpose (Gasteiger et al. 2003). It was found by HHpred version HHsuite-2.0.16mod (Söding et al. 2005) that the translated protein of TSA species have homology with structure of the La Crosse Bunyavirus polymerase in complex with the 3' viral RNA (5amr-A), Influenza C Virus RNA-dependent RNA Polymerase (5d98-B) and with L

protein of vesicular stomatitis virus (5a22-A). These finding could support the predicating about TSA species that could carry some viral genes that might cause diseases for both human and animals around the areas that they found such as the countries that used the murine animals for breeding and/or feeding.

Moreover, there is some gene similarities between L protein of arenaviruses and translated protein of TSA species at some location when the alignment carried by using the Clustal Omega (1.2.4), which could assist for confirmation the protein homology by designing some gene specific primers that could amplified synthesized cDNA of TSA species and L gene of arenaviruses (Blazewicz et al. 2013) (Ortuño et al. 2015) which could be used for further experimental design in Vitro. In fact, reverse transcription PCR (Paweska et al. 2009) and real time PCR (Cordey et al. 2011) were used for finding novel/existence arenaviruses as degenerate oligonucleotide primers of L (Vieth et al. 2007), NP and GPC (Strigl et al. 1998) genes of arenaviruses were designed and used for virus investigation.

Interestingly, phylogenetic tree for alignment of all proteins sequences that used and obtained at this research were utilized for designing phylogenetic tree by using Bayesian Inference (BI) were MrBayes v3.2.6 interface run for this purpose (Ronquist et al. 2011) and FigTree v1.4.3 available on http://tree.bio.ed.ac.uk/software/figtree/ was used for the final presenting of the tree. The finding expected novel TSA genus of fishes arenaviruses that consists of *Channa punctate, Catostomus commersonii* that potentially be within arenaviruses were located between *Mammareanvirus* and *Reptareanvirus.* Also, the tree shows that the *Rhizarenavirus* located between *Mammareanvirus* and *Rice Stripe virus* (RSV) of *Orthomyxoviridae.* Besides, there are many viruses that used for phylogenetic tree root such as viruses of the family *Orthomyxoviridae*, *Bunyaviridae* and Monoegavirales (table 2). Thus, there is evidence of gene sequencing of TSA living species that have some genes homology of the L protein of arenaviruses during using some bioinformatics tools and software including online databases, which assist for monitoring future outbreak for human and animals cause by these viruses.

References

Adams, M.J. et al., 2016. Ratification vote on taxonomic proposals to the International Committee on Taxonomy of Viruses (2016). *Archives of Virology*, 161(10), pp.2921–2949.

Altman, R.B. & Dugan, J.M., 2005. Defining bioinformatics and structural bioinformatics. In *Structural Bioinformatics*. John Wiley & Sons, Inc., pp. 3–14.

Belshaw, R. & Katzourakis, A., 2005. BlastAlign: A program that uses blast to align problematic nucleotide sequences. *Bioinformatics*, 21(1), pp.122–123.

Blazewicz, J. et al., 2013. G-MSA - A GPU-based, fast and accurate algorithm for multiple sequence alignment. *Journal of Parallel and Distributed Computing*, 73(1), pp.32–41. Available at: http://dx.doi.org/10.1016/j.jpdc.2012.04.004.

Bodewes, R. et al., 2013. Detection of novel divergent arenaviruses in boid snakes with inclusion body disease in The Netherlands. *Journal of General Virology*, 94(PART 6), pp.1206–1210.

Buchmeier, M.J., de la Torre, J.C. & Peters, C.J., 2007. *Arenaviridae: the viruses and their replication, In: Knipe, D.L., Howley, P.M. (Eds.),Fields Virology* 4th ed., Lippincott- Raven, Philadelphia.

Clegg, J.C., 2002. Molecular phylogeny of the arenaviruses.Current Topics in Microbiology and Immunology. , volume 262(CT MICROBIOLOGY), pp.262, 1–24.

Clote, P. & Backofen, R., 2000. *Computational Molecular Biology: An Introduction* 1st ed., West Sussex: John Wiley and Sons Inc.

Cordey, S. et al., 2011. Analytical validation of a lymphocytic choriomeningitis virus real-time RT-PCR assay. *Journal of Virological Methods*, 177(1), pp.118–122. Available at: http://dx.doi.org/10.1016/j.jviromet.2011.06.018.

Ehichioya, D.U. et al., 2010. Lassa fever, Nigeria, 2005-2008. *Emerging Infectious Diseases*, 16(6), pp.1040–1041.

Fichet-Calvet, E. & Rogers, D.J., 2009. Risk maps of lassa fever in West Africa. *PLoS Neglected Tropical Diseases*, 3(3).

Gasteiger, E. et al., 2003. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research*, 31(13), pp.3784–3788.

Gopakumar, O., 2012. *Bioinformatics: Sequence and Structural Analysis*, Oxford:

Alpha Science Intl Ltd.

Hall, B.G., 2011. *Phylogentic Trees Made Easy* Fourth Edi., Sunderland, Massachusetts, USA: Sinauer Associates, Inc.

Kaya, M., Sarhan, A. & Alhajj, R., 2014. Multiple sequence alignment with affine gap by using multi-objective genetic algorithm. *Computer Methods and Programs in Biomedicine*, 114(1), pp.38–49. Available at: http://dx.doi.org/10.1016/j.cmpb.2014.01.013.

Kernéis, S. et al., 2009. Prevalence and risk factors of lassa seropositivity in inhabitants of the Forest Region of Guinea: A cross-sectional study. *PLoS Neglected Tropical Diseases*, 3(11).

Lal, S. et al., 2005. *ZOONOTIC DISEASES OF PUBLIC HEALTH IMPORTANCE*, Delhi. Available at: http://www.ncdc.gov.in/writereaddata/linkimages/zoonotic_manual9201154898.pdf.

Liljas, A., L. et al., 2001. *Textbook of Structural Biology*, Singapore: World Scientific Printe.

Lozano, M.E. et al., 1997. Characterization of arenaviruses using a family-specific primer set for RT-PCR amplification and RFLP analysis its potential use for detection of uncharacterized arenaviruses. *Virus Research*, 49(1), pp.79–89.

McGinnis, S. & Madden, T.L., 2004. BLAST: At the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Research*, 32(WEB SERVER ISS.), pp.20–25.

Ter Meulen, J. et al., 1996. Hunting of peridomestic rodents and consumption of their meat as possible risk factors for rodent-to-human transmission of Lassa virus in the Republic of Guinea. *The American Society of Tropical Medicine and Hygiene*, 55(6), pp.661–662. Available at: https://doi.org/10.4269/ajtmh.1996.55.661.

Morin, B. et al., 2010. The N-terminal domain of the arenavirus L protein is an RNA endonuclease essential in mRNA transcription. *PLoS Pathogens*, 6(9).

Orobitg, M. et al., 2015. High Performance computing improvements on bioinformatics consistency-based multiple sequence alignment tools. *Parallel Computing*, 42, pp.18–34. Available at: http://dx.doi.org/10.1016/j.parco.2014.09.010.

Ortuño, F.M. et al., 2015. Comparing different machine learning and mathematical

regression models to evaluate multiple sequence alignments. *Neurocomputing*, 164, pp.123–136. Available at: http://dx.doi.org/10.1016/j.neucom.2015.01.080.

Paweska, J.T. et al., 2009. Nosocomial Outbreak of Novel Arenavirus Infection , Southern Africa. , 15(10).

Pearson, W.R., 1995. Comparison of methods for searching protein sequence databases. *Protein Society*, 4(6), pp.1145–60. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2143149&tool=pmcentrez&rendertype=abstract.

Perry, W.L., 2002. JavaScript DNA Translator: DNA-Aligned Protein Translations. *BioComputing/BioInformatics*, 33(6), pp.1318–1320.

Pevzner, P.., 2000. *Computational Molecular Biology: An Algorithmic Approach*, Cambridge: The MIT Press.

Rani, R.R. & Ramyachitra, D., 2016. Multiple sequence alignment using multi-objective based bacterial foraging optimization algorithm. *BioSystems*, 150, pp.177–189. Available at: http://dx.doi.org/10.1016/j.biosystems.2016.10.005.

Ronquist, F., Huelsenbeck, J. & Teslenko, M., 2011. MrBayes Version 3.2 Manual: Tutorials and Model Summaries. *Manual MrBayes*, pp.1–103.

Salvato, M., Shimomaye, E. & B.A.Oldstone, M., 1989. The primary structure of the lymphocytic choriomeningitis virus L gene encodes a putative rna polymerase. *Virology*, 169(2), pp.377–384.

Söding, J., Biegert, A. & Lupas, A.N., 2005. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Research*, 33(SUPPL. 2), pp.244–248.

Strigl, S., Conde, S. & Schmitz, H., 1998. Detection of Lassa Virus Antinucleoprotein Immunoglobulin G (IgG) and IgM Antibodies by a Simple Recombinant Immunoblot Assay for Field Use. , 36(11), pp.3143–3148.

Torre, J.C. de la, 2009. Molecular and Cell Biology of the PrototypicArenavirus LCMV: Implications forUnderstanding and Combating HemorrhagicFever Arenaviruses. *Annals of the New York Academy of Sciences*, 1171(Issue Supplement s1, Version of Record online), pp.E57–E64.

Vieth, S. et al., 2007. RT-PCR assay for detection of Lassa virus and related Old World arenaviruses targeting the L gene. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 101(12), pp.1253–1264.

Waterhouse, A.M. et al., 2009. Jalview Version 2-A multiple sequence alignment

editor and analysis workbench. *Bioinformatics*, 25(9), pp.1189–1191.

Zapata, J.C. & Salvato, M.S., 2013. Arenavirus variations due to host-specific adaptation. *Viruses*, 5(1), pp.241–278.

Zhang, Q., Veretnik, S. & Bourne, P.E., 2005. Overview of Structural Bioinformatics. Springer. In *Bioinformatics Technologies*. Heidelberg: Springer, pp. 15–44.