RESEARCH ARTICLE

Epitope specificity and protein signaling interactions driving epidemic occurrences of Ebola disease [version 1; peer review: peer review discontinued]

Daniel Achinko ^{1,2}, Anton Dormer¹

¹PepVax, Inc. 10411 Motor City Drive, Bethesda, MD, 20817, USA

²Current affiliation: NCBI/NLM/NIH, Computational Biology Branch, Bethesda, MD, 20894, USA

V1 First published: 24 Jun 2015, 4:166
<https://doi.org/10.12688/f1000research.6515.1>

Latest published: 24 Jun 2015, 4:166
<https://doi.org/10.12688/f1000research.6515.1>

Peer review discontinued

Peer review at F1000Research is author-driven. Currently no reviewers are being invited. [What does this mean?](#)

Any reports and responses or comments on the article can be found at the end of the article.

Abstract

Ebolavirus has as main hosts, humans and nonhuman primates where its pathogenic effects result to serious hemorrhagic fever with lethal effects. Despite the great advancement in deciphering the clinical course of the virus, specific mechanisms favoring Ebolavirus pathogenicity and transmission, and which genomic structures are most antigenic, are still to be clearly delineated. This study used functional protein phylogenetic analysis, pathway designs and antigenic epitope predictions to respectively; identify viral genomic regions closely related to host proteins, predict protein/genetic interactions favoring viral pathogenesis and identify frequency of MHC class I & II immune related host peptide variants whose transmission intensity value favors disease epidemicity. Viral glycoprotein (VGP) presented the highest genetic variation and though captured on the network with matrix protein (MXP), no direct interaction was observed. The majority of host interacting proteins presented with kinase functions, particularly a protein-signaling role observed in LCK, a Tyrosine-protein kinase with the most dominant interactions and viral related functions implicated in disease shock events. Four VGP and three MXP main antigenic epitopes identified, differentially showed high frequency to two MHC class I types. The same pattern was observed for VGP and MXP antigenic epitopes predicted to MHC class II allele variants, favoring high transmission intensity values within the host population, suggesting their involvement in Ebola epidemic upsurges. Related Ebola species with high transmission values were dominantly non-Zaire Ebolaviruses whose antigenic regions showed several repeats, implicating them in viral antigenic variations. Our analysis show that VGP and MXP are both critical for viral entry and pathogenicity in the host and with their species specific occurrence, their combined role in drug/vaccine design is critical. The identification of several antigenic epitopes in this study will be used in combination for drug/vaccine design and for

better understanding related molecular targets in pathogenic pathways favoring Ebola disease burden.

Keywords

Ebolavirus , pathogenicity , transmission , Major histocompatibility complex (MHC) , Immune epitopes , protein interactions

This article is included in the **Emerging Diseases and Outbreaks** gateway.

This article is included in the **Ebola Virus** collection.

Corresponding author: Daniel Achinko (achdany@yahoo.co.uk)

Competing interests: No competing interests were disclosed.

Grant information: The author(s) declared that no grants were involved in supporting this work.

Copyright: © 2015 Achinko D and Dormer A. This is an open access article distributed under the terms of the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the **Creative Commons Zero "No rights reserved" data waiver** (CC0 1.0 Public domain dedication).

How to cite this article: Achinko D and Dormer A. **Epitope specificity and protein signaling interactions driving epidemic occurrences of Ebola disease [version 1; peer review: peer review discontinued]** F1000Research 2015, 4:166 <https://doi.org/10.12688/f1000research.6515.1>

First published: 24 Jun 2015, 4:166 <https://doi.org/10.12688/f1000research.6515.1>

Introduction

Ebolavirus is responsible for explosive though rare outbreaks of haemorrhagic fever common in equatorial Africa. The virus is transmitted from person to person through infected body fluids such as faeces, vomit and blood (Dowell *et al.*, 1999). Ebolavirus comprises mainly of four subtypes; Zaire, Sudan, Cote d'Ivoire and Reston. Ebola ancestors are not clearly defined, though Marburg virus, which belongs in the same family as Ebola (Filoviridae; a grouping consisting of nonsegmented, negative-stranded RNA viruses) is a possible ancestor (Mayo & Pringle, 1998; Sanchez *et al.*, 1993).

The average rate of case fatalities due to Ebolavirus is 78% (Kuhn, 2008). The Ebola outbreak in 2014 was considered the largest in history and estimates show about 2,473 infections with 1350 deaths occurring (WHO). The 2014 outbreak was first identified in February in Guinea, West Africa (Burke, 2003), then in March it spread into Liberia, to Sierra Leone in May and finally to Nigeria towards the end of July. This outbreak was considered the largest because of its exponential expansion, doubling every 34.8 days. In major cities such as Conakry (Guinea), Monrovia (Liberia), Freetown (Sierra Leone) and Lagos (Nigeria) where viral emergence was observed, the spectre of local and international viral dissemination was raised.

Currently, no approved drugs have been identified against Ebolavirus disease, though rapid diagnosis, intensive care and support is playing a major role in improving survival. Characteristics of the Ebola disease include; high fever, headache, body aches, intense weakness, stomach pain, and lack of appetite. Manifestations of Ebola in patients are: vomiting, diarrhea, rash, impaired kidney and liver function and for critical cases, external and internal bleeding (<http://www.nih.gov/news/health/aug2014/od-29.htm>).

The current Ebola epidemic raised awareness worldwide given its international transmission observed in the US and Spain. Several genetic approaches to understanding the disease are currently being tested, the majority investigating the genomic composition of the virus, pathogenic pathways within the host and immune variants (MHC-HLA class I & II) commonly expressed by Ebola disease survivors and non-survivors (Sullivan *et al.*, 2003). The HIV, ENV glycoprotein (gp160) and Ebola virus envelope glycoprotein (VGP) are genetically closely related hence, on disrupting the glycoprotein transmembrane domain of MHC class I at the cell surface, a facilitated viral evasion of the immune system was observed (Ploegh, 1998). This viral interaction at the cell surface is considered to occur with the CD209 antigen (DC-SIGN), making the VGP a prominent area of research for drug/vaccine development.

The Ebolavirus genome is 19 kb in length, with seven open reading frames that encode structural proteins including virion envelope glycoprotein (VGP), nucleoprotein (NP), and matrix proteins VP24 and VP40; nonstructural proteins, including VP30 and VP35; and

the viral polymerase (Sanchez *et al.*, 2001). Unlike the Marburg virus, the GP open reading frame of Ebolavirus has two gene products, a soluble 60- to 70-kDa protein (sGP) and a full-length 150- to 170-kDa protein (GP) that inserts into the viral membrane (Sanchez *et al.*, 1996; Volchkov *et al.*, 1995), through transcriptional editing.

The adaptive immune and inflammatory systems both respond to viral infection synchronously, such that some types of cells (e.g.: monocytes and macrophages) become principal targets for disease pathogenesis. This infection criteria was first suggested based on observations of Ebolavirus immunohistochemical localization *in vivo*, therefore, mononuclear phagocytes (monocytes and macrophages), endothelial cells, and hepatocytes constitute the main infectious targets (Baskerville *et al.*, 1978; Baskerville *et al.*, 1985). The recognition of specific Ebolavirus VGP epitopes by monoclonal antibodies conveyed immune protection in a murine model (Wilson *et al.*, 2000) and in guinea pigs (Parren *et al.*, 2002) infected with Ebolavirus.

From the current information known about viral interaction with host genes and pathways favoring cytokine activation, this study was undertaken to analyze the relationship between predicted viral antigenic epitopes and their population frequency as a means of transmission risk. This Ebola species-epitope variation was used to decipher viral species diversity and their related pathway, and how this interplay favors disease intensity within the host. The different viral epitopes identified in this study and their related immune peptides served as a means for predicting which viral species could be commonly found in a given Ebolavirus epidemic. This study opens up epidemiological aspects of viral pathogenesis and possible targets for drug/vaccine design.

Results

Phylogenetic analysis

The sequence alignment for all retrieved protein variants of different Ebola sequences showed a 1.3% identity across all sequences and a 14.8% pairwise identity across amino acid residues for all sites. These values further improved for each genomic variant analysed for all Ebola species as follows: 23.3% identity across all sites and 68.8% pairwise identity for nuclear proteins, 17.5% identity across all sites and 57.2% pairwise identity for virion glycoproteins and 23.9% identity across all sites and 70.6% pairwise identity respectively for matrix proteins. The phylogenetic analysis showed two main clusters (Figure 1A) of nuclear proteins on one arm (blue cluster) and the other arm contained glycoproteins in a sub cluster (red), and matrix proteins in another sub cluster (blue). Mammalian related immune protein sequences associated with viral antigenicity composed the rest of the sequences. The red sub cluster described above was further subjected to phylogenetic analysis and four sub clusters were identified (Figure 1B) with the red composed of the Zaire Ebolavirus (29 variants), the black composed of the Reston Ebolavirus (14 variants) the green composed of the Sudan Ebola

Figure 1. Functional-Phylogenetic analysis of Ebola species. 348 amino acid sequences for Ebolavirus species and mammals were aligned and their phylogeny on the unrooted tree showed two distinct clades (**1A**) wherein the blue cluster belonged to nuclear protein sequences of the viral species while the other cluster contained viral glycoprotein sequences (red sub cluster) and viral matrix proteins (green sub cluster) and the rest were dominantly mammalian proteins. **1B**) The viral glycoproteins were subjected to further phylogenetic analysis and the unrooted tree showed four main clusters belonging to different species as follows; the red composed of Zaire Ebolavirus (29 variants), the black composed of the Reston Ebolavirus (14 variants) the green composed of the Sudan Ebolavirus (11 variants) and the blue composed of Bundibugyo and Tai forest Ebolaviruses (8 variants).

virus (11 variants) and the blue composed of Bundibugyo and Tai forest Ebola viruses (8 variants as indicated by the name) (Table 1). The Marburg virus separated independently and appears to be the ancestral strain from which Ebolaviruses diverged. The host-virus phylogenetic clustering based on related functional sequence features associated with viral genomic structures.

Protein-protein interaction analysis

The predicted network showed 69 nodes with 123 edges (Figure 2). The network analysis distinguished from most important to least important genes (nodes) depicted by name size (Figure 2). The three most important nodes in descending order included: GRB2 (Growth factor receptor bound protein 2), IKKE (Inhibitor of

Table 1. Distribution of sequence variants for viral glycoprotein genes within Ebola species extracted from UniProtKB. 29 variants were identified for Zaire Ebolavirus, 14 variants for Reston Ebolavirus, 11 for Sudan Ebolavirus, 3 variants for Bundibugyo Ebolavirus (B) and 5 variants for Tai forest Ebolavirus (T).

Zaire Ebolavirus	Reston Ebolavirus	Bundibugyo & Tai forest Ebolavirus	Sudan Ebolavirus
VGP_EBOZM	VSSGP_EBORE	B8XCN2_9MONO (B)	VSGP_EBOSU
VGP_EBOZ5	VSGP_EBORS	VGP_TAFVC (T)	I7FLM2_9MONO
VSSGP_EBOZM	C6G8F3_EBORE	VSGP_TAFVC (T)	I7F2J9_9MONO
VSGP_EBOZM	C6G8E5_EBORE	B8XCP0_9MONO (T)	VSGP_EBOSM
A0A068J471_9MONO	C6G8D6_EBORE	B8XCN9_9MONO (T)	B0LPL7_9MONO
X5H5A9_9MONO	C6G8F2_EBORE	B8XCN0_9MONO (B)	VGP_EBOSU
A9QPL9_9MONO	VSSGP_EBORR	B8XCN1_9MONO (B)	B0LPL8_9MONO
Q6V1Q5_EBOZ5	C6G8D7_EBORE	B8XCP1_9MONO (T)	VGP_EBOSM
A0A068J949_9MONO	C6G8E4_EBORE		VGP_EBOSB
VSGP_EBOZ5	VSGP_EBORR		VSGP_EBOSB
VSGP_EBOEC	VSGP_EBORE		VSSGP_EBOSU
V9TJ39_9MONO	VGP_EBORS		
VGP_EBOEC	VGP_EBORR		
X5HL92_9MONO	VGP_EBORE		
A9QPM0_9MONO			
G8DB49_9MONO			
G8DB50_9MONO			
A0A068J9C3_9MONO			
X5H5I7_9MONO			
X5HMZ0_9MONO			
A0A068J419_9MONO			
A0A068J4A4_9MONO			
G8DB42_9MONO			
V9TJ34_9MONO			
X5HMX4_9MONO			
VSGP_EBOG4			
A0A068J941_9MONO			
VSSGP_EBOZ5			
VGP_EBOG4

B

Figure 2. Ebola virus protein interaction and pathway analysis. A) The 348 extracted sequences were used to predict a protein interaction map with 69 nodes and 123 edges. Several mammalian sequences captured on the network showed direct interactions with each other with one main protein (LCK) a tyrosine kinase was involved with virus interactions and signalling functions. Name size related to number of interactions each protein had within the network. Two viral proteins were captured in the network; VP40 a matrix protein and ENV a glycoprotein. The red node at top of the network interacting with several other nodes represents a cluster of microRNAs within the network and their related genes they regulate based on their interactions. **B)** Several regulatory elements were seen to interact with the nodes of interest (red circles). MiRNAs (yellow squares), drug targets (purple squares), transcription factors (light green squares).

nuclear factor kappa-B Kinase subunit epsilon) and UBIQ/UBB (Polyubiquitin-B). Plotting neighborhood connectivity (number of neighbors per node) against clustering coefficient (CC) [ratio of connected edges N for neighbors of node n compared to the maximum number of edges M for node n (N/M)] (Figure 3), showed that several nodes ($n = 16$) did not cluster but had several neighbors. The largest cluster had a coefficient of 0.15 to 0.30. The node distribution based on the best-fit line ($4.346 = 8.858x$) showed a correlation value of $r^2 = 0.319$ indicating skewedness towards most connected nodes. Based on all GO molecular functions derived from the network, focus was laid on those genes involved with immune and virus related functions which were; LCK (a tyrosine-protein kinase) and CD209 (a pathogen recognition receptor) respectively. LCK associated functions included: protein kinase activity (GO 4672, P-value = $1.9585E-9$) (Figure 5B), interspecies interaction between organisms (GO 44419, P-value = $1.3293E-7$), T cell activation (GO 42110, P-value = $1.1774E-4$), receptor binding (GO 5102, P-value = $7.5709E-3$), SH2 domain binding

(GO 42169, P-value = $1.2270E-2$), CD8 receptor binding (GO 42610, P-value = $1.5663E-2$) and CD4 receptor binding (GO 42609, P-value = $2.5648E-2$). CD209 associated functions included; Intracellular virion transport (GO 46795, P-value = $2.5648E-2$), endocytosis (GO 6897 P-value = $1.1185E-2$), signal transducer (GO: 4871, P-value = $1.1671E-2$), peptide antigen transport (GO 46968, P-value = $3.4042E-2$) and peptide binding (GO 42277, P-value = $4.4976E-2$). The protein kinase pathway associated with LCK, showed two clusters of interacting proteins (Figure 4A) with the largest (9 genes) containing LCK while that for peptide binding associated with CD209, showed non-interacting proteins (Figure 4B). Three proteins, Bone marrow stromal antigen (BST2), Serine/threonine-protein kinase (TBK1) and Interferon regulatory factor 3 (IRF3), had as function, response to virus (GO 9615, P-value = $2.4498E-2$) and of the three, TBK1 was identified in the protein kinase pathway. Two viral related proteins identified on the network were ENV [envelope glycoprotein gp160 for the HIV-1 virus which binds to CD4 cells (GO 19031, P-value = $4.0529E-2$)]

Figure 3. Neighborhood connectivity and clusterization analysis of the protein interactome. This graph shows that several genes had neighbors up to 14 proteins but no clusters were formed due to no interactions. For those that clustered, majority were between 0.15 and 0.30 defining a high zone of interacting genes within cells.

Figure 4. Pathways and clusterization observed for viral related genes. A) for the several pathways identified with LCK, protein kinase pathway showed three clusters (blue nodes) with LCK involved in the largest cluster (9 genes). **B)** The peptide binding cluster (brown nodes) involving CD209 a pathogen recognition receptor gene showed no direct interactions between each other.

and VP40 (virus matrix protein, a structural protein most commonly found in Ebolavirus and promotes virus assembly and budding) though both showed no protein interactions. Identification of regulatory elements targeting the network showed that, 66 drugs have been developed as targets for several genes within the network, of which proto-oncogene tyrosine-protein kinase (src) and LCK ranked first and second respectively. A few transcription factors like; signal transducer and activator of transcription 1-alpha/beta (STAT1), T-cell leukemia homeobox protein 1 (TLX1), transcription factor PU.1 (SPI1), cyclin dependent kinase inhibitor 1A, and cyclin-dependent kinase inhibitor 1A (CDKN1A), regulated genes including nuclear factor NF- κ -B p105 subunit (NFKB1), p100 subunit (NFKB2), CD209 and SRC respectively (Figure 5D). Several microRNA molecules (n = 1399) targeting several genes within the network, mapped to different disease pathways with the most important being hsa05131 (pathogenic *Escherichia coli* infection) and hsa01430 (cell communication) while major pathway gene related domains were; IPR001664 (intermediate filament protein) and IPR000108 (neutrophil cytosol factor 2).

Epitope prediction and MHC HLA class I & II allele distribution among Ebolavirus species

The most antigenic epitopes for each virus structural proteins (glycoproteins and matrix proteins) were statistically evaluated for their peptide antigenic score against peptide frequency (count) for the HLA allele types (Figure 5A–F). For HLA type I, the VGP epitope peptides showed $t = 0.24338$ while the MXP proteins showed $t = 0.09190$. For the HLA type II (HLA-DRB alleles), the VGP showed $t = 2.8506E-107$ while the MXP showed $t = 2.21965E-49$ and for HLA type II (HLA-DQA-A alleles), the VGP showed $t = 8.96323E-08$ while the MXP showed $t = 0.046921205$. For HLA class I alleles, four main epitopes (FLLQLNETI, FLTNITRGV, FQVDHLTYV and NMTNFLFQV) were predicted for VGP while three main epitopes (AIMLASYTV, ALLTGSYTI and LTMVITPDY) were predicted for MXP. Both viral structural proteins showed high frequency values for HLA*A02:01 and HLA*A02:02 alleles (Figure 6A, Table S2) with VGP higher for the latter than the former and vice versa for MXP. The validated epitopes for VGP showed that; FLLQLNETI was specific to HLA*A02:02 while FQVDHLTYV and NMTNFLFQV was variably distributed for all the selected HLA class I alleles (Figure 6B). The validated epitopes for MXP showed that; AIMLASYTV was variably distributed for selected HLA class I alleles while ALLTGSYTI and LTMVITPDY were allele specific (Figure 6C). FLLQLNETI was specific to Zaire Ebolavirus with a unique occurrence while the other epitopes were either specific or distributed among Bundibugyo and Taï forest Ebolaviruses with multiple occurrences (Figure 6D). AIMLASYTV was more specific to Reston and Sudan Ebolaviruses with multiple occurrences while the other epitopes were either specific to Sudan Ebolavirus or Marburg virus with unique occurrences (Figure 6E).

Several epitopes were predicted for HLA class II-DRB alleles; VGP (n = 221) and MXP (n = 154) showing high frequencies for DRB1*01:01 and DRB1*01:04 alleles (Figure 7A, Table S3) with MXP higher for the latter than the former and vice versa for VGP. The four epitopes with highest VGP frequencies and coverage (COV) per HLA-DRB1 allele were; IILFQRTFSIPLGVI (7.05%) & COV = 6/8, RWGFRSGVPPKVVNY (3.52%) & COV = 3/8, WVIILFQRTFSIPLG (3.52%) & COV = 3/8 and IILFQRAISMLGIV (3.40%) & COV = 6/8 (Figure 7B). The four epitopes with the highest MXP frequencies and coverage per HLA-DRB1 allele were; PQYFTFDLTALKLIT (7.25%) & COV = 5/8, ITHFGKATNPLVRVN (4.38%) & COV = 5/8, TTAAIMLASYTITHF (3.00%) & COV = 2/8 and QKTYSFSDSTTAAIML (2.75%) & COV = 2/8 (Figure 7C). Based on VGP and HLA-DRB1 allele epitope distribution per Ebola species; IILFQRTFSIPLGVI was dominantly Zaire Ebolavirus, WVIILFQRTFSIPLG and RWGFRSGVPPKVVNY were solely Zaire Ebolavirus and IILFQRAISMLGIV was dominantly Reston Ebolavirus (Figure 7D). For the MXP epitopes and HLA-DRB1 allele distribution per Ebola species; PQYFTFDLTALKLIT was relatively equally distributed among the different Ebolaviruses, ITHFGKATNPLVRVN was solely Zaire Ebolavirus, TTAAIMLASYTITHF and QKTYSFSDSTTAAIML were dominantly Zaire Ebolavirus (Figure 7E).

Several epitopes were predicted for HLA class II-DQA-B alleles; VGP (n = 41) and MXP (n = 11) showing high frequencies for HLA-DQA1_03_01-DQB1_03_01 and HLA-DQA1_05_04-DQB1_05_03 (Figure 8A, Table S4) with VGP higher for the former than for the latter. The four epitopes with highest VGP frequencies per HLA-DQA-B allele coverage were; ETEYLFQVDNLTIVQ (13.55%) & COV = 3/9, FHKEGAFFLYDRLAS (8.64%) & COV = 1/9, RGVNFAEGVIAFLIL (7.24%) & COV = 3/9 and WVIILFQRTFSIPLG (6.78%) and COV = 1/9 (Figure 8B). The four epitopes with highest MXP frequencies per HLA-DQA-B allele coverage were; YSFDSTTAAIMLAS (33.33%) & COV = 2/7, HTPSGVASAFIEAT (33.33%) & COV = 5/7, IYSFDSTTAAIMLAS (8.33%) & COV = 1/7 and HTPSGVASAFIEAK (8.33%) & COV = 4/7 (Figure 8C). Based on VGP epitope and HLA-DQA-B distribution per Ebola species; ETEYLFQVDNLTIVQ and FHKEGAFFLYDRLAS was dominantly Zaire Ebolavirus, RGVNFAEGVIAFLIL was dominantly Sudan and Reston Ebolaviruses and WVIILFQRTFSIPLG was solely Zaire Ebolavirus (Figure 8D). Based on MXP epitope and HLA-DQA-B distribution per Ebola species; YSFDSTTAAIMLAS was relatively distributed equally among the different Ebolaviruses, HTPSGVASAFIEAT was dominantly Reston Ebolavirus, IYSFDSTTAAIMLAS was solely Reston Ebolavirus and HTPSGVASAFIEAK was equally distributed between Reston and Taï Ebolaviruses (Figure 8E).

Figure 5. Analysis of peptide antigenicity of viral sequences and their related frequency within a host population. Antigenic epitopes were predicted for MHC class I and MHC class II. Four (A) against three (B) epitopes for VGP and MXP respectively were validated for MHC class I alleles, 221 (C) against 154 (D) epitopes for VGP and MXP respectively were validated for MHC class II HLA-DRB1 alleles and 41 (E) against 11 (F) epitopes for VGP and MXP respectively were validated for MHC class II HLA-DQA-B alleles. Generally their peptide antigenicity was higher than host population frequency.

Figure 6. MHC class I alleles distribution for antigenic epitopes and different viral species. Frequency of antigenic peptides for both VGP and MXP were dominant for HLA*A02:01 and HLA*A02:02 alleles (A) with VGP frequency higher in the latter than in the former and vice versa. Epitope distribution for the different immune alleles showed that antigenic sites for each genomic structure (VGP and MXP) could be recognised by particular immune alleles while others were recognized by all (B, VGP and C, MXP). The identified epitopes were seen to occur variably among species with some epitopes being specific to a given viral species and others common to a species (D, VGP and E, MXP).

Figure 7. MHC class II HLA-DRB1 allele distribution for antigenic epitopes and different viral species. Frequency of antigenic peptides for both VGP and MXP were dominant for DRB1*01:01 and DRB1*01:04 alleles (A) with VGP frequency higher in the former than in the latter and vice versa. For the top four selected Epitopes, frequency distribution for the different immune alleles showed that antigenic sites for each genomic structure (VGP and MXP) could be recognised by particular immune alleles, others were recognized by all and some could not be recognized by any of the immune alleles (B, VGP and C, MXP). The identified epitopes were seen to occur variably among species with some epitopes being specific to a given viral specie and several others common to a species and for some species, none had the epitope site (D, VGP and E, MXP).

Figure 8. MHC class II HLA-DQA-B allele distribution for antigenic epitopes and different viral species. Frequency of antigenic peptides for both VGP and MXP were dominant for HLA-DQA1_05_04-DQB1_05_03 allele with that of VGP higher than MXP but only MXP presented a high frequency for HLA-DQA1_03_01-DQB1_03_01 allele (A). For the top four selected epitopes, frequency distribution for the different immune alleles showed that antigenic sites for each genomic structure (VGP and MXP) could be recognised by particular immune alleles, others were recognized by all and some could not be recognized by any of the immune alleles (B, VGP and C, MXP). The identified epitopes were seen to occur variably among species with some epitopes being specific to a given viral species and several others common to a species and for some species, none had the epitope site (D, VGP and E, MXP).

Prediction of transmission intensity (TRI) values in the wake of Ebola epidemic

Based on the data collected, peptide scores and peptide frequencies expressed in the viral population (simulating host population) were used to calculate the transmission intensity ratio predicting how fast or slow the virus will spread within its host population (Table 2).

The majority of the MHC HLA class I peptides showed a value greater than 1 while the majority of the MHC HLA class II peptides showed a value less than one. In general, $TRI < 1$, showed multiple expressed epitopes within each species and high distribution within the population and vice versa for $TRI > 1$. Zaire and Reston Ebolaviruses showed the lowest TRIs in the selected data.

Table 2. Transmission intensity values for Ebola disease. Epitopes predicted for MHC class I and II were predicted and their antigenic scores (PS)/peptide frequency within the host population were analyzed to determine values which could be used to determine epidemic outcomes of the disease. PS/PF > 1 was considered high transmission risk value. M = mixed, S = specific.

	Epitope	Peptide antigenic score (PS)	Peptide frequency (PF)	TRI = PS/PF	Species (Ebola virus)
HLA-class I VGP	FLLQLNETI	0,604656471	1	0,604656471	Zaire (S)
	FLTNTIRGV	1	0,074074074	13,5	Bundibugyo and Tai forest
	FQVDHLYV	0,928897512	0,888888889	1,045009701	Bundibugyo and Tai forest
	NMTNFLFQV	0,636955033	0,333333333	1,910865099	Bundibugyo and Tai forest
HLA-class I MXP	AIMLASYTV	1	1	1	Reston and Sudan
	ALLTGSYTI	0,849088454	0,2	4,245442269	Sudan & Marburg
	LTMVITPDY	0,915091155	0,3	3,050303849	Sudan & Marburg
HLA-DRB1 VGP	IILFQRTFSIPLGVI	0,314844562	1	0,314844562	Zaire (M)
	WVILFQRTFSIPLG	0,397193612	0,5	0,794387225	Zaire (S)
	RWGFRSGVPPKVNY	0,591798167	0,5	1,183596334	Zaire (S)
	IILFQRAISMLGIV	0,326136256	0,482758621	0,675567959	Reston
HLA-DRB1 MXP	PQYFTFDLTALKLIT	0,411019225	1	0,411019225	All
	ITHFGKATNPLVRVN	0,277306412	0,603448276	0,45953634	Zaire (S)
	TTAAIMLASYTITHF	0,354419504	0,413793103	0,856513801	Zaire (M)
	QKMYSFDSTTAIML	0,360970895	0,034482759	10,46815594	Zaire (M)
HLA-DQA-B VGP	ETEYLFEVDNLTYVQ	0,333333333	1	0,333333333	Zaire (M)
	FHKEGAFFLYDRLAS	0,41535682	0,637931034	0,65109988	Zaire (M)
	RGVNFAEGVIAFLIL	0,427280939	0,534482759	0,799428855	Sudan & Reston
	WVILFQRTFSIPLG	0,506293285	0,5	1,01258657	Zaire (S)
HLA-DQA-B MXP	YSFDSTTAIMLAS	0,439248283	1	0,439248283	All
	HTPSGVASAFLEAT	0,363570654	1	0,363570654	Reston (M)
	IYSFDSTTAIMLAS	0,454644019	0,25	1,818576075	Reston (S)
	HTPSGVASAFLEAK	0,53552584	0,25	2,142103361	Reston & Tai forest

Discussion

In this study, Ebolavirus amino acid sequence data was obtained for Zaire, Sudan, Côte d'Ivoire (Taï forest), Uganda (Bundibugyo) and Philippines (Reston) species. These sequences were subjected to phylogenetic analysis, which took into consideration all known functional motifs and domains for each sequence to construct the tree. It was observed that, species variation was higher for VGP compared to Nuclear and Matrix proteins of the viral genome. This observation could have resulted from pairwise alignments, whose identity scores accounted for species specific clustering observed.

The first two Ebolavirus species, Zaire and Sudan, identified in 1976 and later, Reston and Côte d'Ivoire, in 1989 and 1994, respectively, revealed after comparison that the four viral genomes differed from each other by 36.7–42.3% (Towner *et al.*, 2008). In 2008, Uganda Ebolavirus species was identified and its genomic sequence differed from previously known viruses by 31.7–42.4% (Towner *et al.*, 2008). Work done by Leroy *et al.* (2002) showed that the mean amino acid distances between the same NP subtypes was about 45%, compared to 65–66% in VGP. These values were higher than those observed in this work with probable difference being that we performed an amino acid sequence alignment, whilst the authors carried out a nucleotide sequence alignment. The longest branch on the tree associated with Marburg virus suggest an ancestral lineage which could be a consequence of evolutionary time wherein, Marburg virus and Ebolavirus were isolated in 1967 (Siebert *et al.*, 1967) and 1976 (Bowen *et al.*, 1977; Johnson *et al.*, 1977), respectively. They also show a genomic difference $\geq 50\%$ at the nucleotide level (Towner *et al.*, 2006). The presence of one gene overlap in Marburg virus compared to several in Ebolavirus delineates the main genomic difference between them (Kuhn, 2008).

The observed functional relationship between related hosts (mammal) and Ebolavirus proteins, identified within the interactome, showed direct interactions between the former, but none was observed between VP40 and ENV (gp160), and likewise to host proteins. Given the identification of *gag*, *pol*, and *env* (ENV) retroviral genes, the *pol* gene, which encodes reverse transcriptase (RT), is considered the most conserved of the retroviral elements (Simmons *et al.*, 2002). The polypeptide encoded by the *env* gene is cleaved into two proteins; a surface protein for receptor recognition, and a transmembrane subunit, which anchors the ENV complex to the membrane. The ENV transmembrane unit interacts directly with the cell membrane, favoring viral penetration (Bénil *et al.*, 2001) into the cell which has been elucidated for human immunodeficiency virus type 1 (HIV-1) (Connolly *et al.*, 1999; Xu *et al.*, 1998) and Ebolavirus (Sanchez *et al.*, 2001). The ENV-dependent pathways for HIV accounts for almost 50% of antigen presentation mediated principally by the lectin DC-SIGN (CD209). CD209 molecules, form part of dendritic cells equipped with molecular diversity for HIV-1 capture, which binds the viral envelope gp120 with high affinity. Efficient cross presentation of HIV antigens from incoming virions activates anti-HIV CD8⁺ cytotoxic T lymphocytes. The major role of CD209 in viral entry into the host cell, solidifies its identity within the predicted interactome. Though the observed functional roles include intracellular virion transport, protein signaling and peptide identification, associated with CD209, no direct interaction was observed with the viral structural genes

predicted on the network. This could suggest a domain specific function in CD209 for viral-host interactions, and though similar in Ebolaviruses, the functional pathway may be different.

The Ebolavirus VGP is synthesized as a secreted (sGP) or full-length transmembrane form, having distinct biochemical and biological properties. Multiple lines of evidence suggest that VGP plays a prominent role in clinical manifestations of Ebolavirus infection. A change of immune response could result from inhibition of neutrophil activation by sGP, while the transmembrane form could contribute to symptoms of hemorrhagic fever due to viral attack of reticuloendothelial and blood vessel cells (Ströher *et al.*, 2001; Yang *et al.*, 1998). The different host cell targets and roles played by sGP and the transmembrane form, makes them potential targets that could be exploited for control of Ebolaviral burden. The type of preferred cells by each VGP form, should determine the type of pathogenic manifestation and nature of transmission intensity observed within the population. Therefore, drug targets common across viral species or specific to most virulent forms will be vital for quick clearance of an infection.

The assembly and budding of Ebolaviruses from the cell membrane is a process directed by the viral matrix protein, VP40 (Harty *et al.*, 2000; Panchal *et al.*, 2003). VP40 alone is capable of assembling and budding filamentous virus-like particles from cells (Geisbert & Jahrling, 1995; Johnson *et al.*, 2006; Noda *et al.*, 2002). The specific interaction of VP40 with membrane-associated VGP and VP24 during the budding process, is a possible occurrence. Work done by Bornholdt *et al.* (2013), showed that binding of VP40 to RNA forms a ring structure at a perinuclear location. Observation of these rings was only in infected cells, but not mature purified Ebolaviruses (Gomis-Rüth *et al.*, 2003), which suggests that, the ring structure was not necessary for matrix assembly. Therefore the VP40 RNA-binding rings could possess a critical function within infected cells (Gomis-Rüth *et al.*, 2003; Hoenen *et al.*, 2010). Previous data showed that, recombinant viruses that have an RNA-binding knock-out mutation for VP40 (R134A) failed to replicate (Hoenen *et al.*, 2005), suggesting a dual role for VP40, which includes: viral transcription with assembly, and budding, for the release of new viral particles. Given that VP40 is structural and shows possible interactions with VGP, suggests a contributory role to host immune activation pathway, and though directing different cellular processes, their immune-interaction makes them potential targets for vaccine design and control of Ebolavirus disease.

The close relationship between Ebola genomic structures; VP40, CD209 and ENV and between HIV-1 and Ebolavirus, motivated our use of VP40 and VGP for the prediction of antigenic peptide epitopes required for generating potent host antibodies against the virus. The interactive nature between ENV and CD209 proteins, could explain the downstream signaling reactions necessary for cytokine activation. Identifying target signaling molecules activated by interaction of CD209 with Ebolavirus antigens, could suggest a possible path for curbing the loss of body fluids during epidemic outbreaks. The presence of LCK and its kinase function suggests a protein signaling role in driving the infection and loss of fluid pathway associated with Ebolavirus. Its involvement in CD4 and CD8 binding makes its pathway, an interesting target

area for Ebolavirus disease control. The signaling of T cell receptors (TCR), which results from peptide-MHC binding, requires the phosphorylation of immunoreceptor, tyrosine-based activation motifs (ITAMs) and CD3 complex, associated with SFK Lck (Palacios & Weiss, 2004). Generally, it is accepted that, the activation of T cells favors Lck activation by pMHC binding. Given the stable association of Lck to CD4 and CD8 coreceptors (Rudd *et al.*, 1988; Veillette *et al.*, 1988), other interaction models propose that MHC binding activates Lck by trans-autophosphorylation. Basing on this information, the role-played by LCK within the network and its interaction with immune cells becomes very clear. The implication of the LCK gene in Ebolavirus disease intensity could be related to a hyper activation cascade leading to massive release of cytokines (Cytokine storm), which is an event common to patients with septic shock (Mackenzie & Lever, 2007). The Lck protein kinase pathway identified in the network as a cluster, showed interaction with proteins including SRC (Proto-oncogene tyrosine-protein kinase), MERTK (Tyrosine-protein kinase), TYRO3 (Tyrosine-protein kinase receptor), UFO (Tyrosine-protein kinase receptor), ACK1 (Activated CDC42 kinase 1), BMPR2 (Bone morphogenic protein receptor type-2), YES (Tyrosine-protein kinase) and FYN (tyrosine protein kinase). The UniProt Knowledge Base (KB) (www.uniprot.org), associates these proteins to downstream ligand binders and signal transducers, which activates other proteins in a phosphokinase related interaction type. Several of these are involved in disease pathways including SRC (colon carcinoma), MERTK (retinitis pigmentosa, MIM: 613862), UFO (thyroid tumorigenesis) and BMPR2 (pulmonary hypertension primary, PPHL 1, MIM: 178600). This cluster suggests that, the role played by LCK in kinase function and implicates it to epidemic levels observed for Ebolavirus disease.

MiRNAs are potentially known to regulate several proteins (Grimson *et al.*, 2007) and also modulate their concentrations in a dose-dependent manner (Baek *et al.*, 2008; Selbach *et al.*, 2008). After Grigoryev *et al.* (2011) identified several miRNAs significantly differentially expressed in T-lymphocytes, they showed a decrease in mRNA expression for target genes of upregulated miRNA molecules. This was confirmed by inhibiting two of these miRNA targets (miR-221 and miR-155) in CD4⁺ cells, which increased lymphocyte proliferation by liberating four (PIK3R1, FOS, IRS2, and IKBKE) gene targets, hence favouring their proliferation and survival. MiR-155 overexpression has been shown to increase *in vivo* antiviral CD8⁺ T cell responses (Gracias *et al.*, 2013). Among the highlighted miRNA gene targets above, IKKE (IKBKE) was the only one captured in this study and identified in the top three proteins. MiRNAs and their related disease pathway, hsa05131 (microbial infection, shigellosis) and cellular process, hsa01430 (cell communication), could suggest the possibility of successful viral transmission and disease occurrence through gene regulation within each pathway. Knowledge of other regulatory elements such as transcription factors and drug targets for related genes within the network, suggests a complex regulatory platform for the control of Ebolavirus disease. Transcription factors to gene targets like CD209 known for antigen binding and immune presentation could be further molecularly analysed as potential targets for gene regulation and disease prevention within the host.

Gene products in the HLA region (chromosome 6) facilitate antigenic protein processing and peptides presentation to CD8⁺ cytotoxic

T lymphocytes (CTLs) and CD4⁺ T helper cells (Coffin, 1996; Dyer *et al.*, 1997). HLA allele's multiplicity has favoured great species diversity as individuals differentially adapt to the wide range of microbial agents such as HIV-1. Optimal immune defence against HIV-1 could depend on the combined intervention of CD4⁺ and CD8⁺ lymphocytes (Rosenberg *et al.*, 1997) wherein each polymorphic class I and class II loci contributes separately to the resultant immune response. CD209 favors specificity of HIV antigen presentation to MHC-II alleles while dendritic cells promote HIV-specific CD4⁺ cells activation and eventual infection. Similarity between HIV-1 and Ebolavirus in the use of antigen-immune interactions to penetrate the host cell could be seen from related genes identified in the network. Components of the immune system with protective role against infection of Ebolavirus are not clearly known due to the fact that the pathogen replicates at unexpected high rates, overwhelming host infected cells and immune defenses with related protein synthesis mechanisms (Sanchez *et al.*, 2001).

From our study, HLA class I peptide antigenicity and frequency values for VGP and MXP within the pathogen population, showed no significant statistical difference. Their specificity was tied to HLA*A02:01 and HLA*A02:02 alleles for all the validated peptides, unlike differential selectivity observed for other tested host alleles. This was in conformity with host expression of HLA*A02 allele against the HIV-1 virus expressed in the Botswana (Novitsky *et al.*, 2001) and Rwandan population (Tang *et al.*, 1999). The FLLQLNETI epitope for VGP was specific to the Zaire Ebolavirus with a single occurrence while the other VGP and all the MXP epitopes were related to other viral species with multiple occurrences. Although the Zaire strain has been considered the most lethal (Feldmann *et al.*, 1994; Sanchez *et al.*, 1996), the observed transmission intensity (TRI) for the FLLQLNETI epitope was relatively low compared to the FLTNTIRGV epitope for VGP (Bundibugyo and Tai forest), with a TRI twice that of the former. This brings two questions to mind; 1) Is the Ebolavirus epidemic caused mainly by the Zaire Ebola species? and, 2) What causes the sudden change in antigenicity that results in epidemic trends currently observed in West Africa (Liberia, Sierra Leone and Equatorial Guinea). One possible solution could be due to the high phylogenic interspecies differences observed within the VGP genomic structure, suggesting a species specific (pairwise identity) pathogenic driven occurrence. The grouping of Zaire Ebolavirus species with other related species and not vice versa could suggest a derivation of other species from the Zaire virus, hence the host immunity may have adapted to keeping the Zaire strain below epidemic levels. Introduction of high antigenic variants from related strains could be the driving force of the epidemic levels observed.

HLA class II antigenic against frequency values for identified epitopes, presented with high statistical differences for VGP and MXP genomic structures. The validated peptides showed high antigenic values for HLA II thus favouring a high TRI. The most common host allelic frequencies for VGP and MXP were HLA-DRB1*01:01 and HLA-DRB1*01:04 and were consistent with that of the Botswana population carrying the HIV-1 virus though different for other alleles observed with high frequencies in that population. The viral species identified for selected HLA-II peptide epitopes were predominantly and specifically Zaire Ebolavirus. All epitopes presented with low TRIs except for QKMYSFDST-TAAIML which had a TRI > 10 for the MXP genomic structure.

This was contrary to the HLA-I high TRI data observed for VGP genomic structure of the Zaire Ebolavirus. Given that specificity of antigen presentation to HLA-II immune proteins is facilitated by CD209, MXP should contribute highly to immune response and signal transmissions downstream. MXP has not always been associated with antigenic interactions but the fact that it could interact with VGP at the viral surface makes its contribution to immune response of considerable importance.

Antigenicity of peptide epitopes for HLA-DQA-B were statistically significant for VGP and not MXP. The VGP validated epitopes were dominant for Zaire Ebolavirus species while those of MXP were dominant for Reston Ebolavirus. Both genomic structures showed high frequencies for HLA-DQA1_05_04-DQB1_05_03 while only frequencies for MXP epitopes were high for HLA-DQA1_03_01-DQB1_03_01. Only HLA-DQA1-B1*03 was in conformity with the alleles identified for the Botswana HIV-1 infected individuals. This allele was specific to MXP and dominantly Reston Ebolavirus, suggesting a possible contribution to immune activation. The study from Botswana genotyped humans for HLA-II allele frequencies and from our study suggests immune specificity to different Ebola genome structures. This observation highlights the difference in antigenicity between the Ebolavirus and HIV-1 virus. Though the low TRI and lack of significant difference for MXP epitopes, the observed antigenic specificity with that of VGP genomic structure could play vital roles for viral pathogenesis within the host.

The multiple occurrences of several antigenic epitopes for a given viral species could suggest repeat sequences in tandem within the genome known as retrotransposons, which favor high genome variation and transmission. Previous work done by several authors on the similarity of reverse transcriptase (*pol* gene) in retroelements including retrotransposons and retroviruses such as Ebolavirus, suggest that retroviruses could have evolved from long terminal repeat (LRT) retrotransposons by acquiring a functional *env* gene (McClure, 1991; Xiong & Eickbush, 1990). Therefore, the *env* gene (VGP) should be critical for viral transmission in its host and could be targeted with anti-retroviral technology currently used in HIV patients. Its identification in the interactome suggests an interaction with several human proteins involved in signal transduction. The interplay between the MXP, VGP and their related interacting human proteins makes all these various genes potential targets for Ebolavirus control through drug/vaccine design or the use of genetic technology through interference of viral replication.

Conclusion

This study is the first of its kind that has exploited all possible viral antigenic targets and possible systemic routes in the host that could be targeted individually or together for the control of Ebolavirus disease. From this study, the phylogenetic relationship distinguished in the viral envelope of Ebolaviruses and their similarity to HIV-1 viral genomic structures shows that common pathways to pathogenic manifestations within the host are eminent. The ability to clearly understand cellular pathways driving pathogenicity is critical for arresting rapid and abundant cytokine activation leading to shock and organ malfunction. Though the Zaire Ebolavirus has been identified as the most virulent; the TRI for this species generally shows a low transmission rate compared to others, which is in

accordance with viral-host cohabitation and adaptation. This suggests that, epidemic occurrences of the Ebolavirus disease could result from related species which have no genetic adaptation with its host. The sequence repeats predicted within antigenic epitopes, occurred more for the non-Zaire Ebola viral species. The identified antigenic epitope targets for this species could be used for drug/vaccine designs. Potential microRNA's, transcription factors and known drugs identified for most gene targets, suggest possible points in the viral pathogenic pathway that could be the focus of modern genetic technology for arresting viral replication and epidemics.

Methods

Phylogenetic analysis of retrieved Ebola and related human protein sequences

The Geneious software (V.6.1.8) (Kearse *et al.*, 2012) and its routine plugin packages were used for extracting and analyzing obtained sequences for various Ebolavirus species. The key word "Ebolavirus", was used to search the UniProtKB (Magrane & Consortium, 2011) (Table S1), resulting in 348 protein sequence variants (database searched in September, 2014), from the five main regional strains highly associated with Ebola Hemorrhagic Fever (Dowell *et al.*, 1999), which are: Zaire Ebolavirus, Sudan Ebolavirus, Bundibugyo Ebolavirus (from Uganda), Tai Ebolavirus (from Côté d'Ivoire) and Reston Ebolavirus (from the Philippines). The sequences also included related mammalian protein sequences (*Homo sapiens*, *Mus musculus*, *Rattus norvegicus*) associated with the viral genome based on genetic motifs, domains and Gene Ontology (GO) terms (Ashburner *et al.*, 2000) captured in various phyla. Sequences were aligned using a novel method for rapid multiple sequence alignment based on fast Fourier transform (MAFFT V7.017) (Katoh *et al.*, 2002). The parameters involved an auto alignment algorithm, a Blosum 62 (Mihalek *et al.*, 2007) scoring matrix, a gap penalty of 1.5 and an offset value = 0.123. The aligned sequences were then subjected to a Geneious tree builder which uses pairwise genetic distances from the sequence alignment, for tree construction, while basing on the following parameters: Jukes & Cantor (1969) genetic distance model for amino acid substitution, Unweighted Pair Group Method with Arithmetic mean (UPGMA) (Yu-Shiang *et al.*, 2014), for tree building and jackknifing (Jones, 1974) while sampling for the best tree (random seed = 932, 120; replicates = 500; support threshold = 80%). Sequence clusters identified from the tree were exported in FASTA format for MHC HLA class I and II epitope predictions. UniProt accession numbers were extracted and used for protein interaction predictions.

Protein interaction predictions

UniProt accession numbers for the 348 sequences were loaded into Cytoscape software (V.2.8.1) (Shannon *et al.*, 2003), a bioinformatics platform for molecular interaction networks visualization with integration of identified interactions to other biological data, and its routine plugins were used to predict and analyze the network. APID2NET (Hernandez-Toro *et al.*, 2007), was used to predict the protein interaction map. This tool queries interactome data from the Agile Protein Interaction Data Analyzer (APID) server (<http://bioinfow.dep.usal.es/apid/>) providing an interactive analysis of individual protein-protein interaction information into a network. During the interaction, the program considers parameters like GO terms,

InterPro domains (Hunter *et al.*, 2012), PFAM domains (Bateman *et al.*, 2002), experimental data for validated interactions and all necessary genetic information captured in the APID database for each protein. Network Analyzer, a Cytoscape plugin was used to evaluate node attributes for the complex network and it considers parameters such as: degree distribution (Barabási & Oltvai, 2004; Diestel, 2005), neighborhood connectivity (Maslov & Sneppen, 2002), shortest paths (Watts & Strogatz, 1998), clustering coefficients (Barabási & Oltvai, 2004), shared neighbors, topological coefficients (Stelzl *et al.*, 2005), stress centrality (Shimbel, 1953), betweenness centrality (Brandes, 2001) and closeness centrality (Newman, 2003) to name a few. The Biological Network Genes Ontology tool (BiNGO) was used to determine full GO terms significantly overrepresented in our gene set (Maere *et al.*, 2005) with species target *Homo sapiens*. The plugin CyTargetLinker (Kutmon *et al.*, 2013) was used to extend the interaction network with regulatory elements such as: miRNA-target (Hsu *et al.*, 2014), drug-target or TF-target interactions and TargetScan (Friedman *et al.*, 2009) to identify genes regulated by these elements. Identified microRNA molecules were associated to particular pathways based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa *et al.*, 2014).

Epitope predictions, antigenic evaluation and MHC HLA class I & II distribution

The various Ebolavirus sequence sub clusters, [virion envelope glycoprotein, VGP (N = 63) and Matrix protein, MXP (N = 19)] were separately submitted for: 1) MHC class I epitope predictions at the HLA restrictor 1.2 Server (Erup Larsen *et al.*, 2011), a tool for patient-specific predictions of HLA restriction elements and optimal epitopes (8 – 11 amino acids) within peptides and 2) MHC class II epitope predictions at Net MHCIIpan 3.0 server (Karosiene *et al.*, 2013) which predicts binding of peptides (9 – 19 amino acids long) to MHCII alleles for three human isotypes (HLA-DR, HLA-DP and HLA-DQ). Given the vacuum of information on the MHC HLA class I & II allele types in Ebola stricken populations, selected alleles associated with the HIV virus in African related populations (Tang *et al.*, 1999; Novitsky *et al.*, 2001) were considered for use in this study since both viral species have been shown to possess the homologous hairpin structural envelope glycoprotein (gp160) (Watanabe *et al.*, 2000), which is a target for several HIV vaccines. Based on previous work done by our group (Ojwang *et al.*, 2014), nine monomers predictions were better preferred for MHC class I, HLA-A alleles while 15 monomers were better preferred for MHC class II, HLA-DRB and HLA-DQ alleles (Table 3).

Peptides were selected on two criteria; threshold affinity for strong binders (0.5 for percentage rank and 50 for IC50) and threshold affinity for weak binders (1 for percentage rank and 100 for IC50). Only strong binders were considered for further analysis. Files containing strong binder epitopes for each Ebola structural genomic class variant and for each species was generated in FASTA format and submitted to Vaxijen software (V.2.0) (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>). Viral antigenic prediction filter was used at a cutoff of 0.5 and all epitopes with antigenicity values (peptide score) ≥ 0.5 separately for MHC class I and II alleles were further considered for analysis.

Table 3. MHC class I and II allele variants used to predict Ebola antigenic peptide epitopes. Nine monomers and fifteen monomers were predicted for HLA-I and II alleles respectively.

MHC class I	MHC class II	
HLA-A alleles	HLA-DRB allele	HLA-DQ alleles
HLA-A01:01	DRB-01:01	DQA1-01:01:DQB1-02:01
HLA-A01:02	DRB-01:04	DQA1-01:04:DQB1-02:04
HLA-A02:01	DRB-02:01	DQA1-03:01:DQB1-03:01
HLA-A02:02	DRB-02:04	DQA1-03:03:DQB1-03:04
HLA-A02:04	DRB-03:01	DQA1-04:01:DQB1-04:01
	DRB-03:04	DQA1-04:04:DQB1-04:04
	DRB-04:01	DQA1-05:01:DQB1-05:01
	DRB-04:04	DQA1-05:04:DQB1-05:03
	DRB-05:01	DQA1-06:02:DQB1-06:04
	DRB-05:04	
	DRB-06:01	
	DRB-06:04	
	DRB-07:01	
	DRB-07:04	

Calculating Transmission Intensity (TRI) for selected epitopes

For every validated epitope in our data, the peptide scores and peptide frequencies were considered for the calculation of a novel value called TRANSMISSION INTENSITY (TRI) which translates how high or low the disease transmission will develop into an epidemic situation within the population. The value was based on the ratio of peptide score (PS)/peptide frequency (PF). If TRI was around one, it means the transmission is stable and could easily be contained to elimination. If $TRI > 1$, it means, the transmission poses a bigger risk to the infected individuals given that high antigenicity will activate a high kinetic and signalling response pathway (Yamashita *et al.*, 1998) reaching epidemic levels due to the host immune system generating high antibody levels against the virus possibly leading to hypotensive shock, characterized by; fever, myalgia, malaise, severe bleeding and coagulation abnormalities, including gastrointestinal bleeding, rash, and a range of hematological irregularities, such as lymphopenia and neutrophilia (Colebunders & Borchert, 2000). If $TRI < 1$, it means, the transmission poses a low risk within the population and could easily be eliminated by the host immune system. Given that antigenicity is low and peptide frequency is high, the host quickly develops immunity against the pathogen. The different transmission intensities will depend on the MHC HLA allele type and its frequency within the population.

Statistical analysis

Batch analysis for evaluating node attributes for the network was used to fit the best line between the data points using the least squares method (Weisstein, www.mathworld.wolfram.com). GO terms for the predicted interactome were validated based on the Hypergeometric test after Benjamini & Hochberg False Discovery Rate (FDR) (Benjamini & Hochberg, 1995) correction and at a

significance level of 0.05. Files for each structural genomic class variant per species was used to evaluate frequency distribution for epitopes against peptide scores, HLA allele frequency and frequency within each species protein sequence (supporting [Table S2](#), [Table S3](#) and [Table S4](#)). After normalizing every value to the largest in each dataset (epitope peptide antigenic score and epitope peptide frequency) of different Ebola sequences, paired student one tailed t-test was performed to compare means between the groups ($\alpha = 0.05$). This was to evaluate how epitope targets were distributed across different Ebola species and their epidemiological implications.

Data availability

The complete list of UniProt KB protein sequence codes for extracted Ebolavirus sequences used for analysis can be found in [Table S1](#).

Author contributions

All authors contributed to the ideas that led to the development of the manuscript. A.D.O and D.A.A initiated the research idea, D.A.A did the primary research, analysed the data and wrote the manuscript. A.D.O did the proof reading and edited the manuscript for submission.

Competing interests

No competing interests were disclosed.

Grant information

The author(s) declared that no grants were involved in supporting this work.

Supplementary Tables

Table S1.

UniProt KB protein sequence codes for extracted Ebolavirus sequences used for analysis

[Click here to access the data.](#)

Table S2.

Predicted antigenic epitopes evaluated for frequency distribution against MHC class I immune alleles and different Ebola species

[Click here to access the data.](#)

Table S3.

Predicted antigenic epitopes evaluated for frequency distribution against MHC class II HLA-DRB1 immune alleles and different Ebola species

[Click here to access the data.](#)

Table S4.

Predicted antigenic epitopes evaluated for frequency distribution against MHC class II HLA-DQA-B immune alleles and different Ebola species

[Click here to access the data.](#)

References

- Ashburner M, Ball CA, Blake JA, *et al.*: **Gene ontology: tool for the unification of biology.** *The Gene Ontology Consortium. Nat Genet.* 2000; **25**(1): 25–29.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Baek D, Villén J, Shin C, *et al.*: **The impact of microRNAs on protein output.** *Nature.* 2008; **455**(7209): 64–71.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Barabási AL, Oltvai ZN: **Network biology: understanding the cell's functional organization.** *Nat Rev Genet.* 2004; **5**(2): 101–113.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Baskerville A, Bowen ET, Platt GS, *et al.*: **The pathology of experimental Ebola**

virus infection in monkeys. *J Pathol.* 1978; **125**(3): 131–138.

[PubMed Abstract](#) | [Publisher Full Text](#)

Baskerville A, Fisher-Hoch SP, Neild GH, *et al.*: **Ultrastructural pathology of experimental Ebola haemorrhagic fever virus infection.** *J Pathol.* 1985; **147**(3): 199–209.

[PubMed Abstract](#) | [Publisher Full Text](#)

Bateman A, Birney E, Cerruti L, *et al.*: **The Pfam protein families database.** *Nucleic Acids Res.* 2002; **30**(1): 276–280.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Bénit L, Dessen P, Heidmann T: **Identification, phylogeny, and evolution of**

retroviral elements based on their envelope genes. *J Virol.* 2001; 75(23): 11709–11719.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Benjamini Y, Hochberg Y: **Controlling the false discovery rate: a practical and powerful approach to multiple testing.** *J R Statist Soc.* 1995; 57(1): 289–300. [Reference Source](#)

Bornholdt ZA, Noda T, Abelson DM, *et al.*: **Structural rearrangement of ebola virus VP40 begets multiple functions in the virus life cycle.** *Cell.* 2013; 154(4): 763–774.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Bowen ET, Lloyd G, Harris WJ, *et al.*: **Viral haemorrhagic fever in southern Sudan and northern Zaire. Preliminary studies on the aetiological agent.** *Lancet.* 1977; 309(8011): 571–573.

[PubMed Abstract](#) | [Publisher Full Text](#)

Brandes U: **A faster algorithm for betweenness centrality.** *J Math Sociol.* 2001; 25(2): 163–177.

[Publisher Full Text](#)

Burke JP: **Infection control - a problem for patient safety.** *N Engl J Med.* 2003; 348(7): 651–6.

[PubMed Abstract](#) | [Publisher Full Text](#)

Coffin JM: **HIV viral dynamics.** *AIDS.* 1996; 10(Suppl 3): S75–S84.

[PubMed Abstract](#)

Colebunders R, Borchert M: **Ebola haemorrhagic fever—a review.** *J Infect.* 2000; 40(1): 16–20.

[PubMed Abstract](#) | [Publisher Full Text](#)

Connolly BM, Steele KE, Davis KJ, *et al.*: **Pathogenesis of experimental Ebola virus infection in guinea pigs.** *J Infect Dis.* 1999; 179(Suppl 1): S203–S217.

[PubMed Abstract](#) | [Publisher Full Text](#)

Diestel R: **Graph theory.** Springer-Verlag, Heidelberg, ISBN 3-540-26182-6. 2005. [Reference Source](#)

Dowell SF, Mukunu R, Ksiazek TG, *et al.*: **Transmission of Ebola hemorrhagic fever: a study of risk factors in family members, Kikwit, Democratic Republic of the Congo, 1995.** Commission de Lutte contre les Epidémies à Kikwit. *J Infect Dis.* 1999; 179(Suppl 1): S87–S91.

[PubMed Abstract](#) | [Publisher Full Text](#)

Dyer WB, Geczy AF, Kent SJ, *et al.*: **Lymphoproliferative immune function in the Sydney Blood Bank Cohort, infected with natural nef/long terminal repeat mutants, and in other long-term survivors of transfusion-acquired HIV-1 infection.** *AIDS.* 1997; 11(13): 1565–74.

[PubMed Abstract](#) | [Publisher Full Text](#)

Erup Larsen M, Kloverpris H, Stryhn A, *et al.*: **HLArestrictor—a tool for patient-specific predictions of HLA restriction elements and optimal epitopes within peptides.** *Immunogenetics.* 2011; 63(1): 43–55.

[PubMed Abstract](#) | [Publisher Full Text](#)

Feldmann H, Nichol ST, Klenk HD, *et al.*: **Characterization of filoviruses based on differences in structure and antigenicity of the virion glycoprotein.** *Virology.* 1994; 199(2): 469–473.

[PubMed Abstract](#) | [Publisher Full Text](#)

Friedman RC, Farh KK, Burge CB, *et al.*: **Most mammalian mRNAs are conserved targets of microRNAs.** *Genome Res.* 2009; 19(1): 92–105.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Geisbert TW, Jahrling PB: **Differentiation of filoviruses by electron microscopy.** *Virus Res.* 1995; 39(2–3): 129–150.

[PubMed Abstract](#) | [Publisher Full Text](#)

Gomis-Rüth FX, Dessen A, Timmins J, *et al.*: **The matrix protein VP40 from Ebola virus octamerizes into pore-like structures with specific RNA binding properties.** *Structure.* 2003; 11(4): 423–433.

[PubMed Abstract](#) | [Publisher Full Text](#)

Gracias DT, Stelekati E, Hope JL, *et al.*: **The microRNA miR-155 controls CD8⁺ T cell responses by regulating interferon signaling.** *Nat Immunol.* 2013; 14(6): 593–602.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Grigoryev YA, Kurian SM, Hart T, *et al.*: **MicroRNA regulation of molecular networks mapped by global microRNA, mRNA, and protein expression in activated T lymphocytes.** *J Immunol.* 2011; 187(5): 2233–2243.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Grimson A, Farh KK, Johnston WK, *et al.*: **MicroRNA targeting specificity in mammals: determinants beyond seed pairing.** *Mol Cell.* 2007; 27(1): 91–105.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Harty RN, Brown ME, Wang G, *et al.*: **A PPxY motif within the VP40 protein of Ebola virus interacts physically and functionally with a ubiquitin ligase: implications for filovirus budding.** *Proc Natl Acad Sci U S A.* 2000; 97(25): 13871–13876.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Hernandez-Toro J, Prieto C, Las Rivas DJ: **APID2NET: unified interactome graphic analyzer.** *Bioinformatics.* 2007; 23(18): 2495–2497.

[PubMed Abstract](#) | [Publisher Full Text](#)

Hoenen T, Biedenkopf N, Zielecki F, *et al.*: **Oligomerization of Ebola virus VP40 is essential for particle morphogenesis and regulation of viral transcription.** *J Virol.* 2010; 84(14): 7053–7063.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Hoenen T, Volchkov V, Kolesnikova L, *et al.*: **VP40 octamers are essential for**

Ebola virus replication. *J Virol.* 2005; 79(3): 1898–905.

[PubMed Abstract](#) | [Free Full Text](#)

Hsu SD, Tseng YT, Shrestha S, *et al.*: **miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions.** *Nucl Acids Res.* 2014; 42(Database issue): D78–D85.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Hunter S, Jones P, Mitchell A, *et al.*: **InterPro in 2011: new developments in the family and domain prediction database.** *Nucleic Acids Res.* 2012; 40(Database issue): D306–D312.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Johnson KM, Lange JV, Webb PA, *et al.*: **Isolation and partial characterisation of a new virus causing acute haemorrhagic fever in Zaire.** *Lancet.* 1977; 309(8011): 569–571.

[PubMed Abstract](#) | [Publisher Full Text](#)

Johnson RF, Bell P, Harty RN: **Effect of Ebola virus proteins GP, NP and VP35 on VP40 VLP morphology.** *Virol J.* 2006; 3: 31.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Jones HL: **Jackknife estimation of functions of stratum means.** *Biometrika.* 1974; 61(2): 343–348.

[Publisher Full Text](#)

Jukes TH, Cantor CR: **Evolution of Protein Molecules.** New York: Academic Press. 1969: 21–132.

[Reference Source](#)

Kanehisa M, Goto S, Sato Y, *et al.*: **Data, information, knowledge and principle: back to metabolism in KEGG.** *Nucleic Acids Res.* 2014; 42(Database issue): D199–D205.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Karosiene E, Rasmussen M, Blicher T, *et al.*: **NetMHCIIpan-3.0, a common pan-specific MHC class II prediction method including all three human MHC class II isotypes, HLA-DR, HLA-DP and HLA-DQ.** *Immunogenetics.* 2013; 65(10): 711–724.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Katoh K, Misawa K, Kuma K, *et al.*: **MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform.** *Nucleic Acids Res.* 2002; 30(14): 3059–3066.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Kearse M, Moir R, Wilson A, *et al.*: **Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data.** *Bioinformatics.* 2012; 28(12): 1647–1649.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Kuhn JH: **Filoviruses. A compendium of 40 years of epidemiological, clinical, and laboratory studies.** *Arch Virol Suppl.* Springer, Wien. ISBN 978-3-211-69495-4, 2008; 20: 13–360.

[PubMed Abstract](#) | [Publisher Full Text](#)

Kutmon M, Kelder T, Mandaviya P, *et al.*: **CyTargetLinker: a cytoscape app to integrate regulatory interactions in network analysis.** *PLoS One.* 2013; 8(12): e82160.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Leroy EM, Baize S, Mavoungou E, *et al.*: **Sequence analysis of the GP, NP, VP40 and VP24 genes of Ebola virus isolated from deceased, surviving and asymptotically infected individuals during the 1996 outbreak in Gabon: comparative studies and phylogenetic characterization.** *J Gen Virol.* 2002; 83(Pt 1): 67–73.

[PubMed Abstract](#)

Mackenzie I, Lever A: **Management of sepsis.** *BMJ.* 2007; 335(7626): 929–932.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Maere S, Heymans K, Kuiper M: **BINGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks.** *Bioinformatics.* 2005; 21(16): 3448–3449.

[PubMed Abstract](#) | [Publisher Full Text](#)

Magrane M, Consortium U: **UniProt Knowledgebase: a hub of integrated protein data.** *Database (Oxford).* 2011; 2011: bar009.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Maslov S, Sneppen K: **Specificity and stability in topology of protein networks.** *Science.* 2002; 296(5569): 910–913.

[PubMed Abstract](#) | [Publisher Full Text](#)

Mayo MA, Pringle CR: **Virus taxonomy—1997.** *J Gen Virol.* 1998; 79(Pt 4): 649–657.

[PubMed Abstract](#)

McClure MA: **Evolution of retroposons by acquisition or deletion of retrovirus-like genes.** *Mol Biol Evol.* 1991; 8(6): 835–856.

[PubMed Abstract](#)

Mihalek I, Res I, Lichtarge O: **Background frequencies for residue variability estimates: BLOSUM revisited.** *BMC Bioinformatics.* 2007; 8: 488.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Newman MEJ: **A measure of betweenness centrality based on random walks.** *arXiv cond-mat/0309045.* 2003.

[Reference Source](#)

Noda T, Sagara H, Suzuki E, *et al.*: **Ebola virus VP40 drives the formation of virus-like filamentous particles along with GP.** *J Virol.* 2002; 76(10): 4855–4865.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Novitsky V, Rybak N, McLane MF, *et al.*: **Identification of human**

immunodeficiency virus type 1 subtype C Gag-, Tat-, Rev-, and Nef-specific elispot-based cytotoxic T-lymphocyte responses for AIDS vaccine design. *J Virol.* 2001; **75**(19): 9210–9228.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Ojwang AME, Daniel AA, Anton D, *et al.*: *In silico* identification of universal HLA stimulating B and T-cell restricted mage epitopes for vaccine development. *OJB. Online Journal of Bioinformatics.* 2014; **15**(1): 114–132.

[Reference Source](#)

Palacios EH, Weiss A: **Function of the Src-family kinases, Lck and Fyn, in T-cell development and activation.** *Oncogene.* 2004; **23**(48): 7990–8000.

[PubMed Abstract](#) | [Publisher Full Text](#)

Panchal RG, Ruthel G, Kenny TA, *et al.*: *In vivo* oligomerization and raft localization of Ebola virus protein VP40 during vesicular budding. *Proc Natl Acad Sci U S A.* 2003; **100**(26): 15936–15941.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Parren PW, Geisbert TW, Maruyama T, *et al.*: **Pre- and postexposure prophylaxis of Ebola virus infection in an animal model by passive transfer of a neutralizing human antibody.** *J Virol.* 2002; **76**(12): 6408–6412.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Ploegh HL: **Viral strategies of immune evasion.** *Science.* 1998; **280**(5361): 248–253.

[PubMed Abstract](#) | [Publisher Full Text](#)

Rosenberg ES, Billingsley JM, Caliendo AM, *et al.*: **Vigorous HIV-1-specific CD4⁺ T cell responses associated with control of viremia.** *Science.* 1997; **278**(5342): 1447–1450.

[PubMed Abstract](#) | [Publisher Full Text](#)

Rudd CE, Trevillyan JM, Dasgupta JD, *et al.*: **The CD4 receptor is complexed in detergent lysates to a protein-tyrosine kinase (pp58) from human T lymphocytes.** *Proc Natl Acad Sci USA.* 1988; **85**(14): 5190–5194.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Sanchez A, Khan AS, Zaki SR, *et al.*: **Filoviridae: Marburg and Ebola viruses.** In Knipe DM and Howley PM (ed.), *Fields virology*. Lippincott, Williams and Wilkins, Philadelphia, Pa. 2001: 1279–1304.

Sanchez A, Kiley MP, Holloway BP, *et al.*: **Sequence analysis of the Ebola virus genome: organization, genetic elements, and comparison with the genome of Marburg virus.** *Virus Res.* 1993; **29**(3): 215–240.

[PubMed Abstract](#) | [Publisher Full Text](#)

Sanchez A, Trappier SG, Mahy BWJ, *et al.*: **The virion glycoproteins of Ebola viruses are encoded in two reading frames and are expressed through transcriptional editing.** *Proc Natl Acad. Sci.* 1996; **93**(8): 3602–3607.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Selbach M, Schwanhäusser B, Thierfelder N, *et al.*: **Widespread changes in protein synthesis induced by microRNAs.** *Nature.* 2008; **455**(7209): 58–63.

[PubMed Abstract](#) | [Publisher Full Text](#)

Shannon P, Markiel A, Ozier O, *et al.*: **Cytoscape: a software environment for integrated models of biomolecular interaction networks.** *Genome Res.* 2003; **13**(11): 2498–504.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Shimbel A: **Structural parameters of communication networks.** *Bull Math Biophys.* 1953; **15**(4): 501–507.

[Publisher Full Text](#)

Siegert R, Shu HL, Slenczka W, *et al.*: **Zur Ätiologie einer unbekannten, von Affen ausgegangenen menschlichen Infektionskrankheit.** *Dtsch Med Wochenschr.* 1967; **92**(51): 2341–2343.

[Publisher Full Text](#)

Simmons G, Wool-Lewis RJ, Baribaud F, *et al.*: **Ebola virus glycoproteins induce global surface protein down-modulation and loss of cell adherence.** *J Virol.* 2002; **76**(5): 2518–2528.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Stelzl U, Worm U, Lalowski M, *et al.*: **A human protein-protein interaction network: a resource for annotating the proteome.** *Cell.* 2005; **122**(6): 957–968.

[PubMed Abstract](#) | [Publisher Full Text](#)

Ströher U, West E, Bugany H, *et al.*: **Infection and activation of monocytes by Marburg and Ebola viruses.** *J Virol.* 2001; **75**(22): 11025–11033.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Sullivan NJ, Geisbert TW, Geisbert JB, *et al.*: **Accelerated vaccination for Ebola virus haemorrhagic fever in non-human primates.** *Nature.* 2003; **424**(6949): 681–684.

[PubMed Abstract](#) | [Publisher Full Text](#)

Tang J, Costello C, Keet IP, *et al.*: **HLA class I homozygosity accelerates disease progression in human immunodeficiency virus type 1 infection.** *AIDS Res Hum Retroviruses.* 1999; **15**(4): 317–324.

[PubMed Abstract](#) | [Publisher Full Text](#)

Towner JS, Khristova ML, Sealy TK, *et al.*: **Marburgvirus genomics and association with a large hemorrhagic fever outbreak in Angola.** *J Virol.* 2006; **80**(13): 6497–6516.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Towner JS, Sealy TK, Khristova ML, *et al.*: **Newly discovered Ebola virus associated with hemorrhagic fever outbreak in Uganda.** *PLoS Pathog.* 2008; **4**(11): e1000212.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Veillette A, Bookman MA, Horak EM, *et al.*: **The CD4 and CD8 T cell surface antigens are associated with the internal membrane tyrosine-protein kinase p56^{lck}.** *Cell.* 1988; **55**(2): 301–308.

[PubMed Abstract](#) | [Publisher Full Text](#)

Volchkov VE, Becker S, Volchkova VA, *et al.*: **GP mRNA of Ebola virus is edited by the Ebola virus polymerase and by T7 and vaccinia virus polymerases.** *Virology.* 1995; **214**(2): 421–430.

[PubMed Abstract](#) | [Publisher Full Text](#)

Watanabe N, Yamaguchi T, Akimoto Y, *et al.*: **Induction of M-phase arrest and apoptosis after HIV-1 Vpr expression through uncoupling of nuclear and centrosomal cycle in HeLa cells.** *Exp Cell Res.* 2000; **258**(2): 261–269.

[PubMed Abstract](#) | [Publisher Full Text](#)

Watts DJ, Strogatz SH: **Collective dynamics of 'small-world' networks.** *Nature.* 1998; **393**(6684): 440–442.

[PubMed Abstract](#) | [Publisher Full Text](#)

Wilson JA, Hevey M, Bakken R, *et al.*: **Epitopes involved in antibody-mediated protection from Ebola virus.** *Science.* 2000; **287**(5458): 1664–1666.

[PubMed Abstract](#) | [Publisher Full Text](#)

Xiong Y, Eickbush TH: **Origin and evolution of retroelements based upon their reverse transcriptase sequences.** *EMBO J.* 1990; **9**(10): 3353–3362.

[PubMed Abstract](#) | [Free Full Text](#)

Xu L, Sanchez A, Yang Z, *et al.*: **Immunization for Ebola virus infection.** *Nat Med.* 1998; **4**(1): 37–42.

[PubMed Abstract](#) | [Publisher Full Text](#)

Yamashita H, Xu J, Erwin RA, *et al.*: **Differential control of the phosphorylation state of proline-juxtaposed serine residues Ser725 of Stat5a and Ser730 of Stat5b in prolactin-sensitive cells.** *J Biol Chem.* 1998; **273**(46): 30218–30224.

[PubMed Abstract](#) | [Publisher Full Text](#)

Yang Z, Delgado R, Xu L, *et al.*: **Distinct cellular interactions of secreted and transmembrane Ebola virus glycoproteins.** *Science.* 1998; **279**(5353): 1034–1037.

[PubMed Abstract](#) | [Publisher Full Text](#)

Yu-Shiang L, Chun-Yuan L, Che-Lun H, *et al.*: **GPU-UPGMA: high-performance computing for UPGMA algorithm based on graphics processing units.** *Concurrency Computat Pract Exper.* 2014.

[Publisher Full Text](#)

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research