Genome-wide dynamics of emerging viral diseases

Genography and Genomography

A theoretical computational approach to mapping cross-species adaptability for viral tropism

Misaki Wayengera M.D

Editors: Wilson Byarugaba, Joseph Olobo, Kaddu-Mukasa Mulindwa[3], Henry Kajjumbula[4]

1. Professor of Human and Molecular Genetics Division of Pathology-Faculty of Medicine Makerere University P o Box 7072 Kampala Tel: +256772311699 Email: wbyarugaba@yahoo.co.uk .
2. Associate Professor of Immunology and Microbiology Division of Molecular Biology Department of Microbiology. Faculty of Medicine Makerere University P o Box 7072 TeL: +256772613620 Email: jolobo@makmed.ac.ug
3. Senior Lecturer Microbiology Division of Molecular Biology Department of Microbiology. Faculty of Medicine Makerere University P o Box 7072 TeL: +256772457831 Email: dhkmulindwa@med.mak.ac.ug
4. Lecturer-Virology Division of Molecular Biology Department of Microbiology. Faculty of Medicine Makerere University P o Box 7072 Tel:+256772490268 Email: jumic@hotmail.com

Content

Preface (3)

Abbreviation

Introduction(chapter1)

- Definition
- Principle

Methods(Chapter2)

- Design
- Materials, tools and Software

Measured Variables(chapter 3)

Algorithms for Interpreting Genographic Finding(Chapter 4)

E-acceptors: What is there inference?(Chapter5)

A glimpse into the Future of Genography(Chapter6)

References

Examples 1,2 and 3

Supplement; the Evolutionary adaptation cross-species theorem

Preface

Emerging and re-emerging viral infectious diseases (Ewald 1994; Cassell 1994; Ampel 1991) pose one of the greatest challenges to the future of global human health (Hughes and Montagne 1994; Krause 1992; Kaplan 1994; Levins 1994). This is mainly because newly emerging pathogens are more likely to meet a naïve herd immunity, and re-emerging ones a waned one (Ewald 1993). This explains the morbidity and mortality that's often associated with outbreaks of human disease caused by these sorts of pathogens -underling the epidemic and pandemic connotations. Many of the emerging diseases in nature can be classified as zoonotic- being communicable from animals to man (Krause 1992; Kaplan 1994; Morse 1994; Morse and Schluederberg 1990).The control and prevention of many of these diseases-also known as diseases in nature communicable to man (DNCM) has made extremely hard for reasons inclusive of: (1) Poor knowledge and delineation of the exact natural reservoirs hosts; (2) A lack of defined effective vaccines; (3) inadequate resources and response mechanisms at the first point of outbreak; (4) Quick travel and globalization; and many others (McSweegan 1996; Wilson 1995; WHO 1996-2000).

On the part of our currently poor knowledge of the exact natural reservoirs for these pathogens, the rapid passing of epidemics not allowing enough time to study the diseases dynamics (WHO 1996-2000) such as in Ebola (WHO/EPR 2007 ; Bausch 2003) and Marburg(WHO/EPR 2007) is partly to blame. Minly to blame though is the high infectiousness and virulence exhibited by some of these pathogens as well as the diversity of plausible natural host candidates that has delayed the warranted ecological studies as this rendering them unsafe and complex to conduct respectively. A certain minimal levels of infection control and a huge amount of resources are often required by scientist on ground to conduct these "mapping" ecological studies-conditions that have been demonstrated by previous outbreaks not to be readily available until the multilateral collaborative ventures are in place (WHO 1996-2000; WHO/EPR 2007). This book aims to describe a a theoretical in-Silico model for conducting ecological mapping studies across host species for their adaptability to sustain viral tropism called genography. Genography is a computational model that we have developed as an alternative for conducting insilico ecological mapping exercises for the susceptibility of higher organisms to viruses. As genomics and its related down stream sciences increasingly become a vital tool for evaluating evolutionary relationships, it's predictable that we will be able to use the same approach to evaluate the evolutionary adaptability of various hosts to sustain various viral tropisms. This book details for the first time the

design and methodology for conducting such genographic studies. It also details the measurable variables and algorithms for interpreting the results obtained by genography. Herein, we debate as to what may be the proteomic inference of E-acceptors, questioning if they are immunogenetic variants, Predictors of In-vivo viral integration acceptors or residual intron/viral RNA or DNA.

Genography draws its foundation in the fact that viruses are obligate intracellular microbes whose integration is necessary for viral replication and survival. The exact in-vivo acceptors for majority of viral integration are poorly known. Recent studies have however availed evidence that RNA and in particular retroviral(HIV)viral integration is favoured within ALU or LINE elements(Stevens and Griffith, 1994; Stevens and Griffith, 1996; Pryciak et al., 1992; Vijaya et al., 1986).We recently demonstrated that homogeneity between viral LTR and the flanks of host intragenomic intron material is a likely major predictor of viral integration (Misaki, W et al 2007a), further supporting the earlier hypothesis that viral integration follows the possibly same target DNA primed reverse transcription as occur for case in Group II intron insertion(Zimmerly S, Guo H, Eskes R et al 2005). LINE elements (Stevens and Griffith, 1994; Pryciak et al., 1992; Vijaya et al., 1986) are ancestors of Group II introns- a novel class of self splicing and selectively mobile RNAs distributed across all taxa genomes from viruses, prokaryota, archaeal, Eukaryota to vertebrates(inclusive of Humans- where they are existent as spliceosomes or non-long terminal repeats-LTR or Long interspesed Nuclear elements-LINE), where they are to be found interspersing genes within the genomes . The proportion of those interrupted genes is low in yeast and increases in the lower eukaryotes; with fewer genes being uninterrupted in higher eukaryotes (Arkhipova I and Meselson M 2000; Arkhipova I and Meselson M 2005; Michel F and Ferat JL 1995). The lengths of the introns, which can extend from 0.1±100 kb in vertebrates, determine the gene size. Retroposons (also referred to as non-long Terminal Repeats (LTR) or long interspersed nuclear elements-LINE) form the highly mobile class of Group II introns, the transposons. This Group has a unique feature of encoding a reverse transcriptase (RT) open reading flame (ORF) moiety in its IV arm which they use to insert into a predefined sites at high efficacy(retrohoming) or unrelated sites at low rates (retrotransposing)(Arkhipova I and Meselson M 2000). Depending on the function of the protein coded by this moiety, restroposons can be classified as Restriction endonuclases(REase)-like, or apurinic apyramidinic endonuclease(APE)-like (Arkhipova I and Meselson M 2005; Michel F and Ferat JL 1995). The RT ORF has an endonuclease motif within its C terminal Zn^{2+} fingerlike moiety, whose function is like that of all other REases, and is mediated through recognising 6-12 bp nt in the target sequence, and cleaving within, or close to it. While the RNA component of the Ribonucleoprotein often catalyses sense strand cleavage, the antisense strand is cleaved by this DNase(Zimmerly S, Guo

H, Eskes R et al 2005).These specificity sites are to be found within flanking regions of retrohoming introns(also known as exons).

If, as widely believed, RNA viral integration is mediated by the same Target -DNA primed Reverse transcription (TPRT) as in the case for Transposons, then there must exist a certain level of homogeneity between the intra-host genome splicesosomal intron elements and proviral genome to facilitate EBS-IBSI/EBS-IBS2 interactions. It may thus be possible to assess the distribution of this homogeneity in various potential hosts of any virus. Such measurements would enable us to predict which veterinary viruses are potential pathogens to humans; and which veterinary hosts are potential sources of Human viruses. We hypothesized that by directing a section of the virus's gene to the likely hosts, it is possible to measure the prevailing ratio of homogeneity by way of insilico similarity searches employing the BLAST algorithms (Autschul 1991; Autschul 1993; Autschul et al 1994). This proprietarily theoretical computational approach for mapping cross species adaptability to facilitate viral integration and tropism has been termed Comparative Insilico Genography and genomography(Cis-GNG/GNMG), a deviation from true similarity searches of comparative genomics in that it uses a third party viral gene as a "geomic energy" to scan for virus- host intron homogeneity within the potential hosts genome.

Genography may thus be a viable insilico tool for finger printing hosts for historical events (induced by retroelement mobile activity) that may favour RNA viral tropism and integration. To further support this TPRT theory (and cut and past theory for DNA viruses) of viral integration, we note that a number of naturally existing intragenomic antiviral pathways such as the RNAi and cosuppression pathways present in many organisms appear to have evolved primarily for purposes of protecting the host genome against invasion by mobile genetic elements such as transposons and viruses, which produce aberrant RNA or dsRNA in the host cell when they become active (Dalmay,et al 2000; Jensen, et al 1999; Ketting,1999; Ratcliff,et al 1999; Tabara,et al 1999; Malinsky,et al 2000; Anandalakshmi,et al 2000). Specific mRNA degradation prevents transposon and virus replication, although some viruses are able to overcome or prevent this process by expressing proteins that suppress PTGS (Lucy,et al 2000; Voinnet, et al 2000). From the Target DNA Primed Reverse transcription viral integration theory, it can be deduced that genography achieves its primary aim by assessing the homogeneity between host genomic ancestral intron element regions and the virus genome; the main reason the hits to DB obtained are referred to here as evolutionary (E-) acceptors. This theory is referred to as the "acceptor homology theory; in which its hypothesised that acceptors of viral integration on the genomic level are not 3-D conformational lock and key fits as is the case often on the proteomic level, but rather predefined areas of the host genome populated with *viral homologs existent as ancestors of Group II introns* "The theorem postulates that these would

create hot spots for target primed reverse transcription(splicing) as occurs in the case of retrohoming(Misaki, W); a process by which Group II RTE elements insert at high frequency in predefined sites.

In addition to the above "acceptor homology theory" based on target DNA primed reverse transcription as a likely integration model for RNA viruses, we have used 2 other theories to support genograpgy

1. Histoimmunologic acceptance/rejection theory

Selective pressure characterizes almost all host parasites relationship, being more obvious in terms of the parasitic effects on the host. Since viruses integrate intragenomically, such pressure will be exerted on the genetic level, and will be geared at converting the parasitic relationship into a more viable one for both host and viruses (in which case, the viruses will acquire new immune evasion abilities, AND the hosts will acquire viral pathogenesis evasion abilities. It can thus be said that both directions of this pressure work to establish an equilibrium in which the pathogenic relationship is converted into a more symbiotic one. The only way that can occur at the genetic level is when the host acquires the ability to accept the virus as part of its own genetics, and the virus causes no pathology. Just as host graft histoimmunologic compatibility is necessary for allograft or zenograft acceptance, all transposing genetic elements should have the close relationship of the genome into which they integrate to favour integration. This school of thought is explained by the same reasons given above, that across all taxa, there exist evolutionary relatives of retrotransposing elements interspersed within genes. Also, the requirements for TPRT are that retroposon reverse splicing occurs more frequently in predefined sites(retrohoming) than it occurs in any other sites not related to the HNHc or HHVR recognizable nt sequence (retrotransposing)

2.Pathogenesis mediation theory

In Host such as Rhesus and man for filoviruses, or birds/pigs and man for Influenza where both hosts exhibit viral diseases, it may be assumed that mediators of pathogenesis on the proteomic level (antibody antigen interactions) are the same. Since these originate on the genomic level, it can be hypothesised that the antigen-antibody elicitation response abilities across similarly susceptible hosts is the same at the genetic level, reducing gradually in more susceptible to hosts.

E-acceptors, following interpretation by special algorithms can be used to assess the evolutionary adaptability of the studied hosts to favour viral integration.

Abbreviation

BLAST- Basic local alignment sequence tool
Cis- Comparative insilico
DB-Database
Ea-Evolutionary acceptors
EBO-Ebola
GNG/GNMG- Genography/Genomography
HA-hemagglutinase
HGSC-Human Genome Sequencing Center
HSP-High Scoring Pairs
Ia-Inivio Integration acceptor
MBG-Marburg
MIC-Microbial Integration Complex
N,P,X,M-nucleotide, polypeptide, Transcript, metabolite
NA-Neuraminidase
NCBI-National Center for Biotechnology Information
PIC-Pre-Integration Complex
VIC-Viral Integration Complex
VHF-Viral haemorrhagic fevers

Chapter 1

Introduction

Despite the increasing incidence of emerging and re-emerging viral infectious diseases over the past two decades(Hughes and Montagne 1994; Krause 1992; Kaplan 1994; Levins 1994), the exact ecological determinants of viral evolution and emergence are poorly understood. We recently developed the evolutionary adaptation cross-species (EACs) theorem as a succinct summary of the dynamics surrounding viral crossovers (Misaki 2007b see supplement)[1]. The first postulate of this hypothesis holds that since viruses are obligate intracellular organisms, other animals are more likely to be the source of a new viral emergence rather than just spontaneous evolution. This means that all viruses are zoonotic, originating from an animal reservoir; a postulate that would prompt any molecular epidemiologist to ask the question: *What is the spectrum of higher mammals or animals that may be the likely origin of a given virus?* Answering this question is vital in that preventing, or controlling human-host interactions is a vital link in controlling viral emerging infectious diseases.

The only available way of answering the above question has been through ecological mapping exercises. This chapter introduces a novel "trial" theoretical model for conducting such experiments by way of genome wide computational fingerprinting for residual effects of viral tropism- *Genography*

Definition

Genography is a form of genomic scan for mapping the evolutionary adaptability across higher animals to act as sustainable hosts to a given viral; tropism. At the moment, genography is only theoretical, being conducted by way of computational methods employing viral genes as the geomic energy , host genome databases, as well as Software programs developed by Karlin and Altschul as the tools. (Autschul 1991; Autschul 1993)

Principle

Genome wide homology studies aim at establishing similarities between a query from one species(gene or protein, transcript or metabolite) to a genome database of another species by comparing either nucleotide by nucleotide or amino acid by amino acids alignment between the query and database. Because of the wide evolutionary distance between viruses and their usual mammalian hosts, it is highly unlikely that protein similarities are existent. Genography thus uses the nucleotide searches to search for viral homology to intra-genomic ancestral elements of Group II intron within the host that would favour integration in-vivo(Misaki, W"acceptor homology theory) . Much as the basic principle of Genograpgy is to use a similarity search primarily based on same the methods of Karlin and Altschul as the tools (Autschul 1991; Autschul 1993 Autschul et al 1994), genography diverges from the true similarity searches in that it takes advantage of a spectrum of findings that are not truly orthologs (genes that have evolved by vertical descent from same ancestor, so more likely similar both in structure and function) or paralogs (genes that have evolved by recombination, and thus have divergent function).For a consideration of any of the above, which are referred to as homologs, one must have at least 75% homology in sequence alignments, findings that can not be obtained across species even when Retroposon sequence tags(PCR amplifiable 400nt highly conserved sequences adequate to identify a retroposon) are used(Misaki, W). Genography takes advantages of Less than 50 base pairing scores, that though hypothesised to be flanks of introns or spliecosomes, are still poorly delineated; explaining the current nomenclature :evolutionary (E-) acceptors(Historical imprints in the host genome, that are really a residual measure of the consequences of years of retroelement activity or even viral gene induced selective pressure).

The basic principle of Genograpgy draws its foundation in scanning the host genome for these fingerprints called E-acceptors. By comparing the fingerprints for any two given Viral Integration Complexes (VIC1, VIC2) across a given host X for instance, we can roughly establish the ratio of highest(at circulation) possible molecular epidemiological variants of VIC_1 and VIC_2. We can also establish which VIC is best adapted for the tropism of X tissue. Further still, by using a single VIC_1 across hosts X, Y and Z, we can establish the evolutionary adaptability of the studied hosts over time to act as sustainable hosts to the VIC_1.Further details on how

Genography results are interpreted are listed in the next chapters on algorithms (Misaki, W 2007a,b).

Chapter 2

Methodology

- **Design:** Comparative Insilico Genography(*Cis-GNG*)

We detail the methodologies for conducting "fingerprinting of host genomes for E-acceptors of viral integration" by way of genography- a study design we now refer to as *Comparative insilico genography*(because rather than compare just the two host genomes, in which case this would be comparative genomics, we use a third party viral integration complex-*VIC. Note that where a whole virus' genome is used as the VIC, then the design is referred to as *Genomography*(*Cis-GNMG*). Further Still, because we are currently using only computer Software and web based gene and genome databases-the all approach is insilico)

- **Materials, tools and Software:**

The model genographic study is done using[1]nucleotide(nt) Sequences of the VIC gene or whole viral genome if a genomgraphic study is intended[2] **The Blastn Software [3] Host Genome nucleotide database(DB)

*The VIC is often the pathogenesis protein Coding gene of the virus in genography, although in genomography, the entire viral genome is used. For instance, in-vivo pathogenesis of viral haemorrhagic fever in Marburg and Ebola are mediated by the envelope glycoprotein (GP)(Bausch , et al 2006; Towner et al 2006; Mayo & Pringle 1998; Sanchez et al 1993, Khan et al 1997;Dowell et al 1999; Feldmann et al 1993). GP thus forms the model VIC for these two filoviruses. On the other hand, the neuraminidase and haemaglutinin genes form the model VIC for Influenza viruses(Treannor 2004; Vilchez yet al 2002; Nicholson 1998; WHO 2007c; CDC 2007; Russell & Webster 2005; Belshe 2005)

** BLAST (Basic Local Alignment Search Tool) is the heuristic search algorithm employed by the programs blastp, blastn, blastx, tblastn, and tblastx; these programs ascribe significance to their findings using the statistical methods of Karlin and Altschul ((Autschul 1991; Autschul 1993 Autschul et al 1994) with a few enhancements. The BLAST programs were tailored for sequence similarity searching ; for example to identify homologs to a query sequence. The programs are not generally useful for motif-style searching (Autschul 1991; Autschul 1993 Autschul et al 1994). Blastn compares a nucleotide query sequence against a nucleotide sequence database. The Blast tool has four calibration algorithms: Filtration (F), Expect (E), Description (D) and alignment (A). While in a true similarity BLAST search, the program is adjusted to stringent values of filtration(F), Expect(E) , Description(D) and alignment(A), ensuring a high specificity and low sensitivity, in genograpgy-especially because were compare viruses and higher organisms, the values are set a lesser stringency(low specificity, High sensitivity). A True Blast similarity search sets all at stringency values, while in the Cis GNG, because comparing organisms of distant taxa that are likely not to be similar, parameters are set at less stringent values. The values used in these studies can span from: none filtration to default filtration, E= 100 downgrading to 0.01, both A=10 upgrading to 100 and D at 10 to 100

- Interventions:

Genographic scanning or fingerprinting is done by feeding the VIC of the specified virus into the *appropriately calibrated* Blastn tool set to search a host nucleotide genome database for E-acceptors. The E-acceptors are measured by way of proportionate prediction from of measurable variables discussed in the next chapter

Chapter 3

Measured Variables

In a True BLAST similarity search, the two most meaningful measures for homology are sequences producing significant alignment-SPSA(by way of E-value and Score-bits) and High scoring Segment pairs(HSP) (Autschul 1991; Autschul 1993 Autschul et al 1994; Autschul et al 1990). The later, the High-scoring Segment Pair (HSP) is the fundamental unit of BLAST algorithm output. An HSP consists of two sequence fragments of arbitrary but equal length whose alignment is locally maximal and for which the alignment score meets or exceeds a threshold or cutoff score(Autschul 1991; Autschul 1993 Autschul et al 1994; Autschul et al 1990).

While the above two variables are similarly of significance in genography, the large evolutionary distance between the taxa of viruses and higher

mammals will often not allow the generation of many of these. It's thus necessary, as described for calibration of the BLAST tool used in genography, that the stringency of reading the results is also lowered. Thus in the absence of SPSA, and HSPs that are gapped and those not gapped can be used as a predictive measure of the E-acceptors, in which case, one will only concern themselves with the evolutionary distance between the studied hosts rather than the annotating meaning to E-acceptors. Further Still, in absence of any of the above, as often occurs in Genomography(whole virus genome queried against host genome) studies, nucleotide Hits to DB may be used. Thus the descending priorities of the Measured variables for genography studies that are predictive of E-acceptors are:
[1]SPSA-Used to infer what E-acceptors may really regardless of the E-values and Score bits obtained
[2]HSP Gapped successfully
[3]HSP not gapped
[4] nucleotide Hits to DB

These four are all used as a proportionate measure of the distribution of E-acceptors in the studies Species

Chapter 4

Algorithms for interpreting E-acceptors

The algorithms for interpreting Genographic findings were developed using results from Genography and genomograpy(whole H5N1 genome) studies of Influenza A virus Hemagglutinin gene (A/Goose/Guangdong/1/96(H5N1)) hemagglutinin (HA)gene' complete cds) and Neuraminidase (A/Goose/Guangdong/1/96(H5N1)) neuraminidase (NA)gene's complete cds.) across Human, pig and chicken genomes. Influenza viruses provide us with such a model since their virology, genetics and molecular epidemiology has been widely studied, and the only model in which we can apparently claim to thus have appropriate knowledge base to which to compare genographic findings-calibration of

the fingerprints or E-acceptors (NCBI 2007a,b,c &d; Trans-NIH Gallus initiative 2007).

In This Study, the numbers of HSP's gapped (taken to be directly proportional to the number of integration acceptors in-vivo) produced by BLASTN homology survey in humans, pigs and chicken for HA were 62 in humans; 12 in pigs and 306 in chicken compared to 25 in humans; 3 in pigs and 190 in chicken produced by Influenza A neuraminidase (NA) gene. These figures corresponded to the number of successful extensions for each query. There were no HSP's successfully gapped, HSP's better than 0.01 without gapping, or sequences better than 0.01 KARLIN-ALTSCHUL Statistical Constant values were Lambda= 1.37, H= 0.711, and K= 0.31(for Details of the Studied, see Misaki, W EID 2007, Misaki, W Afr J Biotech)

Table 3 Ratio of HA and NA across the 3 Studies Species

Human build 36 genome	Pig genome build	Gallus gallus (chicken) build 2 genome
62/25	12/3	306/190

Note the relationship between the distribution of E-acceptors to the molecular currently documented epidemiological distribution of antigenically distinct hemagglutinin (HA) and neuraminidase (NA) types (15/9) across chicken and man

Our results served to show, that as expected from the distant evolutionary distance between viruses and the 3 hosts studied(Pigs, chicken and Man), there were completely no homologs(paralogs or orthologs). This makes these findings, when viewed from the perspective of the true BLAST similarity searches, thus useless. However, on the part of accessing the evolutionary adaptability of the studied hosts, this study has a lot of significance. For instance, here we demonstrate that the ratio of E-acceptors for HA and NA across the three studies species is directly proportionate to the current molecular epidemiological data on currently circulating antigenically distinct hemagglutinin (HA) and neuraminidase (NA) types (15/9) (Treannor 2004; Vilchez yet al 2002; Nicholson 1998; WHO 2007c; CDC 2007; Russell & Webster 2005; Belshe 2005) . This ratio is also approximately maintained across all influenza A hosts analyzed (62/25 for Homo sapiens; 12/3 for Sus scrofa and 306/190 for gallus gallus); although the relationship is more obvious in humans and chicken (***see tables 1, 2 and 3).*** This lead us to our first equation and algorithm for interpreting Cis-GNG results:

For any two VICs(VIC1;VIC2) studied in species X; the ratio $X.E_{a(Vic1)}$: $X.E_{a(Vic2)}$ approximates the molecular epidemiological distribution of the current circulation forms of VIC1 and VIC2. …… (E-Algorithm1)

From equation 1, and the fact that this ratio is maintained across all the 3 studies hosts, it may mean that that E-acceptors (E_a) may actually be proportional to In-vivo integration acceptors (I_a). It means that this ratio

shows the distributions of the selective pressure that each gene has had on the hosts over years of tropism, and since viruses are obligate intracellular microbes, this ratio shows in a way, the ratio of the naturally existing molecular distribution of these genes (VIC). This brings us to our second algorithm or equation for Cis-GNG

In any particular species; $E_a \sim I_a$ and $E_a = \lambda I_a$; where λ is a proportionality constant defining the relationship between Evolutionary acceptors (E_a) and in-vivo Integration acceptors (I_a). …… (E-Algorithm2)

Basing on the above E-algorithms (1 and 2), it means that Chicken have more In-vivo Integration acceptors for both HA and NA than man and pigs. If we unquestioningly held the hypothesis that a natural reservoir should be one that is most well adapted for viral integration and tropism, and that adaptation is proportional to the number of In-vivo Integration acceptors (I_a) innately expressed; then It would mean that birds (chicken) may thus be more closely related to the natural reservoir of Influenza viruses than man and pigs-basing on the hypothesis that the natural reservoir is. This brings us to our third equation or algorithm (E- Algorithm3)

For any queried viral integration complex (VIC) across host species X, Y and Z; if $X.E_a > Y.E_a > Z.E_a$; then X is more likely to be related to the virus' natural reservoir (NR) than Y and Z. ……..(E-Algorithm3)

However, considering the acceptor homology theory used here seems applicable to only RNA viruses, and the exact relationship between In-vivo acceptors for DNA viruses and evolutionary (E-)acceptors is yet to be defined; equation 2 ($E_a \sim I_a$)is debatable as that the reverse may equally questionable hold, that the natural reservoir is one that has the least number I_a. The later thesis is emphasized by the fact that a natural reservoir does not exhibit disease, and yet both the VICs used in this study mediate diseases. If I_a were mediators of disease as from the pathogenesis mediator homology theory, then inverse relationship would hold($E_a \sim 1/I_a$. Holding onto the later, it may thus mean that pigs are more closely related to Influenza NR in the evolutionary tree, since pigs from this study have fewer E_a(and thus I_a) acceptors, and the NR is that organism in nature that although habors the virus, does not exhibit associated disease. This question can only be resolved on research geared at establishing the actual in-vivo integrations acceptors, and comparing their prevalence across the hosts studied.

However, despite the fact that our discussion here limits the interpretation of E-acceptor to In-vivo integration acceptors, a broader consideration of the implications of E-acceptors is discussed in the Next chapter.

Chapter 5

E-acceptors: What is there inference?

From the second algorithm above, we limit the implication of E-acceptors to possibly in-vivo Integration acceptors. However, as we describe from the preference, E-acceptors; inline with the "acceptor homology theory"

are an indiscrete measure of evolutionary adaptation of the host genome in light of historical retro homing activities. However, from the other theories described in the preface, one may argue that E-acceptors are immunogenetic variants (IGvar) or residual intron or viral RNA and DNA.

In a preliminary study that aimed to use the University of California genome browser to infer the meaning of these E-acceptors from Ebola and Marburg genrographic studies in Human and NHP(Rhesus macaca), a broad of range of enzymatic peptides, tumour mediating promoter and inhibitor genes previously not described in the mediation of viral haemorrhagic fever were found to be related to this residual effect(Misaki, W Afr J Biotech). That was expected as 25-35% of entire human genome is comprised of spliceosomes interspersing such genes. Much as it may seem that the only way to answer this will be to conduct in-vitro and in-vivo experiments to map the exact meaning of E-acceptors. Genome browsers are complex insilico search Software that combine information on the genomics of a host, and it's related down stream sciences inclusive of proteomics, transcrptomics, and metabolomics. In a way, these programs and softwares bring us very close to in-vivo experiments, but none the less, can not replace in-vivo gene expression experiments.

This takes us to our one of the 3 the future prospects for genography discussed in the next chapter, that of conducting genography on DNA chips/microarrays and interpreting the results by bioinformatics software

For now, E-acceptors may be inferred as:
[1]A proportionate measure of the Distribution of In-vivo Intergration acceptors (or mediators of disease in-vivo)
[2]Immunogenetic variants
[3] Residual intron or viral RNA/DNA or intron RNA/DNA

Chapter 6

A glimpse into the Future of Genography

In the previous chapter, the question as to what is the exact inference of E-acceptors remains widely unanswered, despite their value in genography to offers us an indiscrete measure of the evolutionary distance across species with respective to a given VIC. It's the need to answer this question that will in a way determine what the future of genography will be. At this point, we predict that:

[1]Within the Insilico model, genopgraphy will intergrate other down stream sciences by the use of genome browsers. We predict that this integration of related downstream sciences (transcriptomics, proteomics and metabolomics) will not only broaden applications of Cis-GNG/GNMG to include other microbial forms of life(: Archaea, Bacteria, Eukaryotae, Viruses, Viroids, and Plasmids); but also enable us to make discrete descriptions of the in-vivo acceptors of pathogen integration into hosts-in other words; the mediators of diseases pathogenesis. This will become a sort of molecular fingerprinting on the genomic level
For antigen-antibody interactions(in the immune disease model)
[2]Beyond the in-silco model, genography will move to the be conducted on DNA micro array chips and interpreted by way on Bioinformatic software.

References

- Altschul, Stephen F., Warren Gish, Webb Miller, Eugene W.Myers, and David J. Lipman .1990. Basic local alignment search tool. J. Mol. Biol. 215:403-10.
- Altschul, SF. 1991. Amino acid substitution matrices from an information theoretic perspective. J. Mol. Biol. 219:555-65.
- Altschul, S. F. 1993. A protein alignment scoring system sensitive at all evolutionary distances. J. Mol. Evol. 36:290-300.
- Altschul, S. F., M. S. Boguski, W. Gish and J. C. 1994. Wootton Issues in searching molecular sequence databases. Nature Genetics; 6:119-129.
- Ampel NM. 1991. Plagues--What's past is present: thoughts on the origin and history of new infectious diseases. Rev Infect Dis 13:658-665.
- Anandalakshmi, R., Marathe, R., Ge, X., Herr, J.M., Jr., Mau, C., Mallory, A., Pruss, G., Bowman, L., and Vance, V.B. 2000. A calmodulin-related protein that suppresses posttranscriptional gene silencing in plants. *Science* 290: 142-144
- Baize, S., Leroy, E. M., Georges-Courbot, M.-C., Capron, M., Lansoud-Soukate, J., Debré, P., Fisher-Hoch, S. P., McCormick, J. B. & Georges, A. J. 1999. Defective humoral responses and extensive intravascular apoptosis are associated with fatal outcome in Ebola virus-infected patients. *Nature Medicine.* 5: 423-426.
- Bausch, D. G. , Nichol, S. T. , Muyembe-Tamfum, J. J., Tugume, B., Colebunders,R KondÃ, K.M.,et al.. 2006. Marburg hemorrhagic fever associated with multiple genetic lineages of virus. N. Engl. J. Med., August 31; 355(9): 909 - 919.
- Bausch DG, Borchert M, Grein T, Roth C, Swanepoel R, Libande ML, et al. Risk factors for Marburg hemorrhagic fever, Democratic Republic of the Congo. Emerg Infect Dis [serial online] 2003 Dec [3rd AUG 2007]. Available at http://www.cdc.gov/ncidod/EID/vol9no12/03-0355.htm
- Belshe RB. 2005. The origins of pandemic influenza—Lessons from the 1918 virus. *N Engl J Med*; 353: 2209-2211.
- Bor, Y.-C., F. Bushman, and L. Orgel. In vitro integration of human immunodeficiency virus type 1 cDNA into targets containing protein-induced bends. Proc. Natl. Acad. Sci. USA 1995; 92:10334-10338
- Bowen, E. T., Platt, G. S., Lloyd, G., Raymond, R. T. & Simpson, D. I. 1980. A comparative study of strains of Ebola virus isolated from southern Sudan and northern Zaire in 1976. *Journal of Medical Virology.* 6: 129-138.
- Bushman, F. D. Tethering human immunodeficiency virus 1 integrase to a DNA site directs integration to nearby sequences. Proc. Natl. Acad. Sci. USA 1994;91:9233-9237
- Bushman, F. D., and R. Craigie. Integration of human immunodeficiency virus DNA: adduct interference analysis of required DNA sites. Proc. Natl. Acad. Sci. USA 1992; 89:3458-3462

- Carteau S, Hoffmann C, Bushman F. Chromosome structure and human immunodeficiency virus type 1 cDNA integration: centromeric alphoid repeats are a disfavored target. J Virol. 1998;72(5):4005-14.
- Cassell GH. 1994. New and emerging infections in the face of a funding crisis. ASM News 60:251-254.
- CDC. Key facts about avian Influenza and Human influenza virus A. CDC Facts sheets http://www.cdc.gov/flu/avian/gen-info/facts.htm
- Dalmay, T., Hamilton, A., Rudd, S., Angell, S., and Baulcombe, D.C. 2000. An RNA-dependent RNA polymerase gene in *Arabidopsis* is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. *Cell* 101: 543-553
- Dowell, S. F., Mukunu, R., Ksiazek, T. G., Khan, A. S., Rollin, P. E. & Peters, C. J. 1995. Transmission of Ebola hemorrhagic fever: a study of risk factors in family members, Kikwit, Democratic Republic of the Congo. *Journal of Infectious Diseases.*1999;179 (Suppl. 1): S87–S91
- Ewald PW. *The Evolution of Infectious Disease*. Oxford University Press, 1994.
- Ewald PW. 1993. The evolution of virulence. Scientific American. April: 86-93.
- Feldmann, H., Klenk, H. D., & Sanchez, A. 1993. Molecular biology and evolution of filoviruses. *Archives of Virology.* 7: 81-100.
- Felsenstein, J. 1993. PHYLIP: Phylogeny Inference Package; version 3.5c. University of Washington, Seattle, WA, USA.
- Howard, M. T., and J. D. Griffith. A cluster of strong topoisomerase II cleavage sites is located near an integrated human immunodeficiency virus. J. Mol. Biol. 1993;232:1060-1068
- Hughes JM, La Montagne JR. 1994. The challenges posed by emerging infectious diseases. ASM News 60:248-250.
- Jensen, S., Gassama, M.P., and Heidmann, T. 1999. Taming of transposable elements by homology-dependent gene silencing. *Nat. Genet.* 21: 209-212
- Kaplan RD. 1994. The coming anarchy. The Atlantic Monthly. Feb: pp 44-76.
- Ketting, R.F., Haverkamp, T.H., van Luenen, H.G., and Plasterk, R.H. 1999. Mut-7 of *C. elegans*, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. *Cell* 99: 133-141
- Khan, A. S., Tshioko, F. K., Heymann, D. L., Le Guenno, B. Nabeth p. ; Kerstiens b., et al. 1993. The reemergence of Ebola hemorrhagic fever, Democratic Republic of the Congo, 1995. *Journal of Infectious Diseases.* 179 (Suppl. 1):S76–S86.
- Krause RM. 1992. The origin of plagues: old and new. Science 257:1073-1078..
- Leroy, E. M., Baize, S., Volchkov, V. E., Fisher-Hoch, S. P., Georges-Courbot, M.-C., Lansoud-Soukate, J., Capron, M., Debré, P.,

McCormick, J. B. & Georges, A. J. 2000. Human asymptomatic Ebola infection and strong inflammatory response. *Lancet.* 355: 2210-2215.

- Levins R, Members of the Harvard Working Group on New and Resurgent Diseases. 1994. The emergence of new diseases. American Scientist 82:52-60.
- Lucy, A.P., Guo, H.S., Li, W.X., and Ding, S.W. 2000. Suppression of post-transcriptional gene silencing by a plant viral protein localized in the nucleus. *EMBO J.* 19: 1672-1680
- Malinsky, S., Bucheton, A., and Busseau, I. 2000. New insights on homology-dependent silencing of I factor activity by transgenes containing ORF1 in *Drosophila melanogaster. Genetics* 156: 1147-1155
- Mayo, M. A. & Pringle, C. R. 1998. Virus taxonomy - 1997. *Journal of General Virology.* 79: 649-657.
- McSweegan E. 1996. The infectious disease impact statement: a mechanism for addressing emerging infectious diseases. Emg Infect Dis 2:103-108.
- Milot, E., A. Belmaaza, E. Rassart, and P. Chartrand. 1994. Association of a host DNA structure with retroviral integration sites in chromosomal DNA. Virology 201:408-412
- Misaki, W A model for mapping of Ebola and Marburg RNA integration sites in rhesus *Macaca mulatta* genome in silico: Ebola virus acceptor sites probably located on chromosomes 4, 6, 7, 8, 9, 14 and 15 AJB-07-557 Accepted 31st OCT 2007
- Misaki, W The evolutionary adaptation cross species theorem Hypothesis Journal- No MS No. Underpress 2007.
- Moalic Y, Blanchard Y, Félix H, Jestin A. 2006. Porcine endogenous retrovirus integration sites in the human genome: features in common with those of murine leukemia virus. J Virol. . 80(22):10980-10988.
- Morse SS. 1994. Crossing over: the interspecies traffic of emerging infections. J NIH Res 6(10):52-56.
- Morse SS, Schluederberg A. 1990. Emerging viruses: the evolution of viruses and viral diseases. J Infect Dis 162:1-7.
- Muller, H.-P., and H. E. Varmus. 1994. DNA bending creates favored sites for retroviral integration: an explanation for preferred insertion sites in nucleosomes. EMBO J. 13:4704-4714
-
- Nicholson KG. 1998. Human influenza. In: Nicholson KG, Webster RG, Hay AJ, eds. *Textbook of Influenza*. Oxford, England: Blackwell Science,: 219-264
- Peterson, A.T., Carroll, D.S., Mills, J.N., Johnson, K.M. 2004. Potential of mammalian filovirus reservoirs. EID 10(2):2073-2081
- Pruss, D., F. D. Bushman, and A. P. Wolffe. 1994. Human immunodeficiency virus integrase directs integration to sites of severe DNA distortion within the nucleosome core. Proc. Natl. Acad. Sci. USA 91:5913-5917

- Pruss, D., R. Reeves, F. D. Bushman, and A. P. Wolffe. 1994. The influence of DNA and nucleosome structure on integration events directed by HIV integrase. J. Biol. Chem. 269:25031-25041
- Pryciak, P., H.-P. Muller, and H. E. Varmus. Simian virus 40 minichromosomes as targets for retroviral integration in vivo. Proc. Natl. Acad. Sci. USA 1992; 89:9237-9241
- Pryciak, P. M., A. Sil, and H. E. Varmus. Retroviral integration into minichromosomes in vitro. EMBO J. 1992; 11:291-303
- Ratcliff, F.G., MacFarlane, S.A., and Baulcombe, D.C. 1999. Gene silencing without DNA: RNA-mediated cross-protection between viruses. *Plant Cell* 11: 1207-1216
- Russell CJ, Webster RG. 2005. The genesis of a pandemic influenza virus. *Cell*; 123: 368–371
- Sanchez, A., Kiley, M. P., Holloway, B. P. & Auperin, D. D. 1993. Sequence analysis of the Ebola virus genome: organization, genetic elements, and comparison with the genome of Marburg virus. *Virus Research.* 29: 215-240
- Scherdin, U., K. Rhodes, and M. Breindl. 1990. Transcriptionally active genome regions are preferred targets for retrovirus integration. J. Virol. 64:907-912
- Scottoline, B. P., S. Chow, V. Ellison, and P. O. Brown. 1997. Disruption of the terminal base pairs of retroviral DNA during integration. Genes Dev. 11:371-382
- Sels, F. T., S. Langer, A. S. Schulz, J. Silver, M. Sitbon, and R. W. Friedrich. 1992. Friend murine leukaemia virus is integrated at a common site in most primary spleen tumours of erythroleukaemic animals. Oncogene 7:643-652
- Stevens, S. W., and J. D. Griffith. 1994. Human immunodeficiency virus type 1 may preferentially integrate into chromatin occupied by L1Hs repetitive elements. Proc. Natl. Acad. Sci. USA 91:5557-5561
- Stevens, S. W., and J. D. Griffith. 1996. Sequence analysis of the human DNA flanking sites of human immunodeficiency virus type 1 integration. J. Virol. 70:6459-6462
- Tabara, H., Sarkissian, M., Kelly, W.G., Fleenor, J., Grishok, A., Timmons, L., Fire, A., and Mello, C.C. 1999. The rde-1 gene, RNA interference, and transposon silencing in *C. elegans. Cell* 99: 123-132
- Towner, J. S. , Khristova, M. L., Sealy, T. K. , Erickson, B. R. , Bawiec, D., Hartman, A. L. ,et al.. 2006. Marburg virus genomics and association with a large hemorrhagic Fever outbreak in angola. J. Virol. 80(13): 6497 - 6516.
- Treanor JJ. 2004. Influenza virus. In Mandell GL, Bennett JE, Dolin R, eds. *Principles and Practice of Infectious Diseases*, 6th Ed. New York: Churchill Livingstone Inc.,: 2060–2085.
- Varmus, H. E., and Brown, P. O. 1989. Retroviruses, p. 53-108. *In* D. E. Berg, and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.

- Vilchez RA, Fung J, Kusne S. 2002. The pathogenesis and management of influenza virus infection in organ transplant recipients. *Transpl Infect Dis*; 4: 177–182.
- Vijaya, S., D. L. Steffan, and H. L. Robinson. Acceptor sites for retroviral integrations map near DNase I–hypersensitive sites in chromatin. J. Virol. 1986; 60:683–692
- Voinnet, O., Lederer, C., and Baulcombe, D.C. 2000. A viral movement protein prevents spread of the gene silencing signal in *Nicotiana benthamiana. Cell* 103: 157–167
- Watanabe, S., Takada, A., Watanabe, T., Ito, H., Kida, H. & Kawaoka, Y. 2000. Functional Importance of the coiled-coil of the Ebola virus glycoprotein. *Journal of Virology.* 74: 10194–10201.
- Wilson M. 1995. Travel and the emergence of infectious diseases. Emg Infect Dis 1(2):39–46.
- Withers-Ward, E. S., Y. Kitamura, J. P. Barnes, and J. M. Coffin. 1994. Distribution of targets for avian retrovirus DNA integration in vivo. Genes Dev. 8:1473–1487
- World Health Organization. Strategic Plan Outline 1996–2000. Emerging and Other Communicable Diseases Surveillance and Control, WHO EMC/96.1.
- WHO.2007a. Epidemic and pandemic alert and response (EPR). Ebola haemorrhagic fever http://www.who.int/csr/disease/ebola/en/
- WHO. 2007b. Epidemic and pandemic alert and response(EPR). Marburg haemorrhagic fever http://www.who.int/csr/disease/marburg/en/index.html
- WHO. 2007c. Influenza facts sheets. http://www.who.int/mediacentre/factsheets/fs211/en/index.html
- Yang, Z.-Y., Delgado, R., Xu, L., Todd, R. F., Nabel, E. G., Sanchez, A. & Nabel, G. J. 1998. Distinct cellular interactions of secreted and transmembrane Ebola virus glycoproteins. *Science.* 279: 1034–1037.

Examples 1

Comparative in-silico genography and genomography as in-silico tools for predicting the adaptability across species to act as suitable hosts for zoonotic pathogens

Wayengera Misaki

1. Makerere University faculty of Medicine (wmisaki@yahoo.com

Zoonotic diseases-or rather diseases in nature communicable to man are increasingly becoming a noticeable threat to the future of global human health. Several of these diseases have emerged from yet to be defined hosts-and the dynamics surrounding their crossover are currently poorly understood-mainly due to the high infectiousness of the associated microbes and the complexity of the warranted ecological studies. Genomics and its related downstream sciences offer us a great tool for predicting species adaptability to act as sustainable hosts to a zoonotic microbe in-Silico. Here, we describe Comparative In-silico genography and genomography(Cis-GNG/GNMG)as novel form of geomic "ultrasound" for exploring the consistency across host species' genomes in relation to a defined microbial integration complex (MIC): a tool for of generating insights regarding the evolutionary adaptability of the studied species to act as sustainable hosts to the microbe whose MIC was used in-Silico.

Key Words: Emerging pathogens; re-emerging microbes; host adaptability; Microbial integration complexes, genomics, evolution;, cross species, zoonoses, xenoses. **Abbreviations:** MIC-microbial integration complex; PIC-pre-integration complex; DB-database; HSP-High scoring pair; Cis-GNG-Comparative In-Silico genography; Cis-GNMG-Comparative Insilico genomography, DNCM-Diseases in nature communicable to man.

Introduction

Emerging and re-emerging infectious diseases[1, 2, 3] pose one of the greatest challenges to the future of global human health[4-7]-mainly because newly emerging pathogens are bound to meet a naïve herd immunity, while re-emerging ones a waned

one[8]. Consequently, the morbidity and mortality associated with outbreaks of human disease caused by these sorts of pathogens are in many cases extremely maximal-underling the epidemic and pandemic connotations. Many of the emerging diseases in nature can be classified as zoonotic- being communicable from animals to man[9-12]. The control and prevention of many of these diseases-also known as diseases in nature communicable to man (DNCM) is made extremely hard for reasons inclusive of: (1) Poor knowledge and delineation of the exact natural reservoirs hosts; (2) A lack of defined effective vaccines; (3) inadequate resources and response mechanisms at the first point of outbreak; (4) Quick travel and globalization; and many others[1-12, 13, 14].
On the part of our currently poor knowledge of the exact natural reservoirs for these pathogens, the rapid passing of epidemics not allowing enough time to study the diseases dynamics[15] such as in Ebola[16, 17] and Marburg[18] is partly to blame. But mainly, it is the high infectiousness and virulence exhibited by some of these pathogens as well as the diversity of plausible natural host candidates that has delayed theoretically possible and warranted ecological studies by rendering them unsafe and complex respectively
.

Thus certain levels of infection control and a huge amount of resources must be available to the scientist on ground to conduct these "mapping" ecological studies-something that has not been readily available until the multilateral collaborative ventures are in place[1-18]. We aimed to develop an in-Silico model for conducting ecological mapping studies across host species.

PRINCIPLE AND HYPOTHESIS:
The combination of genomics, genomics related downstream sciences (transcriptomics, proteomics and metabolomics)[19-28] as well as theoretical sciences employing computational modeling strategies[19-41] offers us an avenue to conduct in-Silico genome wide scale "mapping" ecological exercises regarding cross-species host suitability to act as a sustainable hosts to a query microbe[42]. In light of the completion of the sequencing of the human genome[43,44], several related primate and non primate genomes[45-47], as well as a spectrum of microbial genomes[48-53], we hypothesized that by directing a defined microbial gene, protein, transcript or metabolite or even whole genome [aka model integration complex-MIC)][54,55] as "*a geomic energy of standard wavelength*" towards any two or more study host Databases(DB) , and analyzing the resulting levels of absorbance, transmittance, refraction ,reflection and deflection in terms of hits to DB, HSPs, and significant alignments produced[20-41]; we may generate indiscrete insights on the current adaptability of the queried host species to act as sustainable host to a query pathogen. It is this novel concept we have termed Comparative In-Silico genography(*Cis*-GNG) or genomography(*Cis*-GNMG). Viruses, being obligate intracellular microbes, and the lowest forms of life necessitating genomic integration into hosts for survival form the basic model for genography that is purely based on genomics and the Basic local alignment sequence tool-

nucleotide(blastn) homology tools, although, we predict that the integration of related downstream sciences (transcriptomics, proteomics and metabolomics) will not only broaden applications of Cis-GNG/GNMG to include other microbial forms of life(: Archaea, Bacteria, Eukaryotae, Viruses, Viroids, and Plasmids); but also enable us to make discrete descriptions of the in-vivo acceptors of pathogen integration into hosts- in other words; the mediators of diseases pathogenesis. Because we are describing some concepts on a nano-scale level, that is atomic or molecular levels; these two concepts may also be labeled as nano-genography(NGNG) or nano-genomograpgy(NGNMG). Further still because the approach being explored is based purely on the use of computer and web based tools, that's to say in-silico, these concepts may be further refined as in-silico nano-genography(ISNGNG) or in-silico nano-genomograpgy(ISNGNMG). The comparative term is coined because more than one species is studied.

RATIONALE AND JUSTIFICATION: EXISTING SUPPORT FOR THE PRINCIPLE OF GNG AND GNMG
Nature-echolation and hearing in bats; Echolocation is a type of sonar that microbats use to detect prey, locate roosting crevices and avoid close obstacles in the dark. These bats emit a very loud and short 'shout' of sound and listen for the echo that bounces back when it hits an object. Put simply, this can tell the bats how far away the object is by estimating the time it takes before they hear the echo. The longer it takes the echo to return, the further away the object[56-58].A section of nocturnal butterflies also employ this mechanism[59]. At this point, we may regard ourselves to be to some extent blind in regard to our current limited knowledge on the evolutionary and innate adaptability of various species to act as sustainable hosts for emerging pathogens. What we are trying to do is direct "a constant wavelength of a microbial genome (energy)" towards a spectrum of possible host species; and rely on the echo to make deductions as to which may best act as a suitable host. Other sciences exploring a related principle are: sonography for ship navigation[60-62], and ultrasonography for radiological imaging tool in medicine[63]

current areas of applications for *cis*-gng and *cis*-gnmg
Comparative In-Silico genography and genomography are currently applicable for the following purposes:
1. Cross species mapping for adaptability as hosts to a specified emerging pathogen or microbe (see current examples)
2. Prediction of which species is most likely to be natural reservoirs to a specified emerging pathogen (see current examples)
3. Both 1 and 2 can be used in pre-cross species tissue and organ transplantation and regenerative medicine to predict and prevent xenoses

Materials and tools required to conduct comparative in-silico genographic and genomographic studies

1. A defined MIC or query: The query or MIC may be a specified pathogen's gene, an attachment protein or whole microbial genome

2. Host database (nucleotide, protein, transcriptic, metabolic): The basic DB for of the current model described are based on nucleotide databases of the host organisms

3. Appropriate computational software: In the case of the basic genography based on nucleotide searches, the tool is the Basic local Alignment search tool-BLAST based on nucleotide search (BLASTN). **BLAST (Basic Local Alignment Search Tool) is the heuristic search algorithm employed by the programs blastp, blastn, blastx, tblastn, and tblastx**; these programs ascribe significance to their findings using the statistical methods of Karlin and Altschul (1990, 1993) with a few enhancements.The BLAST programs were tailored for sequence similarity searching -- for example to identify homologs to a query sequence. The programs are not generally useful for motif-style searching[20-41]. **BLASTN** compares a nucleotide query sequence against a nucleotide sequence database.

Current examples of studies exploiting the concept of comparative in-silico genography and genomography

1. Genome wide homology survey for influenza A acceptors in Homo sapiens(humans), Sus scrofa(pigs) and Gallus gallus(chicken): Gallus gallus(chicken) originating Influenza A strains better adapted to Homo sapiens(humans) than Sus scrofa(pigs). *[Influenza A, the subtype responsible for large human pandemics, infects a spectrum of species ranging from humans, swine, equine, avian, to marine mammals. In the past, pigs have been denoted as mixing vessels for this family of viruses. We conducted a genome wide Cis-GNG survey for influenza A acceptors in Homo sapiens (humans), Sus scrofa(pigs), and Gallus gallus(chicken)to trace the evolutionary trends acceptors of influenza A among these three hosts by querying both HA and NA against whole genomes of Homo sapiens (humans), Sus scrofa(pigs), and Gallus gallus(chicken). Of all 2.85 Gb nucleotides in the human genome, 2.7Gb in Sus scrofa, and 1.2 Gb in Gallus gallus: HA hits to DB were 6,351,123; 347,716 and 2,723,004 compared to 4,868,784; 283,394 and 2,088,939 hits to DB for NA. Numbers of HSP's gapped (paralleled for integration acceptors in-vivo) were 62; 12 and 306 for HA respectively compared to 25; 3 and 190 for NA in humans, pigs and chicken respectively. Avian derived influenza strains seem to be more innately adapted for integration in humans than pigs, making the need for porcine as mixing vessels obsolete, meaning that, viral recombination and mutations necessary for bird-to-human crossovers occur more readily within birds rather than an intermediary porcine host]*[64].

2. A Comparative genomics survey of acceptors for filoviridae viruses in the Rhesus macaque (macaca mulatta) and Homo sapiens genomes: Filoviruses are still not well adapted to establish sustainable integration within the humans. *[Viral integration into the host genetic material is necessary for replication and survival, since viruses are obligate intracellular organisms. We hypothesized that homogeneity between viral integration complex and host genome may be a major predictor of integration. To investigate and map the loci of integration of the two major geneses of this family of viruses within NHP and Human genomes, we queried both Ebola and Marburg whole genomes and Glycoprotein (GP) gene sequences against the whole genomes of rhesus macaque and the Homo sapiens using BLAST-N analysis. Whole filoviridae genomes had no compatible acceptors in both rhesus and human genomes. Of all the contigs length 2.87 Gb (2,863,665,185) bases in the genome of rhesus macaque, Marburg GP blast hits to DB were 6,451,736 compared to 4,012,901 for Ebola; while, of all 2.88 Gb(2,851,330,913) nucleotides in the human genome, Marburg GP blast hits were 1,013,987 compared to 1,013,492 for Ebola. Marburg GP genomic RNA had 18(66.6%efficiency) alignments located on undefined scaffolds compared to 7(100%efficiency) of Ebola located on chromosomes 4, 6, 7, 8, 9, 14 & 15 within the rhesus macaca where but, within the human genome Marburg GP genomic RNA had 8(62.5%efficiency) alignments located on undefined scaffolds compared to 3(100% efficiency) of Ebola. Our results serve to demonstrate that although Marburg GP RNA acceptors are more prevalent in the Rhesus genome than ebola; their loci of integration are vaguely defined compared to Ebola. If*

the level of homogeneity between acceptors and PIC has no effect of integration, then Marburg may be better adapted to integrate into Rhesus that Ebola. Alternatively, chromatic DNA might be a more effective target for future Ebola genomic vaccines sequestered at a nuclear location inaccessible to incoming Pre-integration Complexes (PICs-which in this model are Ebola glycoprotein gene complexes) than Marburg.]

THE FUTURE OF THE CONCEPT OF COMPARATIVE IN-SILICO GENOGRAPHY AND GENOMOGRAPHY

The exact host integration acceptor sites for many viruses are still ill defined, most not expected to be present as naked DNA but rather associated with histones and other DNA-binding proteins in chromatin. DNA packaging in vivo is expected to influence integration site selection, and the choice of integration site may have profound effects on the virus[54, 55]. Fine tuning of genography by integrating other related genomics downstream sciences such as proteomics, trascriptomics and metabolomics may thus enable the identification and description of exact acceptors of MIC in-silico without the need for further in-vitro and –vivo studies.

Conclusion

We present here for the first time ever, Comparative In-Silico genography and genomography ((Cis-GNG/GNMG) as a quick, easy to use, and safe in-silico approach for "mapping" cross species adaptability to act as suitable hosts for emerging and re-emerging pathogens classifiable as diseases in nature communicable to man (DNCM) or zoonoses

Funding: The author received no specific funding towards this work

References

1. Ewald PW. *The Evolution of Infectious Disease*. Oxford University Press, 1994.
2. Cassell GH. New and emerging infections in the face of a funding crisis. ASM News 1994; 60:251-254.
3. Ampel NM. Plagues--What's past is present: thoughts on the origin and history of new infectious diseases. Rev Infect Dis 1991; 13:658-665.
4. Hughes JM, La Montagne JR. The challenges posed by emerging infectious diseases. ASM News 1994; 60:248-250.
5. Krause RM. The origin of plagues: old and new. Science 1992; 257:1073-1078..
6. Kaplan RD. The coming anarchy. The Atlantic Monthly, 1994; Feb: pp 44-76.
7. Levins R, Members of the Harvard Working Group on New and Resurgent Diseases. The emergence of new diseases. American Scientist 1994; 82:52-60.
8. Ewald PW. The evolution of virulence. Scientific American, 1993; April: 86-93.
9. Kaplan RD. The coming anarchy. The Atlantic Monthly, 1994; Feb: 44-76.
10. Krause RM. The origin of plagues: old and new. Science 1992;257:1073-1078
11. Morse SS. Crossing over: the interspecies traffic of emerging infections. J NIH Res 1994;6(10):52-56.
12. Morse SS, Schluederberg A. Emerging viruses: the evolution of viruses and viral diseases. J Infect Dis 1990; 162:1-7.
13. McSweegan E. The infectious disease impact statement: a mechanism for addressing emerging infectious diseases. Emg Infect Dis 1996; 2:103-108.
14. Wilson M. Travel and the emergence of infectious diseases. Emg Infect Dis 1995; 1(2):39-46.
15. World Health Organization. Strategic Plan Outline 1996-2000. Emerging and Other

Communicable Diseases Surveillance and Control, WHO EMC/96.1.
16. WHO. Epidemic and pandemic alert and response (EPR). Ebola haemorrhagic fever http://www.who.int/csr/disease/ebola/en/
17. Bausch DG, Borchert M, Grein T, Roth C, Swanepoel R, Libande ML, et al. Risk factors for Marburg hemorrhagic fever, Democratic Republic of the Congo. Emerg Infect Dis [serial online] 2003 Dec [3rd AUG 2007]. Available at http://www.cdc.gov/ncidod/EID/vol9no12/03-0355.htm
18. WHO. Epidemic and pandemic alert and response(EPR). Marburg haemorrhagic fever http://www.who.int/csr/disease/marburg/en/index.html
19. Arcady R. Mushegian, James R. Garey, Jason Martin, and Leo X. Liu .Large-Scale Taxonomic Profiling of Eukaryotic Model Organisms: A Comparison of Orthologous Proteins Encoded by the Human, Fly, Nematode, and Yeast Genomes. Genome research, 1998;8(6): 590-598
20. Banfi, S., G. Borsani, E. Rossi, L. Bernard, A. Guffanti, F. Rubboli, A. Marchitiello, S. Giglio, E. Coluccia, M. Zollo, O. Zuffardi, and A. Ballabio. Identification and mapping of human cDNA homologous to *Drosophila* mutant genes through EST database screening. *Nature Genet.* 1996;13: 167-174
21. Bassett, D.E., Jr., M.S. Boguski, F. Spencer, R. Reeves, S. Kim, T. Weaver, and P. Hieter. Genome cross-referencing and XREFdb: Implications for the identification and analysis of genes mutated in human disease. *Nature Genet.* 1997;15: 339-344
22. Mushegian, A.R., D.E. Bassett Jr, M.S. Boguski, P. Bork, and E.V. Koonin.. Positionally cloned human disease genes: Patterns of evolutionary conservation and functional motifs. *Proc. Natl. Acad. Sci.* 1997;94: 5831-5836
23. Koonin, E.V., A.R. Mushegian, M.Y. Galperin, and D.R. Walker. Common and distinctive features of bacterial and archaeal genomes revealed by computer analysis of protein sequences. *Mol. Microbiol.* 1997;25: 619-637
24. Fitch, W.M.. Distinguishing homologous from analogous proteins. *Syst. Zool.* 1970;19: 99-106
25. Henikoff, S., E.A. Greene, S. Pietrokovski, P. Bork, T.K. Attwood, and L. Hood. Gene families: The taxonomy of protein paralogs and chimeras. *Science* 1997;278: 609-614
26. Tatusov, R.L., A.R. Mushegian, P. Bork, N.P. Brown, M. Borodovsky, W.S. Hayes, K.E. Rudd, and E.V. Koonin. Metabolism and evolution of *Haemophilus influenzae* deduced from a whole genome sequence comparison to *Escherichia coli*. *Curr. Biol.* 1996; 6: 279-291
27. Feng, D.F., G. Cho, and R.F. Doolittle. Determining divergence times with a protein clock: Update and reevaluation. *Proc. Natl. Acad. Sci.* 1997; 94: 13028-13033
28. Doolittle, R.F., D.F. Feng, S. Tsang, G. Cho, and E. Little. Determining divergence times of the major kingdoms of living organisms with a protein clock. *Science* 1996; 271: 470-477
29. Expert Protein Analysis System (Revised 2007).ExPASy UniProt Knowledgebase: Swiss-Prot and TrEMBL http://expasy.org/sprot/ Accessed 27th July 2007
30. Altschul, Stephen F.Amino acid substitution matrices from an information theoretic perspective. J. Mol. Biol.1991; 219:555-65.
31. Altschul, S. F. A protein alignment scoring system sensitive at all evolutionary distances. J. Mol. Evol.1993; 36:290-300.
32. Altschul, S. F., M. S. Boguski, W. Gish and J. C. Wootton Issues in searching molecular sequence databases. Nature Genetics.1994; 6:119-129.
33. Altschul, Stephen F., Warren Gish, Webb Miller, Eugene W.Myers, and David J. Lipman . Basic local alignment search tool. J. Mol. Biol.1990; 215:403-10.
34. Claverie, J.-M. and D. J. States. Information enhancement methods for large scale sequence analysis. Computers in Chemistry. 1993; 17:191-201.
35. Gish, W. and D. J. States . Identification of protein coding regions by database similarity search. Nature Genetics. 1993; 3:266-72.
36. Henikoff, Steven and Jorga G. Henikoff. Amino acid substitution matrices from protein blocks. Proc. Natl. Acad. Sci. USA.1992; 89:10915-19.
37. Karlin, Samuel and Stephen F. Altschul. Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. Proc. Natl. Acad. Sci. USA. 1990; 87:2264-2268.
38. Karlin, Samuel and Stephen F. Altschul. Applications and statistics for multiple high-scoring segments in molecular sequences. Proc. Natl. Acad. Sci. USA. 1993; 90:5873-5877.
39. States, D. J. and W. Gish Combined use of sequence similarity and codon bias for coding region identification. J. Comput. Biol. 1994; 1:39-50.
40. States, D. J., W. Gish and S. F. Altschul . Improved sensitivity of nucleic acid database similarity searches using application specific scoring matrices. Methods: A companion to Methods in Enzymology. 1991; 3:66-70.
41. Wootton, J. C. and S. Federhen. Statistics of local complexity in amino acid sequences and sequence databases. Computers in Chemistry. 1993; 17:149-163.
42. Jeffrey M. Drazen, Elizabeth G. Phimister.. Publishing Genomewide Association Studies. NEJM 2007 July 18, 2007 Volume 357(5):496
43. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome *Nature* 21 October 2004; 431: 931-945
44. National Centre for Biotechnology Information (NCBI). 2007.Homo sapiens (human) genome view Build 36.2 statistics Accessed 21st July 2007. http://www.ncbi.nlm.nih.gov/genome/guide/human/
45. National Centre for Biotechnology Information (NCBI). Revised 2007.Sus scrofa (pig) genome view Build 1.1 statistics Accessed 21st July 2007. http://www.ncbi.nlm.nih.gov/genome/guide/pig/
46. National Centre for Biotechnology Information (NCBI). Revised 2007Gallus gallus (chicken) genome view Build 2.1 statistics Accessed 21st

July 2007. National Centre for Biotechnology Information (NCBI). 2007. 540 Complete Microbial Genomes selected: [A] – 44, [B] – 496 Accessed 21st July 2007. http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi
47. International Human Genome Sequencing Consortium Rhesus Genome Project. Revised 2006. http://www.hgsc.bcm.tmc.edu/projects/rmacaque/ Accessed 21st July 2007.
48. National Center for Biotechnology information (NCBI). Revised 2007.Genbank-viruses. Ebola & Marburg genes NCBI; Accessed May 2007 http://www.ncbi.nlm.nih.gov/genomes/static/vis.html
49. HIV Sequence database. Revised 2007. HIV-1, HIV-2 and SIV ENV proteins. http://www.hiv.lanl.gov/content/hiv-db/ALIGN_CURRENT/ALIGN-INDEX.html Accessed 18th AUG 2007
50. S. T. Cole, R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence *Nature* 11 June 1998;393:537-544
51. Hoffman,S.L., Bancroft,W.H., Gottlieb,M., James,S.L., Burroughs,E.C., Stephenson,J.R. and Morgan,M.J. Funding for malaria genome sequencing. *Nature*, 1997;387, 647
52. Gardner,M.J., Tettelin,H., Carucci,D.J., Cummings,L.M., Aravind,L., Koonin,E.V., Shallom,S., Mason,T., Yu,K., Fujii,C. et al. Chromosome 2 sequence of the human malaria parasite *Plasmodium falciparum*. *Science*, 1998; 282, 1126-1132.
53. Bowman,S., Lawson,D., Basham,D., Brown,D., Chillingworth,T., Churcher,C.M., Craig,A., Davies,R.M., Devlin,K., Feltwell,T. et al. The complete nucleotide sequence of chromosome 3 of *Plasmodium falciparum*. *Nature*, 1999; 400, 532–538.
54. Coffin, J. M. Retroviridae: the viruses and their replication, p. 1767-1848. *In* B. N. Fields, D. M. Knipe, and R. M. Howley (ed.), Virology. Lippincott-Raven Publishers, Philadelphia, Pa. 1996
55. Varmus, H. E., and P. O. Brown. Retroviruses, p. 53-108. *In* D. E. Berg, and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C. 1989
56. Jens Rydell, Gareth Jones, Dean Waters. Echolocating Bats and Hearing Moths: Who Are the Winners? *Oikos*, Sep., 1995;73(3): 419-424
57. Johannes Schul, Felix Matt, Otto von Helversen Listening for bats: the hearing range of the bushcricket *Phaneroptera falcata* for bat echolocation calls measured in the field Proceedings of the Royal Society B: Biological Sciences 2000; 267(1454): 1711-1715
58. Gerlinde Höbel and Johannes Schul Listening for males and bats: spectral processing in the hearing organ of *Neoconocephalus bivocatus* (Orthoptera: Tettigoniidae) Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology 2007;139(3): 917-925
59. Yack, Jayne E.; Fullard, James H. Ultrasonic hearing in nocturnal butterflies. Nature, 2000; 403 (6767): 265-266
60. Neil C. MitchellProcessing and analysis of Simrad multibeam sonar data Marine Geophysical Researches 1996;18(6): 729-739
61. Daniel L. Orange, Michael Angell, and David LappApplications of multibeam mapping to exploration and production; detecting seeps, mapping geohazards, and managing data overload with GIS *The Leading Edge*; April 1999; 18(4):495-501
62. Nohara, T.J. Haykin, S. AR-based growler detection in sea clutter Signal Processing 1993;41(3): 1259-1271
63. Paul G. Newman and Grace S. Rozycki. The history of ultrasound Surgical Clinics of North America 1998;78(2): 179-195
64. Misaki., W. A **model for mapping of Ebola and Marburg RNA integration sites in rhesus *Macaca mulatta* genome in silico: Ebola virus acceptor sites probably located on chromosomes 4, 6, 7, 8, 9, 14 and 15**[AJB-07-557 unpublished]
65. Comparative genomics survey of acceptors for filoviridae viruses in the Rhesus macaque (macaca mulatta) and Homo sapiens genomes: Filoviruses are still not well adapted to establish sustainable integration within the humans. AJB[unpublished]
66. Misaki., W. A Comparative genomics survey of acceptors for filoviridae viruses in the Rhesus macaque (macaca mulatta) and Homo sapiens genomes: Filoviruses are still not well adapted to establish sustainable integration within the humans. AJB[unpublished]

Example 2

Algorithms for comparative in-silico genomography (*Cis*- GNMG)- a theoretical computational approach for mapping cross-species adaptability as hosts to emerging viral pathogens.

Misaki, Wayengera

Faculty of Medicine Makerere University. P o Box 7072. Kampala, Uganda

Abstract

We have recently described Comparative Insilico genomography (Cis-GNMG) as a novel theoretical computational approach for conducting pathogen-host adaptability ecological mapping exercises Insilico. Here, we define the algorithms for interpreting Cis-GNMG results: Evolutionary acceptors (E_a); in-vivo Integration acceptors (I_a); where $E_a \approx I_a$ and $E_a = \lambda I_a$ and λ is a proportionality constant defining the relationship between Evolutionary acceptors (E_a) and in-vivo Integration acceptors (I_a). For any queried viral integration complex (VIC) across host species X,Y and Z; if $X.E_a > Y.E_a > Z.E_a$; then X is more likely to be closer in the evolutionary tree to the virus' NR than Y and Z. For any two VICs(VIC1;VIC2) studied in species X; the ratio $X.E_{a(Vic1)} : X.E_{a(Vic2)}$ approximates the molecular epidemiological distribution of the current circulation forms of VIC1 and VIC2.

Key words: Emerging viral pathogens; Host adaptability; Theoretical /computational sciences, Comparative Genomography Cis-GNMG)

Introduction

Viruses are obligate intracellular pathogens whose genetic integration into the host genome is necessary for sustainable viral tropism[1]. It is therefore imperative that to some extent, the phenotypic manifestation of viral integration, metabolism and in-vitro pathogenesis are correspondingly represented on the genotypic level within the susceptible host cell[2]. A level of homogeneity between viral genomic material and integration site may be existent. Susceptible host cells to viral tropism are thus in a way innately genotypically adapted to sustain a specified viral tropism[3,4]. On the other hand, the virus may either be innately adapted to use a specified host cell type for replication, or it must undergo evolutionary adaptability to affect that ability[5].

Since viruses are obligate intracellular organisms, it is also true that viruses can not survive outside a host beyond certain time limits. Most Viruses known today can thus be connoted as zoonotic-with a natural reservoir (host that does not exhibit diseases) being a more likely source of the virus than just spontaneous Evolution[5]. Ecological mapping exercises for delineating the natural reservoir (NR) of emerging viral pathogens have in the past been made cumbersome by virtue of the associated high virulence of the viruses involved, scarcity of resources, and the absence of a theoretical model for such work. Until recently when we defined Comparative genography and genomogrpahy(Cis-GNG/GNMG), it was impossible to assess the cross species adaptability to act as sustainable hosts to any particular virus using theoretical computational models[Misaki, W EID-07-1122 Unpublished][6].

The later approach-theoretical computational ecology is however central in evaluating the adaptability across species suspected to act as hosts to a newly described virus in resource limited settings; information that is vital in the public health control and prevention of further outbreaks by avoiding human host contact[7,8]. Through Cis-GNMG, we evaluate in discretely the viral tropism adaptability across two or more hosts by comparing the overall evolutionary homology of a specified viral integration complex (VIC) to the studied host genomic databases using statistical methods of Karlin and Altschul (1990, 1993)[9,10] integrated in the **Basic Local Alignment Search Tool** (BLAST) ,

and by so doing, we in a way measure the level of genetic absorbance across all host databases for the VIC. The Values obtained are not orthologs or paralogs as in a true BLAST homology survey, but are rather denoted as evolutionary acceptors (E_a)[9, 10]. Further Still, E_a are not in-vivo Integration acceptors (I_a); although E_a may be indescretely predictive of I_a for any given VIC and host. In this study, we try to describe standard algorithms for interpreting Cis-GNMG results using a *Genomographic survey for influenza A evolutionary acceptors in Homo sapiens(humans), Sus scrofa(pigs) and Gallus gallus(chicken).*

The Study:

Influenza A, the subtype responsible for large human pandemics, infects a spectrum of species ranging from humans, swine, equine, avian, to marine mammals. Influenza disease has been found to occur occurs in both avian species and swine, meaning that does may not be the natural reservoirs of Influenza viruses[11-13]. We have developed Comaprative In-Silico genography and genomography(Cisd-GNG/GNMG) as tools for predicting cross-species adaptability such "known virus carriers" to act as hosts for viral tropism; and evaluating the evolutionary distance from the natural reservoir. We conducted a Cis GNG survey for influenza A evolutionary-acceptors (aka E acceptors) in Homo sapiens (humans), Sus scrofa(pigs), and Gallus gallus(chicken)to trace the evolutionary trends of influenza A among these three hosts.

Design: *Comparative Insilico genograpgy(Cis-GNG)* **Materials and Software Program:** (1)BLASTN [9,10](2)Sus Scrofa(pig) genome view build 1.1[*http://www.ncbi.nlm.nih.gov/genome/guide/pig*] (3)Gallus Gallus(Chiken) genome[http://www.ncbi.nlm.nih.gov/genome/guide/chicken](4)Human genome build 36.2 Statistics[http://www.ncbi.nlm.gov/genome/guide/human](5)Influenza A 1996 H5N1 variant Neuraminidase(NA) and Hemagglutinin(HA) genes(see tables 1 and 2 for Accession numbers).**Calibration of the BLASTN tool algorithms:** The Blast tool has four calibration algorithms: Filtration (F), Expect(E), Description(D) and alignment(A). A True Blast similarity search sets these paradigms at high stringency values, while in the Cis GNG, because we compared organisms of distant taxa, the values were set at lower values(no filtration, E=100, both A and D at 10) .**Intervention:** Both NA and HA genes' complete cds were queried against the whole genomes of Man, pig and Chicken using a BLASTN calibrated as above. **Measured variables**: Four parameters were available to us for estimating the E-acceptor ratio across the studies hosts(1)Hits to DB(2)HSP(high Scoring pair)-gapped(3)HSP successfully gapped(4) and sequences producing significant alignment; named in ascending order of priority

RESULTS: Number of hits to DB: Of all 2.85 Gb nucleotides in the human genome, 2.7Gb in Sus scrofa, and 1.2 Gb in Gallus gallus and a query Length of 1760 nucleotides in the Influenza A virus(A/Goose/Guangdong/1/96(H5N1)) hemagglutinin (HA)gene' complete cds: HA hits to DB were 6,351,123 in the human; 347,716 in the pigs and 2,723,004 in chicken genome compared to 4,868,784 in humans; 283,394 in pigs and 2,088,939 hits to DB in chicken for query length of 1458 nucleotides of Influenza A virus (A/Goose/Guangdong/1/96(H5N1)) neuraminidase (NA)gene's complete cds.

High-scoring Segment pairing the fundamental unit of BLAST algorithm output is the High-scoring Segment Pair (HSP). An HSP consists of two sequence fragments of arbitrary but equal length whose alignment is locally maximal and for which the alignment score meets or exceeds a threshold or cutoff score. The numbers of HSP's gapped (taken to be directly proportional to the number of integration acceptors in-vivo) produced by BLASTN homology survey in humans, pigs and chicken for HA were 62 in humans; 12 in pigs and 306 in chicken compared to 25 in humans; 3 in pigs and 190 in chicken produced by Influenza A neuraminidase (NA) gene. These figures corresponded to the number of successful extensions for each query. There were no HSP's successfully gapped, HSP's better than 0.01 without gapping, or sequences better than 0.01 KARLIN-ALTSCHUL Statistical Constant values were Lambda= 1.37, H= 0.711, and K= 0.31

(see tables 1 and 2 for further details of Blastn homology survey characteristic for each query experiment)

Table 1 results of the distribution of E acceptors for Influenza HA across Homo sapiens (human), Sus scrofa(pig) and gallus gallus(chicken) whole genomes)[*,+]

BLASTN homology search Characteristics	Human build 36 genome	Pig genome build	Gallus gallus (chicken) build 2 genome
Matrix: blastn matrix:	1 -3	1 -3	1 -3
Gap Penalties: Existence: Extension	5: 2	5: 2	5: 2
Number of Sequences:	7067	3836	17507
Number of Hits to DB:	6351123	347716	2723004
Number of extensions:	365601	21910	161893
Number of successful extensions:	62	12	306
Number of HSP's gapped:	62	12	306
Length of query:	1760	1760	1760
Length of database:	5865913461	681739921	1078736841

Query= gi|5805286|gb|AF144305.1|Length=1760 DEFINITION Influenza A virus (A/Goose/Guangdong/1/96(H5N1)) hemagglutinin (HA)gene, complete cds.[*]**Databases:** human build 36 genome database (reference and alternate assemblies)7067 sequences; 5,865,913,461 total letters/ Sus scrofa HTGS 3836 sequences; 681,739,921 total letters/ chicken build 2 genome database (reference assembly only)17,507 sequences; 1,078,736,841 total letters

Table 2 results of distribution of E- acceptors for Influenza NA in Homo sapiens (human), Sus scrofa(pig) and gallus gallus(chicken) whole genomes)[*,+]

BLASTN homology search Characteristics	Human build 36 genome	Pig genome build	Gallus gallus (chicken) build 2 genome
Matrix: blastn matrix:	1 -3	1 -3	1 -3
Gap Penalties: Existence: Extension	5:2	5:2	5:2
Number of Sequences:	7067	3836	17507
Number of Hits to DB:	4868784	283394	2088939
Number of extensions:	249736	18299	112291
Number of successful extensions:	25	3	190
Number of HSP's gapped:	25	3	190
Length of query:	1458	1458	1458
Length of database:	5865913461	681739921	1078736841

[+]**Query=** gi|5805284|gb|AF144304.1| Length=1458 DEFINITION Influenza A virus (A/Goose/Guangdong/1/96(H5N1)) neuraminidase (NA)gene, complete cds.***Database:** human build 36 genome database (reference and alternate assemblies)7067 sequences; 5,865,913,461 total letters/ Sus scrofa HTGS 3836 sequences; 681,739,921 total letters/ chicken build 2 genome database (reference assembly only)17,507 sequences; 1,078,736,841 total letters

CONCLUSIONS

Our results serve to Show, that as expected from the distant evolutionary distance between viruses and the 3 hosts studied(Pigs, chicken and Man), there are completely no homologs(paralogs or orthologs). This makes this study when viewed from the perspective of the true BLAST similarity searches useless. However, on the part of accessing the evolutionary adaptability of the studied hosts, this study has a lot of significance. For instance, here we demonstrate that the ratio of E-acceptors for HA and NA across the three studies species is directly proportionate to the current molecular epidemiological data on currently circulating antigenically distinct hemagglutinin (HA) and neuraminidase (NA) types (15/9)[11-13]. This ratio is also approximately maintained across all influenza A hosts analyzed (62/25 for Homo sapiens; 12/3 for Sus scrofa and 306/190 for gallus gallus); although the relationship is closer in humans and chicken (***see tables 1, 2 and 3).*** This leads us to our first equation and algorithm for interpreting Cis-GNG results:

Table 3 Ratio of HA and NA across the 3 Studies Species

Human build 36 genome	Pig genome build	Gallus gallus (chicken) build 2 genome
62/25	12/3	306/190

Note the relationship between the distribution of E-acceptors to the molecular currently documented epidemiological distribution of antigenically distinct hemagglutinin (HA) and neuraminidase (NA) types (15/9) across chicken and man

For any two VICs(VIC1;VIC2) studied in species X; the ratio $X.E_{a(Vic1)}: X.E_{a(Vic2)}$ *approximates the molecular epidemiological distribution highest possible circulation forms of VIC1 and VIC2.*
……..(E-Algorithm1)

From equation 1, and the fact that this ratio is maintained across all the 3 studies hosts, it may mean that that E-acceptors (E_a) may actually be proportional to In-vivo integration acceptors (I_a). This brings us to our second algorithm or equation for Cis-GNG

In any particular species; $E_a \propto I_a$ *and* $E_a = \lambda I_a$ *; where λ is a proportionality constant defining the relationship between Evolutionary acceptors (*E_a*) and in-vivo Integration acceptors (*I_a*).*
……..(E-Algorithm2)

Basing on the above E-algorithms(1 and 2), it means that Chicken have more In-vivo Integration acceptors for both HA and NA than man and pigs. If we unquestioningly held the hypothesis that a natural reservoir should be one that is most well adapted for viral integration and tropism, It would mean that birds (chicken) may thus be more

closely related to the natural reservoir of Influenza viruses than man and pigs-basing on the hypothesis that the natural reservoir is innately adapted to express more In-vivo Integration acceptors (I_a). This brings us to our third equation or algorithm (E-Algorithm3)

For any queried viral integration complex (VIC) across host species X,Y and Z; if $X.E_a > Y.E_a > Z.E_a$ then X is more likely to be closer in the evolutionary tree to the virus' natural reservoir(NR) than Y and Z. ……..(E-Algorithm3)

However, considering that the exact I_a for many viruses are yet to be defined; exact relationship between In-vivo acceptors and evolutionary (E-)acceptors defined here is debatable, and that the reverse is an equally questionable hypothesis, that the natural reservoir is one that has the least number I_a . The later thesis is emphasized by the fact that a natural reservoir does not exhibit disease, and yet both the VICs used in this study mediate diseases. If I_a were mediators of disease, then this hypothesis would hold. Holding onto this hypothesis, it may thus mean that pigs are more closely related to Influenza NR in the evolutionary tree, since pigs from this study have fewer E(and thus I) acceptors, and the NR is that organism in nature that although habors the virus, does not exhibit associated disease. This question canonly be resolved on research geared at establishing the actually in-vivo integrations acceptors, and comparing their prevalence across the hosts studied.

For many viruses, the exact in-vivo integration acceptors are currently undefined. Most are not expected to be present as naked DNA but rather associated with histones and other DNA-binding proteins in chromatin. DNA packaging in vivo is expected to influence integration site selection, and the choice of integration site may have profound effects on both the virus and the host.[14,15] The determinants of integration efficiency in vivo thus remain incompletely defined, despite their predicted importance in defining targets for future therapeutic and vaccine strategies, pathogenesis elucidation, as well as a modeling the evolutionary trends of successful viral cross over[15]. Past experiments for delinating in-vivo viral integration acceptors (I-acceptors) for other retroviruses like Porcine Endogenous virus(PEV), Murine Leukamia virus(MLV) and Human immunodeficiency virus(HIV), have employed in-vitro, in-vivo and bio-informatic in-silico methods[14,15].

Here, we hold that *Cis-GNMG* and the above 3 algorithms may be used as a theoretical tool for conducting ecological mapping across species for adaptability to act as hosts to any given viral pathogen insilico, and that, as demonstrated here; Avian derived influenza strains seem to be innately adapted for integration in humans than pigs, making the need for porcine as mixing vessels obsolete, meaning that, viral recombination and mutations necessary for bird-to-human crossovers occur within birds rather than an intermediary porcine host. Integrating Knowledge of the actual Ia for various viruses into this model will generate a theoretical tool

able to assess Insilco the actual risk that a host is at in terms of viral transmission.

Conflict of Interests: None

References

1. Morse SS, Schluederberg A. Emerging viruses: the evolution of viruses and viral diseases. J Infect Dis 1990; 162:1-7.
2. Ewald PW. The evolution of virulence. Scientific American, 1993; April: 86-93.
3. Ewald PW. *The Evolution of Infectious Disease*. Oxford University Press, 1994.
4. Ampel NM. Plagues--What's past is present: thoughts on the origin and history of new infectious diseases. Rev Infect Dis 1991; 13:658-665.
5. Wayengera.M., Olobo.J.,Kajjumbula. H., Kaddu. M .2005. Global Bird flue threat. What broader prospects to consider? http://www.bmj.com/cgi/eletters/331/7523/975
6. **Misaki, W. Comparative in-silico genography and genomography as in-silico tools for predicting the adaptability across species to act as suitable hosts for zoonotic pathogens** EID-07-1122 Unpublished
7. McSweegan E. The infectious disease impact statement: a mechanism for addressing emerging infectious diseases. Emg Infect Dis 1996; 2:103-108.
8. World Health Organization. Strategic Plan Outline 1996-2000. Emerging and Other Communicable Diseases Surveillance and Control, WHO EMC/96.1.
9. Altschul, S. F., M. S. Boguski, W. Gish and J. C. 1994. Wootton Issues in searching molecular sequence databases. Nature Genetics; 6:119-129.
10. Altschul, Stephen F., Warren Gish, Webb Miller, Eugene W.Myers, and David J. Lipman. Basic local alignment search tool. J. Mol. Biol. .1990; 215:403-10.
11. Treanor JJ. 2004. Influenza virus. In Mandell GL, Bennett JE, Dolin R, eds. *Principles and Practice of Infectious Diseases*, 6th Ed. New York: Churchill Livingstone Inc.,: 2060–2085.
12. Vilchez RA, Fung J, Kusne S. 2002. The pathogenesis and management of influenza virus infection in organ transplant recipients. *Transpl Infect Dis*; 4: 177–182.
13. Nicholson KG. 1998. Human influenza. In: Nicholson KG, Webster RG, Hay AJ, eds. *Textbook of Influenza*. Oxford, England: Blackwell Science,: 219–264
14. **Moalic Y**, **Blanchard Y**, **Félix H**, **Jestin A**. Porcine endogenous retrovirus integration sites in the human genome: features in common with those of murine leukemia virus. J Virol. . 2006 Nov; 80(22):10980-8. Epub 2006 Aug 23.
15. **Pryciak, P. M., A. Sil, and H. E. Varmus.** Retroviral integration into minichromosomes in vitro. EMBO J. 1992; **11:**291-303

Example 3

Comparing the evolutionary adaptability of humans and rhesus macaca as hosts to filoviruses by way of proprietary Insilco algorithms

Misaki, Wayengera[δ,+], Henry, Kajjumbula[α], Wilson, Byarugaba[β]

[δ]Research fellow., [β]Assoc. Prof and Head, Division of Human Genetics- Faculty of Medicine Makerere University. P o Box 7072. Kampala, Uganda [α]Lecturer, Dept Of Microbiology Faculty of Medicine Makerere University. P o Box 7072. Kampala, Uganda

Abstract

Viral homogeneity to flanking sequences of introns may be a facilitator of RNA viral integration by target DNA primed reverse transcription(TPRT). To an extent, this minimal homogeneity between viral genes to the host genomic intronic or exonic material may thus be used to evaluate likely host susceptibility. Ebola and Marburg form the two geneses of a family of negatively stranded viruses called filoviridae. Although the exact natural reservoir (NR) of both viruses still remains elusive, with the *African Fruit bat species Rousettus aegyptiacushighly suspected, it is clear that these viruses* are the cause fulminating outbreaks of viral hemorrhagic fever (VHF) in both man and monkeys within equatorial Africa. We aimed to ascertain the evolutionary adaptability of both man and Rhesus macaca as susceptible hosts to filovirus tropism, as well as define algorithms for a classical comparative Insilico genography(Cis-GNG), the method employed by querying the Glycoprotein genes and Whole genomes of EBO and MBG against the Genomes of Homo-sapiens and Rhesus macaca using low specificity BLASTN algorithms. Numbers of HSP's better than 2.0 with no gapping were used as a proportionate measure of the distribution of evolutionary (E) acceptors (E_a). 21 E_a for Marburg GP compared to only 7 E_a for Ebola GP were found in the rhesus genome, while 8 E_a for MBGV compared to 3 E_a for Ebola GP were found in human genome; implying that circulating epidemiological distribution of EBOV and MBGV may currently approximate 3/1(being maintained across both man and Rhesus macaca as 8/3 or 21/7). We predict that Marburg may be better adapted to use human and Rhesus Macaca cells for tropism than Ebola. However, Rhesus macaca may actually be more susceptible (about >2-3 times) to both Marburg and Ebola tropism (and associated VHF) than man.

Key words: Emerging viral pathogens; Host adaptability; Theoretical /computational sciences, Comparative Genomography Cis-GNMG); Ebola Virus (EBOV), Marburg Virus (MBGV), Viral haemorrhagic fever (VHF)

Introduction

Viruses are obligate intracellular microbes whose integration is necessary for viral replication and survival. The exact in-vivo acceptors for viral integration are poorly known. However, recent studies have availed evidence that RNA and in particular retroviral(HIV)viral integration is favoured within ALU or LINE elements(Stevens and Griffith, 1994; Stevens and Griffith, 1996; Pryciak et al., 1992; Vijaya et al., 1986).We recently demonstrated that homogeneity between viral LTR and the flanks of host intronic or exonic material is a likely major predictor of viral integration (Misaki, W et al 2007), further supporting the earlier hypothesis that viral integration follows a possibly same target DNA primed reverse transcription as is the case in Group II introns(Zimmerly S, Guo H, Eskes R et al 2005). This observation is drawn from previous in-vivo and invitro finding that viral integration is favoured around LINE elements (Stevens and Griffith, 1994;

Pryciak et al., 1992; Vijaya et al., 1986), which are ancestors of Group II introns- a novel class of self splicing and selectively mobile RNAs distributed across all taxa genomes from viruses, prokaryota, archaeal, Eukaryota to vertebrates(inclusive of Humans- where they are existent as reffered to as spliceosomes or non-long terminal repeats-LTR or Long interspesed Nuclear elements-LINE), interspersing genes within the genomes . The proportion of those interrupted genes is low in yeast and increases in the lower eukaryotes; with fewer genes being uninterrupted in higher eukaryotes [Arkhipova I and Meselson M 2000; Arkhipova I and Meselson M 2005; Michel F and Ferat JL 1995]. The lengths of the introns, which can extend from 0.1±100 kb in vertebrates, determine the gene size. Retroposons(also referred to as non-long Terminal Repeats(LTR) or Long interspersed nuclear elements-LINEs) form the highly mobile class of Group II introns, the transposons.This Group has a unique feature of encoding a reverse transcriptase (RT) open reading flame (ORF) moiety in its IV arm which they use to insert into a predefined sites at high efficacy(retrohoming) or unrelated sites at low rates (retrotransposing)(Arkhipova I and Meselson M 2000). Depending on the function of the protein coded by this moiety, restroposons can be classified as Restriction endonuclases(REase)-like, or apurinic apyramidinic endonuclease(APE)-like (Arkhipova I and Meselson M 2005; Michel F and Ferat JL 1995). The RT ORF has an endonuclease motif within its C terminal Zn^{2+} fingerlike moiety, whose function is like that of all other REases, and is mediated through recognising 6-12 bp nt in the target sequence, and cleaving within, or close to it. While the RNA component of the Ribonucleoprotein often catalyses sense strand cleavage, the antisemnse strand is cleaved by this DNase(Zimmerly S, Guo H, Eskes R et al 2005).These specificity sites are to be found within flanking regions of retrohoming introns(also known as exons).

If, as widely believed viral integration is mediated by the same Target -DNA primed Reverse transcription (TPRT) as in the case for Transposons, then there must exist a certain level of homogeneity between the host genome and the virus genes to facilitate EBS-IBSI/EBS-IBS2. It may thus be possible to assess the distribution of this homogeneity in various potential hosts of any virus. Such measurements would enable us to predict which veterinary viruses are potential pathogens to humans; and which veterinary hosts are potential sources of Human viruses. By directing a section of the viruses gene to the likely hosts, its possible to measure the e prevailing ratio of homogeneity by way of insilico similarity searches employing the BLAST algorithms (Autschul 1993). This theoretical computational approach for mapping cross species adaptability to facilitate viral integration and tropism has been termed Comparative Insilico Genography and genomography(Cis-GNG/GNMG), a deviation from true similarity searches of comparative genomics in that it uses a third party viral gene as a “geomic-intron-exonic energy” to scan for virus-exonic homogeneity within the potential hosts genome.

Ebola viruses (EBOV) and Marburg viruses (MBGV) form the major two geneses of a class of negative-stranded RNA virus family called Filoviridae (Varmus and Brown. 1989; Bausch et al 2006). Both EBOV and MBGV are members of the family *Filoviridae*, order *Mononegavirales*, which groups together nonsegmented, negative-stranded RNA viruses, namely filoviruses, paramyxoviruses, rhabdoviruses and bornaviruses (Towner,et al 2006; Mayo 1998) [3,4]. Generally, filoviruses (Ebola, Marburg) are very similar in morphology, density and sodium dodecyl sulfate - polyacrylamide gel electrophoresis (SDS-PAGE) profile. The particles are pleomorphic, meaning they can exist in many shapes. Their basic structure is long and filamentious, essentially bacilliform, but the viruses often takes on a "U" shape, and the particles can be up to 14,000 nm in length and average 80 nm in diameter. The virus consists of a nucleocapsid, surrounded by a cross-striated helical capsid. There is an axial channel in the nucleocapsid, and the whole virion is surrounded by a lipoprotein unit derived from the host cell (Sanchez, et al 1993; Khan, et al 1993; Dowell,et al 1995; Feldmann,et al 1993;

Felsenstein 1993; Baize, et al 1999; Leroy, et al 2000; Yang, et al 1998; Watanabe, et al 2000; Bowen,et al 1990; Peterson, et al 2004) The main pathogenic protein of the EBOV is the Glycoprotein (GP). EBOV GP is thought to induce endothelial cell disruption and cytotoxicity in blood vessels (Watanabe, et al 2000) and to mediate entry into the host cell (Bowen,et al 1990). GP thus forms the model Pre-integration Complex (PIC) for this family of viruses. Both viruses cause outbreaks of rare but fulminating of Viral Hemorrhagic fevers (VHF) in equatorial Africa, with VHF occurring both in man and chimpanzees(Varmus and Brown. 1989; Bausch et al 2006).

Until now, our knowledge of the natural reservoir and distribution of susceptible hosts to highly virulent viruses such as MBGV/EBOV has been limited mainly because ecological mapping exercises for delineating the natural and susceptible hosts for highly virulent viruses require extreme Infection prevention protocols (level IV); the field trails are expensive to conduct, and there is no theoretical model for conducting such. *The exact natural reservoir of filoviridae viruses is thus still debatable, with recent studies showing the African Fruit bat species Rousettus aegyptiacus as the possible naturally occurring host that does not exhibit disease* (Peterson, et al 2004).

We aimed to ascertain and compare the evolutionary adaptability of both R.macaca and Homo sapiens to act as susceptible hosts to EBOV/MBGV using the CiS-GNG approach described.

METHODS:

Design: *Comparative Insilico Genography(Cis_GNG)*

Materials and Software Programs: *(1)BLASTN* (Autschul 1993). *(2)Rhesus Macaca mullattta genome[*http://www.hgsc.bcm.tmc.edu/projects/rmacaque/*] (3)Human genome build 36.2 Statistics[http://www.ncbi.nlm.gov/genome/guide/human] and (4)Both Ebola and Marburg GP Proteins[EBOV GP Accession:* U23187, version: U23187.1 GI:1041204; MBGV GP Accession: AF005735; Version: AF005735.1 GI:2459879](5) Whole Genomes of MBGV and EBOV[EBOV Genome Accession NC_001608 ; version: NC_001608.2 GI:13489275; MBGV Genome Accession: NC_002549, version NC_002549.1 GI:10313991]

Calibration of the BLASTN tool algorithms: The Blast tool has four calibration algorithms: Filtration (F), Expect (E), Description (D) and alignment (A). A True Blast similarity search sets all at stringency values, while in the Cis GNG, because we aim to compare organisms of distant taxa, parameters are set at less stringent values. The values used in this study were set at: none filtration, E=100, both A=10 and D at 10

Intervention: Both Ebola and Marburg GP gene sequences were queried against the whole genome of the Jan. 2006 rhesus macaque (*Macaca mulatta*) draft assembly — v.1.0, Mmul_051212 (rheMac2), and homo sapiens(Human) genome(build 35) using BLAST-N analysis from Baylor college . The Query of whole Ebola and Marburg genomes

against both Homo sapiens (Human) and Rhesus macaca was done using the BLASTN tool from the National Center for Biotechnology Information (NCBI).

Measured variables: Four parameters were available to us for estimating the E-acceptor ratio across the studied hosts (1) Hits to DB (2)HSP(high Scoring pair)-gapped(3)HSP successfully gapped(4) and sequences producing significant alignment; named in ascending order of priority. The later two were used to calculate E-acceptor ratios. All E-acceptor values were interpreted basing on the Cis-GNG algorithms described else where[Misaki, W unpublished]: (1) In any particular species; $E_a \propto I_a$ and $E_a = \lambda I_a$ and λ is a proportionality constant defining the relationship between Evolutionary acceptors (E_a) and in-vivo Integration acceptors (I_a). (E-Algorithm3)(2) For any queried viral integration complex (VIC) across host species X,Y and Z; if $X.E_a > Y.E_a > Z.E_a$, then X is more likely to be closer in the evolutionary tree to the virus' natural reservoir(NR) than Y and Z.(E-Algorithm2)(3) For any two VICs(VIC1;VIC2) studied in species X; the ratio $X.E_{a(Vic1)}$: $X.E_{a(Vic2)}$ approximates the molecular epidemiological distribution of the current circulation forms of VIC1 and VIC2. (E-Algorithm1)]

RESULTS

Hits by DB: Of all the contigs length 2.87 Gb (2,863,665,185) bases in the genome of rhesus macaque, Marburg GP blast hits to DB were 6,451,736 compared to 4,012,901 for Ebola. Of all 2.88 Gb(2,851,330,913) nucleotides in the human genome, Marburg GP blsat hits were 1,013,987 compared to 1,013,492 for Ebola. NCBI whole genome blasts yielded 315096 blast hits to DB for the Lake Victoria marburgvirus complete genome (Length=19112) compared to 278738 for the Zaire Ebola virus strain Mayinga complete genome (Length=18959).[see tables 3 and 4)

Sequences producing significant alignment: Marburg GP genomic RNA exhibited 18 alignments relative to 7 of Ebola GP within the rhesus macaca mulatta genome compared 8 MBGV alignments relative to 3 EBOV in the human genome **(See tables 1 and 2).** All Ebola aligning Sequences had 100% efficiency compared to 66.6 %(12/18) of Marburg within the Rhesus macaca genome. All EBOV aligning sequences within the Homo sapiens genome had 100% efficiency compared to 62.5 %(5/8) for Marburg. (**See tables 1-4).**

Site mapping of loci of integration Marburg GP genomic RNA had all its18 alignments located on undefined scaffolds of the rhesus macaca genome compared to 7 of Ebola located on chromosomes 4, 6, 7, 8, 9, 14 & 15.(See tables 1 and 2). It was not possible to map the integration sites of both Ebola and Marburg within the Human genome, although we hold that these may take the same pattern as above.

High-scoring Segment pairing the fundamental unit of BLAST algorithm output is the High-scoring Segment Pair (HSP). An HSP consists of two sequence fragments of arbitrary but equal length whose alignment is locally maximal and for which the alignment score meets or exceeds a threshold or cutoff score. 21 HSP's better than 2.0; with no gapping were produced by Marburg GP RNA compared to only 7 for Ebola in the rhesus genome, while 8 HSP's better than 2.0 without gapping were produced by MBGV in the human genome compared to 3 of ebola. KARLIN-ALTSCHUL Statistical Constant values were Lambda= 1.37, H= 0.711, and K= 0.31

Whole genome blast analysis using NCBI BLASTN tool yielded no significant alignments.

(For details of BLASTN results using the Baylor College and NCBI Blast tools, see suppl. Files 2-7)

For purposes of evolutionary Acceptor (E_a) measurements; numbers of HSP's better than 2.0 with no gapping or sequences producing significant alignements used.. 21 E_a for Marburg GP compared to only 7 E_a for Ebola GP were found in the rhesus genome, while 8 E_a for MBGV compared to 3 E_a for Ebola GP were found in human genome

DISCUSSION

"Our results bring to the fore the possibility of using this theoretical computational approach; Comparative Insilico Genography and genomography(Cis-GNG/GNMG) for evaluating of various mammalian hosts' susceptibility to viral tropism. Here, we show that NHPs of the Rhesus macaca type are perhaps more susceptible to both filoviruses studied than man, with Marburg being more adapted to integrate within both Rhesus and human genomes. Further, we show that if viral integration follows the same pattern as group II Retrolements(RTE) do; then this method can be used to map areas within the host genome with highest distribution of integration sites. Note that the method does not serve to describe what the exact integration acceptors are(Stevens and Griffith, 1994; Stevens and Griffith, 1996;Pryciak et al., 1992; Vijaya et al., 1986), for which reason we have decided to label these sites as evolutionary(E-) acceptors. Moreover, because this model only uses a section of the viral genome*(GP gene) as a model pre-integration complex (PIC), it can not serve the purpose of defining the exact in-vivo acceptors, though it may map where they are likely to exist. In nature, the actual in-vivo PIC of the

virus is the entire viral progeny (proviral RNA) (Stevens and Griffith, 1994; Stevens and Griffith, 1996; Pryciak et al., 1992; Vijaya et al., 1986). However, regardless of the length of viral genome used, if as widely believed, TPRT is the method employed by RNA viruses to integrate into host genomes and homogeneity between viral (intron) integration sites and the viral PIC is a determinant of viral integration as hypothesized, any component of viral gene may as well serve the purpose. Basing on the above reasoning, it may be that we are demonstrating for the fist time here that while Marburg GP genomic RNA may have more integration sites within the Rhesus macaca genome (18- with approx 66.6% alignments) the exact location of these integration sites are imperfect and located on vaguely defined scaffolds. On the other hand, Ebola GP genomic RNA exhibits, although fewer (7) relative Marburg, more perfectly aligned [all 100% similar for the short length of Scores values on the query: 42-48] "integrations sites" located on chromosomes 4, 6, 7, 8, 9, 14 and 15 of the Rhesus macaca genome. Up the tree to the Human genome, 8 hits with the MBGV GP were hit relative to 3 hits with EBOV GP. These findings may imply that the Rhesus macaca are a better well adapted reservoir for Ebola virus despite the demonstrable high prevalence of integration sites for Marburg.

Table1. Sequences Producing Significant alignments For Ebola and Marburg GP Query against the Human genome

Sequence With Significant alignment	Score(Bits)	E-Value
A. Ebola GP Queried against the Human Genome(3 Alignments)		
Hsap00000000.project_hcdm_AB 20020223 8787bp	42	0.35
Hsap00000000.project_hccz_AA 20020223 68611bp	42	0.35
Hsap00000000.project_hahl_AD 20020321 12087bp	42	0.35
B. Marburg GP Queried Against Human Genome(8 Alignments)		
Hsap00000000.project_hcxy_AA 20020304 146620bp	46	0.019
Hsap00000000.project_hccy_BG 20020129 1662bp	46	0.019
Hsap00000000.project_hazl_AA 20020312 170871bp	46	0.019
Hsap00000000.project_heow_AV 20030224 1451bp	40	1.2
Hsap00000000.project_heoo_EL 20030224 966bp	40	1.2
Hsap00000000.project_hduh_OF 20020702 718bp	40	1.2
Hsap00000000.project_hdgf_AA 20020425 160708bp	40	1.2
Hsap00000000.project_hbdy_AA 20020725 188251bp	40	1.2

The table shows 3 sequences with significant alignment to Ebola GP in humans compared to 8 for Marburg GP. If these Sequences are predictive of evolutionary acceptors(E_a; and $Ea \infty$ In−vivo integration acceptors (I_a)); then Marburg is more well adapted for human tropism than Ebola.

Genography may thus be a viable insilico tool for finger printing hosts for historical events (induced by retroelement mobile activity) that may favour RNA viral tropism and integration. To further support the TPRT theory of viral integration, we note that a number of naturally existing intragenomic antiviral pathways such as the RNAi and cosuppression pathways present in many organisms appear to have evolved primarily for purposes of protecting the host genome against invasion by mobile genetic elements such as transposons and viruses, which produce aberrant RNA or dsRNA in the host cell when they become active (Dalmay,et al 2000; Jensen, et al 1999; Ketting,1999; Ratcliff,et al 1999; Tabara,et al 1999; Malinsky,et al 2000; Anandalakshmi,et al 2000). Specific mRNA degradation prevents transposon and virus replication, although some viruses are able to overcome or prevent this process by expressing proteins that suppress PTGS (Lucy,et al 2000; Voinnet, et al 2000). It is thus most likely that Genography achieves its primary aim by assessing the homogeneity between intron-extronic regions to the virus-exon in-vivo intergration regions, the main reason hits to DB obtained are referred to here evolutionary (E-) acceptors. Our findings are thus useful on the part of accessing the evolutionary adaptability of the studied hosts to favour viral integratiion. For instance, from the first algorithm of Cis-GNG, which states that: "*For any two VICs(VIC1;VIC2) studied in species X; the ratio X.$E_{a(Vic1)}$: X.$E_{a(Vic2)}$ approximates the molecular epidemiological distribution of the current circulation forms of VIC1 and VIC2*.(E-Algorithm1)"; we can deduce that the currently circulating epidemiological distribution of EBOV and MBGV approximates 3/1(being maintained across both man and Rhesus macaca as 8/3 or 21/7)

This figure is inferred basing on the thinking that both hosts studied have had enough exposure to RTE activity and filoviruses in the past to saturate the homolgy. The second equation for interpreting Cis-GNG results states that: "*In any particular species; $E_a \propto I_a$ and $E_a = \lambda I_a$; where λ is a proportionality constant defining the relationship between evolutionary acceptors*

(E_a) and in-vivo Integration acceptors (I_a)".
Basing on the above E-algorithms (1 and 2), it may mean that Rhesus macaca have more In-vivo Integration acceptors for both EBOV (8) and MBGV (21) than man (in whom EBOV GP I_a =3, and MBGV GP I_a=7).

Table2. Sequences Producing Significant alignments For Ebola and Marburg GP Query against the Rhesus macaca genome

A. Ebola GP Queried against the Rhesus Genome(7 Alignments)		
Sequence With Significant alignment	Score(Bits)	E-Value
gnl|Mmul_051212|Chr9	48	0.023
gnl|Mmul_051212|Chr14	44	0.36
gnl|Mmul_051212|Chr8	44	0.36
gnl|Mmul_051212|Chr15	42	1.4
gnl|Mmul_051212|Chr7	42	1.4
gnl|Mmul_051212|Chr6	42	1.4
gnl|Mmul_051212|Chr4	42	1.4
B. Marburg GP Queried Against Human Genome(18)		
gnl|Mmul_051212|Scaffold1099553000140	46	0.12
gnl|Mmul_051212|Scaffold1099553000132	46	0.12
gnl|Mmul_051212|Scaffold1099548049108	46	0.12
gnl|Mmul_051212|Scaffold1099553000136	44	0.47
gnl|Mmul_051212|Scaffold1099553000090	44	0.47
gnl|Mmul_051212|Scaffold1099548049513	44	0.47
gnl|Mmul_051212|Scaffold1099548049201	44	0.47
gnl|Mmul_051212|Scaffold1099553000123	42	1.8
gnl|Mmul_051212|Scaffold1099553000104_2_2	42	1.8
gnl|Mmul_051212|Scaffold1099553000056	42	1.8
gnl|Mmul_051212|Scaffold1099553000033	42	1.8
gnl|Mmul_051212|Scaffold1099553000020	42	1.8
gnl|Mmul_051212|Scaffold1099548049706	42	1.8
gnl|Mmul_051212|Scaffold1099548049414	42	1.8
gnl|Mmul_051212|Scaffold1099548049409	42	1.8
gnl|Mmul_051212|Scaffold1099548049324	42	1.8
gnl|Mmul_051212|Scaffold1099548049086	42	1.8
gnl|Mmul_051212|Scaffold1099548049085	42	1.8

This table shows 7 sequences with significant alignment to Ebola GP in Rhesus macaca genome compared to 18 for Marburg GP. If these Sequences are predictive of evolutionary acceptors (E_a; and Ea∞ In-vivo integration acceptors (I_a)); then Marburg is more well adapted for macaca tropism than Ebola. From Table 1 and 2; its likely that Rhesus macaca are thus more suitable for tropism by filoviruses than humans.

Table 3 results of the distribution of E- acceptors for Ebola GP across Homo sapiens (human) and Rhesus mulatta whole genomes)[*,+]

BLASTN homology search Characteristics	**Human build 35 genome** Human-phase_1 UnmaskedDB	**rhesus macaque, *Macaca mulatta*** Jan. 2006 draft assembly
Matrix: blastn matrix:	1 -3	1 -3
Gap Penalties: Existence: Extension	5: 2	5: 2
Number of Hits to DB:	1,013,492	4,012,901
Number of extensions:	4502	4012901
Number of sequences better than2.0	3	7
Number of Gapped HSPs:	4	7
Length of query:	2408	2408
Length of database:	715,138,083	2,863,665,185

[*]**Query=** *EBOV GP Accession:* U23187, version: U23187.1 GI:1041204, Length 2408 Zaire Ebola virus Mayinga strain glycoprotein (GP) gene, complete cds.
[*+]**Databases:** human build 35 genome database (Human-phase_1 UnmaskedDB) 155,554 sequences; 715,138,083 total letters; ***Macaca mulatta*** Jan. 2006 draft assembly — v.1.0, Mmul_051212(Mmul20051212-Chr.fa)21 sequences; 2,863,665,185 total letters

Table 4 results of the distribution of E- acceptors for MarburgGP across Homo sapiens (human) and Rhesus mulatta whole genomes)[*,+]

BLASTN homology search Characteristics	**Human build 35 genome** Human-phase_1 UnmaskedDB	**rhesus macaque, *Macaca mulatta*** Jan. 2006 draft assembly
Matrix: blastn matrix:	1 -3	1 -3
Gap Penalties: Existence: Extension	5: 2	5: 2
Number of Hits to DB:	1,013,987	6,451,736
Number of extensions:	1013987	6451736
Number of sequences better than2.0	8	18
Number of HSPs(ungapped vs gapped):	8	21/109
Length of query:	2046	2046
Length of database:	711,871,449	3,014,103,096

[*]**Query=** MBGV GP Accession: AF005735; Version: AF005735.1 GI:2459879 Marburg virus strain M/Germany/Marburg/1967/Ratayczak glycoprotein precursor (GP) gene, complete cds.[*+]**Databases:** human build 35 genome database (Human-phase_1 UnmaskedDB) 155,554 sequences; 715,138,083 total letters; ***Macaca mulatta*** Jan. 2006 draft assembly — v.1.0, Mmul_051212(Mmul20051212-Chr.fa)21 sequences; 2,863,665,185 total

This may also mean that Marburg is perhaps more adapted to use both human and Rhesus tissue for tropism than Ebola. This data seems to tally with recent epidemiological findings that suggest that Marburg may actually be more common that previously presumed[15]. If we unquestioningly held the hypothesis that a natural reservoir should be one that is most well adapted for viral integration and tropism, it would mean that Rhesus macaca are thus more closely related to the natural reservoir of filoviruses than man; and are more susceptible to both EBOV and Marburg VHFs than man. This agrees with our third equation or algorithm (E- Algorithm3) for interpreting Cis-GNG

results that states that:" *For any queried viral integration complex (VIC) across host species X,Y and Z; if* $X.E_a > Y.E_a > Z.E_a$ *then X is more likely to be closer in the evolutionary tree to the virus' natural reservoir(NR) than Y and Z".*

In conclusion, we predict from our theoretical findings here that Marburg may be better adapted to use human and Rhesus Macaca and human cells for tropism than Ebola. However, Rhesus macaca may actually be more susceptible to both Marburg and Ebola tropism (and associated VHF) than man. The Rhesus macaca are probably closer in the evolutionary tree to the natural reservoir of filoviruses than man.

Conflict of Interests: None

Acknowledgment: We thank, for technical advice: Dr Olobo James, Associate Professor of Microbiology, Department of Microbiolobgy-Makerere University faculty of Medicine, Dr Kaddu Mulindwa, and Head of Division of Molecular Biology- Makerere University faculty of Medicine, Uganda

References

- Altschul, S. F. (1993). A protein alignment scoring system sensitive at all evolutionary distances. J. Mol. Evol. 36:290-300.
- Anandalakshmi, R., Marathe, R., Ge, X., Herr, J.M., Jr., Mau, C., Mallory, A., Pruss, G., Bowman, L., and Vance, V.B. 2000. A calmodulin-related protein that suppresses posttranscriptional gene silencing in plants. *Science* 290: 142-144
- Baize, S., Leroy, E. M., Georges-Courbot, M.-C., Capron, M., Lansoud-Soukate, J., Debré, P., Fisher-Hoch, S. P., McCormick, J. B. & Georges, A. J. 1999. Defective humoral responses and extensive intravascular apoptosis are associated with fatal outcome in Ebola virus-infected patients. *Nature Medicine.* 5: 423-426.
- Bausch, D. G. , Nichol, S. T. , Muyembe-Tamfum, J. J., Tugume, B., Colebunders,R KondÃ, K.M.,et al.. 2006. Marburg hemorrhagic fever associated with multiple genetic lineages of virus. N. Engl. J. Med., August 31; 355(9): 909 - 919.
- **Bor, Y.-C., F. Bushman, and L. Orgel**. In vitro integration of human immunodeficiency virus type 1 cDNA into targets containing protein-induced bends. Proc. Natl. Acad. Sci. USA 1995; **92:**10334-10338
- Bowen, E. T., Platt, G. S., Lloyd, G., Raymond, R. T. & Simpson, D. I. 1980. A comparative study of strains of Ebola virus isolated from southern Sudan and northern Zaire in 1976. *Journal of Medical Virology.* 6: 129-138.
- **Bushman, F. D**. Tethering human immunodeficiency virus 1 integrase to a DNA site directs integration to nearby sequences. Proc. Natl. Acad. Sci. USA 1994;**91:**9233-9237
- **Bushman, F. D., and R. Craigie**. Integration of human immunodeficiency virus DNA: adduct interference analysis of required DNA sites. Proc. Natl. Acad. Sci. USA 1992; **89:**3458-3462
- **Carteau S**, **Hoffmann C**, **Bushman F**. Chromosome structure and human immunodeficiency virus type 1 cDNA integration: centromeric alphoid repeats are a disfavored target. J Virol. 1998;72(5):4005-14.
- Dalmay, T., Hamilton, A., Rudd, S., Angell, S., and Baulcombe, D.C. 2000. An RNA-dependent RNA polymerase gene in *Arabidopsis* is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. *Cell* 101: 543-553
- Dowell, S. F., Mukunu, R., Ksiazek, T. G., Khan, A. S., Rollin, P. E. & Peters, C. J. 1995. Transmission of Ebola hemorrhagic fever: a study of risk factors in family members, Kikwit, Democratic Republic of the Congo. *Journal of Infectious Diseases.*1999;179 (Suppl. 1): S87–S91
- Feldmann, H., Klenk, H. D., & Sanchez, A. 1993. Molecular biology and evolution of filoviruses. *Archives of Virology.* 7: 81-100.
- Felsenstein, J. 1993. PHYLIP: Phylogeny Inference Package; version 3.5c. University of Washington, Seattle, WA, USA.
- **Howard, M. T., and J. D. Griffith**. A cluster of strong topoisomerase II cleavage sites is located near an integrated human immunodeficiency virus. J. Mol. Biol. 1993;**232:**1060-1068
- Jensen, S., Gassama, M.P., and Heidmann, T. 1999. Taming of transposable elements by homology-dependent gene silencing. *Nat. Genet.* 21: 209-212
- Ketting, R.F., Haverkamp, T.H., van Luenen, H.G., and Plasterk, R.H. 1999. Mut-7 of *C. elegans*, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. *Cell* 99: 133-141

- Khan, A. S., Tshioko, F. K., Heymann, D. L., Le Guenno, B. Nabeth p. ; Kerstiens b., et al. 1993. The reemergence of Ebola hemorrhagic fever, Democratic Republic of the Congo, 1995. *Journal of Infectious Diseases.* 179 (Suppl. 1):S76–S86.
- Leroy, E. M., Baize, S., Volchkov, V. E., Fisher-Hoch, S. P., Georges-Courbot, M.-C., Lansoud-Soukate, J., Capron, M., Debré, P., McCormick, J. B. & Georges, A. J. 2000. Human asymptomatic Ebola infection and strong inflammatory response. *Lancet.* 355: 2210-2215.
- Lucy, A.P., Guo, H.S., Li, W.X., and Ding, S.W. 2000. Suppression of post-transcriptional gene silencing by a plant viral protein localized in the nucleus. *EMBO J.* 19: 1672-1680
- Malinsky, S., Bucheton, A., and Busseau, I. 2000. New insights on homology-dependent silencing of I factor activity by transgenes containing ORF1 in *Drosophila melanogaster. Genetics* 156: 1147-1155
- Mayo, M. A. & Pringle, C. R. 1998. Virus taxonomy – 1997. *Journal of General Virology.* 79: 649-657.
- **Milot, E., A. Belmaaza, E. Rassart, and P. Chartrand**. Association of a host DNA structure with retroviral integration sites in chromosomal DNA. Virology 1994;**201:**408-412
- **Moalic Y**, **Blanchard Y**, **Félix H**, **Jestin A**. Porcine endogenous retrovirus integration sites in the human genome: features in common with those of murine leukemia virus. J Virol. . 2006 ; 80(22):10980-10988.
- **Muller, H.-P., and H. E. Varmus.**. DNA bending creates favored sites for retroviral integration: an explanation for preferred insertion sites in nucleosomes. EMBO J. 1994; **13:**4704-4714
- Peterson, A.T., Carroll, D.S., Mills, J.N., Johnson, K.M. 2004. Potential of mammalian filovirus reservoirs. EID 10(2):2073-2081
- **Pruss, D., F. D. Bushman, and A. P. Wolffe**. 1994. Human immunodeficiency virus integrase directs integration to sites of severe DNA distortion within the nucleosome core. Proc. Natl. Acad. Sci. USA **91:**5913-5917
- **Pruss, D., R. Reeves, F. D. Bushman, and A. P. Wolffe**. 1994. The influence of DNA and nucleosome structure on integration events directed by HIV integrase. J. Biol. Chem. **269:**25031-25041
- **Pryciak, P., H.-P. Muller, and H. E. Varmus**. Simian virus 40 minichromosomes as targets for retroviral integration in vivo. Proc. Natl. Acad. Sci. USA 1992; **89:**9237-9241
- **Pryciak, P. M., A. Sil, and H. E. Varmus.** Retroviral integration into minichromosomes in vitro. EMBO J. 1992; **11:**291-303
- Ratcliff, F.G., MacFarlane, S.A., and Baulcombe, D.C. 1999. Gene silencing without DNA: RNA-mediated cross-protection between viruses. *Plant Cell* 11: 1207-1216
- Sanchez, A., Kiley, M. P., Holloway, B. P. & Auperin, D. D. 1993. Sequence analysis of the Ebola virus genome: organization, genetic elements, and comparison with the genome of Marburg virus. *Virus Research.*1993;29: 215-240
- **Scherdin, U., K. Rhodes, and M. Breindl**. Transcriptionally active genome regions are preferred targets for retrovirus integration. J. Virol. 1990; **64:**907-912
- **Scottoline, B. P., S. Chow, V. Ellison, and P. O. Brown**. Disruption of the terminal base pairs of retroviral DNA during integration. Genes Dev. 1997; **11:**371-382
- **Sels, F. T., S. Langer, A. S. Schulz, J. Silver, M. Sitbon, and R. W. Friedrich**. Friend murine leukaemia virus is integrated at a common site in most primary spleen tumours of erythroleukaemic animals. Oncogene 1992;**7:**643-652
- **Stevens, S. W., and J. D. Griffith.** 1994. Human immunodeficiency virus type 1 may preferentially integrate into chromatin occupied by L1Hs repetitive elements. Proc. Natl. Acad. Sci. USA **91:**5557-5561
- **Stevens, S. W., and J. D. Griffith**. 1996. Sequence analysis of the human DNA flanking sites of human immunodeficiency virus type 1 integration. J. Virol. **70:**6459-6462

- Tabara, H., Sarkissian, M., Kelly, W.G., Fleenor, J., Grishok, A., Timmons, L., Fire, A., and Mello, C.C. 1999. The rde-1 gene, RNA interference, and transposon silencing in *C. elegans. Cell* 99: 123-132
- Towner, J. S. , Khristova, M. L., Sealy, T. K. , Erickson, B. R. , Bawiec, D., **Hartman**, **A. L. ,**et al.. 2006. Marburg virus genomics and association with a large hemorrhagic Fever outbreak in angola. J. Virol. 80(13): 6497 - 6516.
- **Varmus, H. E., and Brown, P. O.** 1989. Retroviruses, p. 53-108. *In* D. E. Berg, and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.
- **Vijaya, S., D. L. Steffan, and H. L. Robinson**. Acceptor sites for retroviral integrations map near DNase I-hypersensitive sites in chromatin. J. Virol. 1986; **60:**683-692
- Voinnet, O., Lederer, C., and Baulcombe, D.C. 2000. A viral movement protein prevents spread of the gene silencing signal in *Nicotiana benthamiana. Cell* 103: 157-167
- Watanabe, S., Takada, A., Watanabe, T., Ito, H., Kida, H. & Kawaoka, Y. 2000. Functional Importance of the coiled-coil of the Ebola virus glycoprotein. *Journal of Virology.* 74: 10194-10201.
- **Withers-Ward, E. S., Y. Kitamura, J. P. Barnes, and J. M. Coffin.**. Distribution of targets for avian retrovirus DNA integration in vivo. Genes Dev. 1994;**8:**1473-1487
- Yang, Z.-Y., Delgado, R., Xu, L., Todd, R. F., Nabel, E. G., Sanchez, A. & Nabel, G. J. 1998. Distinct cellular interactions of secreted and transmembrane Ebola virus glycoproteins. *Science.* 279: 1034-1037.

Supplement

The evolutionary adaptation cross species "theorem" of viral cross-species jumps: A succinct of the key ecological determinants of the evolutionary trends among emerging viral pathogens

Misaki Wayengera *

Address all correspondences and requests for materials to W.M (wmisaki@yahoo.com).

Viruses are obligate intracellular organisms requiring integration into host cells to replicate and survive. Occasionally, depending on the innate viral adaptability to survive in a new host, or evolutionary trends resulting from mutations, re-assortment or recombination induced by currently ill defined abiotic and biotic ecological determinants, viruses cross from one species to another. Within the new host with naïve immunity, most of these crossovers are associated with viral disease whose propagation and sustainable establishment is dependant at how well the virus is suited to survive. Herein the" evolutionary adaptation cross species theorem, we summarize; using various virus species as examples- the major ecological dynamics of viral crossovers, as well as highlights of their implications on the control of major viral infectious disease outbreaks.

Introduction

The emergence of novel viruses and re-emergence of mutant strains of already described ones has posed some of the worst threat to global human health[1-3]. Viruses are obligate intracellular organisms requiring integration into host cells to replicate and survive. Occasionally, depending on the innate viral adaptability to survive in a new host, or evolutionary trends resulting from mutations, re-assortment or recombination, viruses cross from one species to another. Today, several viruses like HIV1&2, Influenza A,B &C, Ebola & Marburg[1-21], have their zoonotic origins well documented. Despite this, and the wide spectrum of emerging and re-emerging viral infections among humans, the exact ecological determinants of each specific viral cross species jumps are still not well defined, with only a few studies focusing primarily on availing data in this area. Yet understanding of these determinants is central in the control of future outbreaks, say by limiting viral evolutions.

In the past, we have studied some of the possible exogenous ecological determinants of individual virus reassortment and mutations favoring cross species crossovers by elucidating the possible exogenous role of commensal bacteria (such as H. Influenza) retroelement(RTE) derived small noncoding ribonucleic acids (SnRNA) as contributing factors to Influenza genomic diversity(Misaki, W unpublished). We have as well attempted, basing on the hypothesis that RNA viral integration occurs by Target primed reverse transcription as group II mobile retroelements, attempted to develop an Insilco model for assessing reservoir host adaptability for various virus tropism by way of genome wide modelling Insilco approaches. Here, we present for the first time as a full manuscript

documentation of "the evolutionary adaptation cross species theorem" developed over the past 3-5 years using various virus species models and scenarios to summarize the dynamics of viral crossovers, highlighting their implications on the control of major viral infectious disease outbreaks[22, 23].

Postulates of the Evolutionary adaptation cross species "theorem" of viral cross-species jumps

The Evolutionary adaptation cross species (EACS) "theorem" is a succinct summary of the key ecological determinants of the evolutionary trends among emerging viral pathogens. Initially developed following the 1999-2000 outbreaks of Ebola in Uganda, the theorem is applicable to other human emerging and re-emerging viruses. Its main rationale is to summarize; using various virus species crossovers as evidence- the possible major ecological determinants of viral crossovers. The theorem comprises 4 postulates expounded below:

P1. Emerging and re-emerging viral infections have been in existence even before they were isolated or shown to cause disease in humans. They existed in previous hosts called "natural and reservoir" hosts on which they depended for the virus-host cell interaction necessary for viral replication.

The zoonotic origin of viruses is supported by various studies that have mapped the emergence of several known human viruses to contact with lower mammalian or non primate veretebrates. For Instance, in 1981, the centers for disease control described the first clinical case of human infection with the Human immunodeficiency virus (HIV) in Los Angels[1], and 20 years later today, over 40 million worldwide people are living with HIV, with 16,000 acquiring the virus daily[2]., Despite the advancements in the medical understanding of HIV, and the development of several therapeutic interventions, HIV is still incurable, and vaccine research is but been rendered unfruitful[3-5]. It is now clear that HIV originated from nonhuman primates. In 1997, H5N1 was first recognized as a cause of human disease in Hong Kong in connection with a poultry outbreak in live-bird markets. A total of 18 human cases occurred with a 33% mortality. After a quiet period, the virus re-emerged in 2003.[6] Since then, the highly pathogenic H5N1 virus has caused unprecedented outbreaks in poultry stocks throughout Asia.[7] These outbreaks are in fact the largest and most severe outbreaks in domestic poultry in recorded history. In addition, the detection of virus in migratory birds, and the recent spread of the virus to Europe suggest that migratory birds are directly spreading highly pathogenic H5N1 virus. The outbreak of avian influenza (H5N1) through South East Asia and its progressive spread westward have resulted in increasing concern about the threat of a new worldwide influenza pandemic[8-15, 24-31]. Filoviruses form another re-emerging class of viruses of significance to global human health with major epicentres in equatorial Africa. The zoonotic origins are so evident, yet the natural reservoir remains still elusive Ebola viruses (EBOV) and Marburg viruses form the major two geneses of a class of negative-stranded RNA virus family called Filoviridae[16, 17]. EBOV and Marburg virus (MBGV) are members of the family *Filoviridae*, order *Mononegavirales*, which groups together nonsegmented, negative-stranded RNA viruses, namely filoviruses, paramyxoviruses, rhabdoviruses and bornaviruses[18, 19]. Ebola virus (EBOV) is one of the most virulent human pathogens, killing up to 70–80% of patients within 5–7 days [20]. EBOV causes rare but fulminating outbreaks of haemorrhagic fever in equatorial Africa, being transmitted from

person to person via infected body fluids, such as faeces, vomit and blood [21]. The exact natural reservoir of filoviridae viruses is thus still debatable, with recent studies showing the African Fruit bat species Rousettus aegyptiacus as the possible naturally occurring host that does not exhibit disease (Peterson, et al 2004).Other viruses of demonstrated zoonotic origin of significance to Human health include:, Human T-cell lymphotrophic viruses (HTLV) 1-4 [33](from NHP), SARS Corona virus[34](from civet rats), and porcine equine retrovirus (PERV) [35](from pigs).

P2. With an increasing change in variables among the previous hosts and environment (stability of the virus and chance of contact with new host) there is adaptation of these viruses (not necessary) through mutations, recombination and re-assortment to yield new strains able to use new host (human) cells for replication.

Despite many epidemiologic studies for virus sero-prevalence among reservoir hosts, the ecological determinants of viral evolution and adaptability necessary to effect crossovers that are sustainable in humans still remain vaguely defined for several viruses inclusive of HIV & Ebola. Studies of Influenza viruses provide us with some, though still incomplete, insights of the possible determinants of influenza evolution. Within any given micro-environment, these factors can be classified into ex-vivo (external to) and in-vivo (within or internal) to the virus of interest. These then can further be broken into abiotic(non-living) or biotic(living). Some viruses already exhibit innate fitness to successfully integrate in other hosts and replicate without necessary having to undergo any mutations or recombination. On the other hand, viruses like influenza must adapt to acquire fitness to survive in human cells. Often, no single factor can offset an ecosystem into imbalance, but rather interaction of these factors is favored as the cause of change leading to viral adaptation. Until recently however, when we used RNomics(bacteria derived REase palindromics and Group II RTE RT ORF genome wide searches model)
to investigate the possible role of an avian and swine GI tract commensal derived snRNA as exogenous facilitatators of Influenza genomic variation,[36, 37], no single factor exogenous to the influenza virus had been described as a possible facilitator of these antigenic drifts and shifts, with much of the occurrences being attributed to the in-vivo evidence that viruses reproducing by RNA polymerase generally exhibit higher mutations than their DNA polymerase

counterparts. Generally, many viruses with RNA genomes have genetically diverse populations called quasispecies. The representation of any particular sequence within this quasispecies is a result of interactions between the host and environmental factors affecting the replication of the virus. Important biological properties are a direct result of the levels of diversity in the quasispecies 'cloud size', including adaptability and host range. RNA viruses have become the model system for the analysis of viral evolution due to the inherent error-prone nature of their genome-replicating enzymes that lack a proof-reading function[38]. The rationale for our study was in the findings of closely contested mutation in small DNA viruses[39].

P3. The most susceptible new hosts are those whose cellular biochemical environments and genetics favor viral establishment by coding for and producing the energy, metabolites and at most, in some cases all the enzymes required for replication of the adapted new strain of virus.

On the metabolomic level, sinces viruses are obligate intracellular parasites, the sustaianable host ought to be able to encode the energy, metabolites and at most, in some cases all the enzymes required for replication of the adapted new strain of virus. There must exist a certain level of homology between the metabolomics of a natural reservoir and new hosts for effective and sustainable cross over to occur. Depending on the endogenous tissue specificity exhibited by a virus however, viral crossovers can successfully occur between host species of variable biochemical and genetic homology.

On the genomics level however, for many viruses, the exact host integration acceptor sites are still ill defined, most not expected to be present as naked DNA but rather associated with histones and other DNA-binding proteins in chromatin. DNA packaging in vivo is expected to influence integration site selection, and the choice of integration site may have profound effects on both the virus and the host.[40] The determinants of integration efficiency in vivo thus remain incompletely defined, despite their predicted importance in defining targets for future therapeutic and vaccine strategies, pathogenesis elucidation, as well as a modeling the evolutionary trends of successful viral cross over[41]. In the past, such in-vivo and in-vitro surveys of integration sites for retroviruses have led to several proposals for factors influencing site selection. In-vivo studies of Moloney murine leukemia virus have supported a model in which open chromatin regions at transcription units were favored, since associated features such as DNase I-hypersensitive sites [42,43] or CpG islands [44] were apparently enriched near integration sites. Another study proposed that unusual host DNA structures were common near integration sites [45]. A study of avian leukosis virus integration frequencies at several chromosomal sites failed to show any major differences among the regions studied [46], contrary to an earlier report [39]. For human immunodeficiency virus type 1 (HIV-1), it has been proposed that integration may be favored near repetitive elements inclusive of LINE-1 elements [47] or Alu islands [48] or topoisomerase cleavage sites [49]. On the other hand, assays of integration in vitro have revealed several effects of proteins bound to target DNA. Simple DNA-binding proteins can block access of integration complexes to target DNA, creating regions refractory for integration [42, 50, 51]. In contrast, wrapping DNA on nucleosomes can create hot spots for integration at sites of probable DNA distortion [42, 52-54]. Distortion of DNA in several other protein-DNA complexes can also favor integration [55] consistent with the possibility that DNA distortion is involved in the integrase mechanism[56, 57].Some studies using HIV have also demonstrated the absence of integration in vivo into centromeric alphoid repeats, with alphoid repeats being absent in integration site sequences but present in controls, and alphoid sequences being selectively disfavored in the repeat-specific PCR integration assay; thus providing an demonstration of possibility that certain types of chromatin may obstruct cDNA integration[58]. More recent data on the integration of Porcine endogenous retroviruses (PERV) within pig/porcine tissues showed that PERV integration was strongly enhanced at transcriptional start sites and CpG islands and that the frequencies of integration events increased with the expression levels of the genes, except for the genes with the highest levels of expression, which were disfavored for integration[35].

Generally, target primed reverse transcription(TPRT) has been proposed as a possible mechanism for RNA viral integration following finding that RNA viruses integration seems to be favored around ancestral elements of Group II RTE-LINE(long interspersed nuclear elements) or non-LTR(long terminal repeats). If, as widely believed, RNA viruses integrated by TPRT, then historical events of RTE activity may be favoring viral integration. There may thus exist a certain level of homology on the genetic level between the viruses and gene interspersing ancestors of these Group II RTE- spliceosomes. Basing on this school of thought, we hypothesized that, since the spectrum of host to which viral crossover can occur is so wide, it is not wide to use a host to host similarity approach to study the likely host. Rather, by attempting to analayse the intragenomic homology of interspesing elements to viral genetic material acroos various hosts, we may be able to assess the evolutionary adaptability for viral integration acroos these hosts. Using this approach, we have developed, as described here, Comparative Insilco genography as an insilico model for mapping cross species adaptability to acts as sustainable hosts to viral tropism.

By using this approach to for instance map integration sites for Ebola and Marburg within the Rhesus macaca mulatto genome, it was found that Marburg GP genomic RNA had 18 alignments located on undefined scaffolds compared to 7 of Ebola located on chromosomes 4, 6, 7, 8, 9, 14 & 15. Our results serve to demonstrate that although Marburg GP RNA acceptors are more prevalent in the Rhesus genome than ebola; their loci of integration are vaguely defined compared to Ebola. If the level of homogeneity between acceptors and PIC has no effect of integration, then Marburg may be better adapted to integrate into Rhesus that Ebola. Alternatively,, chromatic DNA might be a more effective target for future Ebola genomic vaccines sequestered at a nuclear location inaccessible to incoming Pre-integration Complexes (PICs- which in this model are Ebola glycoprotein gene complexes) than Marburg.

P4. Various mechanisms of interaction between previous and new (human) hosts are required to affect the viral crossover.

These may be direct (as occurs via contact, consumption of meat for ebola VHF or inhalation of respiratory droplets for Avian flue/SARS) OR Indirect (as occurs through vectors like arthropods for arboviruses and amplifier - link hosts like pigs/birds for Avian flue /influenza). Cross species Index cases of human infection with most of the viruses discussed above have been linked with contact with reservoir host. Pandemic influenza caused by the H5N1 was linked to consumption and close droplet aerosol contact with infected avian species[8-15, 24-31].SARS was linked to consumption of civet rats, a practice that until these outbreaks was a common in some provinces of China[34]. Index cases of most Ebola VHF outbreaks have all been successfully traced back contact with an infected NHP or consumption of its game meat[20]. Transplantation (xenotransplanataion) provides one of the mechanisms by which Porcine equine retrovirus (PERV) may become iatrogenic ally introduced into humasn[35]. Shortage of donors has forced medical research to attempt transplantation with cadeveric tissue from pigs (heart valves, whole hearts), bringing the risk of PERV in these transplant patients to reality.

Implications of the postulates of the Evolutionary adaptation cross species "theorem"

Of particular global health siginificance are the deductions to be drawn from these postulations. There are major implications of these postulates (IOPs) on the epidemiological control and prevention of most emerging and re-emerging human viruses, as well as understanding their evolutionary trends and determinants.

I0P.1. All emerging viruses are zoonotic, with a natural reservoir a more likely source of the new virus strain than just spontaneous evolution of a new viral entity. This implies that there are to date various viruses out there in veterinary reservoirs which have not been isolated or shown to cause disease in man, but are potential zoonoses to man. Wild ecologic populations are thus an unexplored source of information regarding viral infectious diseases and criteria to link human and veterinary programs is necessary. Experience with veterinary retroviral and other viral vaccines may be applicable in human situations to control, prevent - let alone develop human vaccines. [24-30], [31]. [32]. [33].

I0P.2. **The control of viral zoonoses requires knowledge of the various variables and how they interplay to increase incidence of zoonotic viral emergence and re-emergence.** There is thus need to strengthen epidemiologic surveillance and concentrate public health activities in epicenters of new emergent viral infections to study these variables - South East Asia for Avian flue/SARS and Gabon & DRC for ebola VHF. While the WHO has increased surveillance for Avian Influenza and SARS in the far east, these efforts may need to be reciprocated by the creation of a VHF surveillance centre(VHFSC) for other viral emergences such as the filoviriade VHF in East, Central and Western Africa [34]. Such a VHFSC Will serve to [1] Conduct continued surveillance for Ebola and Marburg VHF [2] Assess the reservoir prevalence of Marburg and Ebola viruses in NHPs here [3] Avail level IV laboratory diagnostics in case of an outbreak [4] Set in motion research, development and trial of Veterinary and human vaccine for Ebola and Marburg VHFs since these reservoir hosts form a major model for the development of human vaccines as been demonstrated for Ebola[35] [5] Offer training for Health care professionals and policy makers on outbreak handling[6] serve as a WHO/CDC minute watch dog for VHFs outbreaks.

IOP.3. **Innate viral specificity for a particular cell/tissue type remaining a constant, wherever successful viral crossovers have been noticed between different species, these crossovers can play a role in the in the study of evolutionary trends**. Thus restriction of contact between man and his close phylogenetic relatives like NHPs, and infected avian species, civet rats known to be sources of viral zoonoses can play a major role in control of emergent viral infections. Studies of integration sites within reservoir host genomes, and human genomes may be a model for understanding crossovers, as well as identifying targets for future vaccines and therapy. By way of Inslico genome wide approaches, it may be possible to evaluate which host may be likely source of human viruses to man.

IOP.4. **minimizing interactions that increase risk of zoonotic viral crossovers between identified previous hosts and man can greatly reduce incidence of emergent viral infections.** Restricting human contact/consumption (of) with caucus of previous hosts for ebola VHF (NHPs), infected bird for avian flue & civet rats for SARS can play the trick. Respiratory droplet inhalation in the later cases may be prevented by wearing masks as well for Laboratory scientisits, as well as observing level 4 infection prevention protocols. Arboviruses on the other hand, which have sometime now not paused any much danger to global Health over the recents years may be prevented by vector control-spraying with insecticides and use of Insecticide Treated Nets(ITNs). Xeno-transplantation with amplifier-link host tissue (porcine) should carefully be watched histo-immunologically, if not avoided to prevent xenosis from PERV[35].

In conclusion, a deeper understanding of the ecological determinants of viral crossovers, as well as highlighted implications is necessary for the control of major viral infectious disease outbreaks.

Competing Interests: None

References

1. CDC. Pneumocystis pneumonia- Los Angeles. MMWR 1981; 30:250-252
2. UNAIDS. Report on the global HIV-AIDS epidemic July 2006.
3. Bartlett JA, DeMasi R, Quinn J, Moxham C, Rousseau F. Overview of the effectiveness of triple combination therapy in antiretroviral-naive HIV-1 infected adults. AIDS. 2001; 15:1369-77.
4. Bartlett, John A a; Fath, Michael J b; DeMasi, Ralph c; Hermes, Ashwaq b; Quinn, Joseph d; Mondou, Elsa d; Rousseau, Franck d An updated systematic overview of triple combination therapy in antiretroviral-naive HIV-infected adults. AIDS. 20(16):2051-2064, October 24, 2006.
5. Hammer SM, Saag MS, Schechter M, Montaner JS, Schooley RT, Jacobsen DM, Thompson MA, Carpenter CC, Fischl MA, Gazzard BG, Gatell JM, Hirsch MS, Katzenstein DA, Richman DD, Vella S, Yeni PG, Volberding PA; International AIDS Society-USA panel. "Treatment for adult HIV infection: 2006 recommendations of the International AIDS Society-USA panel." *JAMA*. 2006 Aug 16;296(7):827-43.
6. Yuen KY, Chan PK, Peiris M *et al.* (1998)Clinical features and rapid viral diagnosis of human disease associated with avian influenza A H5N1 virus. *Lancet*; 351: 467–471.
7. Ungchusak K, Auewarakul P, Dowell SF *et al.* (2005) Probable person-to-person transmission of avian influenza A (H5N1). *N Engl J Med*; 352: 333–340.
8. Tran TH, Nguyen TL, Nguyen TD *et al.* (2004) Avian influenza A (H5N1) in 10 patients in Vietnam. *N Engl J Med*; 350: 1179–1188.
9. Chotpitayasunondh T, Ungchusak K, Hanshaoworakul W *et al (*2005*).* Human disease from influenza A (H5N1), Thailand, 2004. *Emerg Infect Dis*; 11: 201–209.
10. Cheung CY, Poon LL, Lau AS *et al.* (2002) Induction of proinflammatory cytokines in human macrophages by influenza A (H5N1) viruses: A mechanism for the unusual severity of human disease? *Lancet*; 360: 1831–1837.
11. Thorson A, Petzold M, Nguyen TK, Ekdahl K. (2006) Is exposure to sick or dead poultry associated with flulike illness?: A population-based study from a rural area in Vietnam with outbreaks of highly pathogenic avian influenza. *Arch Intern Med*; 166: 119–123.
12. Horimoto T, Fukuda N, Iwatsuki-Horimoto K *et al.* (2004) Antigenic differences between H5N1 human influenza viruses isolated in 1997 and 2003. *J Vet Med Sci*; 66: 303–305.
13. Kuiken T, Rimmelzwaan G, van Riel D *et al.* Avian H5N1 influenza in cats. *Science* 2004; 306: 241–241.
14. Cox NJ, Subbarao K. (2000) Global epidemiology of influenza: past and present. Annu Rev Med.;51:407-21. Review. PMID: 10774473 [PubMed - indexed for MEDLINE]
15. Smith CB, Cox NJ, Subbarao K, Taber LH, Glezen WP. (2002) Molecular epidemiology of influenza A(H3N2) virus reinfections.J Infect Dis. Apr 1;185(7):980-5. Epub 2002 Mar 13. PMID: 11920323 [PubMed - indexed for MEDLINE]
16. D. G. Bausch, S. T. Nichol, J. J. Muyembe-Tamfum,et al. Marburg hemorrhagic fever associated with multiple genetic lineages of virus. N. Engl. J. Med., August 31, 2006; 355(9): 909 - 919.
17. J. S. Towner, M. L. Khristova, T. K. Sealy,et al. Marburgvirus genomics and association with a large hemorrhagic Fever outbreak in angola. J. Virol., July 1, 2006; 80(13): 6497 - 6516.
18. Mayo, M. A. & Pringle, C. R. Virus taxonomy – 1997. *Journal of General Virology* 1998; 79: 649-657.
19. Sanchez, A., Kiley, M. P., Holloway, B. P. & Auperin, D. D. Sequence analysis of the Ebola virus genome: organization, genetic elements, and comparison with the genome of Marburg virus. *Virus Research* 1993;29: 215-240
20. Khan, A. S., Tshioko, F. K., Heymann, D. L., Le Guenno, B., et al. The reemergence of Ebola hemorrhagic fever, Democratic Republic of the Congo, 1995. *Journal of Infectious Diseases* 1997; 179 (Suppl. 1):S76–S86.
21. Dowell, S. F., Mukunu, R., Ksiazek, T. G., Khan, A. S., Rollin, P. E. & Peters, C. J. . Transmission of Ebola hemorrhagic fever: a study of risk factors in family members, Kikwit, Democratic Republic of the Congo, 1995. *Journal of Infectious Diseases* 1999;179 (Suppl. 1): S87–S91

22. Wayengera Misaki. Emergence of zoonotic viral infections in this era.why? Abstract book and programme of the international students conference on emergencies(ISCE) 11-14th Dec,2001:7
23. Wayengera Misaki. The evolutionary adaptation Hypothesis to explain the origin of ebola VHF. Makerere Medical Journal(MMJ)2002;36:36-37
24. **Apetrei, C., and P. A. Marx.** 2004. Simian retroviral infections in human beings. Lancet **364:**137–138.
25. **Apetrei, C., D. L. Robertson, and P. A. Marx.** 2004. The history of SIVs and AIDS: epidemiology, phylogeny and biology of isolates from naturally SIV infected non-human primates (NHP) in Africa. Frontiers Biosci. **9:**225–254.
26. **Bailes, E., F. Gao, F. Bibollet-Ruce, M. Peeters, P. A. Marx, B. H. Hahn, and P. M. Sharp.** 2003. Hybrid origin of SIV in chimpanzees. Science **300:**1713.
27. **Drucker, E., P. G. Alcabes, and P. A. Marx.** 2001. The injection century: consequences of massive unsterile injecting for the emergence of human pathogens. Lancet **358:**1989.
28. **Hahn, B. H., G. M. Shaw, K. M. DeCock, and P. M. Sharp.** 2000. AIDS as a zoonosis: scientific and public health implications. Science **287:**607–614.
29. **Krauss, H., A. Weber, M. Appel, B. Enders, H. D. Isenberg, H. G. Schiefer, W. Slenczka, A. von Graevenitz, and H. Zahner.** 2003. Zoonoses: infectious diseases transmissible from animals to humans. ASM Press, Washington, D.C.
30. Kaye D, Pringle CR. (2005) Avian influenza viruses and their implication for human health. *Clin Infect Dis*; 40: 108–112.
31. Monto AS. (2005)The threat of an avian influenza pandemic. *N Engl J Med*; 352: 323–325.
32. E. M. Leroy, P. Rouquet, P. Formenty, S. Souquiere,et al Multiple Ebola Virus Transmission Events and Rapid Decline of Central African Wildlife Science, January 16, 2004; 303(5656): 387 - 390.
33. Wolfe ND, Heneine W, Carr JK, Garcia AD, Shanmugam V, Tamoufe U, Torimiro JN, Prosser AT, Lebreton M, Mpoudi-Ngole E, et al Emergence of unique primate T-lymphotropic viruses among central African bushmeat hunters. Proc Natl Acad Sci U S A. 2005 May 31;102(22):7994-9. Epub 2005 May 23.
34. Fiona Fleck .Experts urge action to stop animal diseases infecting humans BMJ 2004; 328:1158
35. **Moalic Y**, **Blanchard Y**, **Félix H**, **Jestin A**. 2006. Porcine endogenous retrovirus integration sites in the human genome: features in common with those of murine leukemia virus. J Virol. Nov; 80(22):10980-8. Epub 2006 Aug 23.
36. Wayengera Misaki. Influenza ha and na antigenic drifts and shifts: what may be the role of h.influenza 2007 RLE
37. Wayengera., M . Heamophilus spp derived restriction enzymes: a synergy to antigenic drifts and shifts in influenza or antimicrobial defense peptides?2007 RLE
38. *A. R. Gould* Virus evolution: disease emergence and spread *Australian Journal of Experimental Agriculture2004 ;* 44(11) :1085–1094
39. Laura A. Shackelton**,** Colin R. Parrish, Uwe Truyen, and Edward C. Holmes (**2005)** High rate of viral evolution associated with the emergence of carnivore parvovirus **PNAS | January 11,; 102(2): 379-384**
40. **Coffin, J. M.** 1996. Retroviridae: the viruses and their replication, p. 1767-1848. *In* B. N. Fields, D. M. Knipe, and R. M. Howley (ed.), Virology. Lippincott-Raven Publishers, Philadelphia, Pa.
41. **Varmus, H. E., and P. O. Brown.** 1989. Retroviruses, p. 53-108. *In* D. E. Berg, and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.
42. **Pryciak, P. M., A. Sil, and H. E. Varmus.** 1992. Retroviral integration into minichromosomes in vitro. EMBO J. **11:**291-303
43. **Vijaya, S., D. L. Steffan, and H. L. Robinson.** 1986. Acceptor sites for retroviral integrations map near DNase I-hypersensitive sites in chromatin. J. Virol. **60:**683-692
44. **Scherdin, U., K. Rhodes, and M. Breindl.** 1990. Transcriptionally active genome regions are preferred targets for retrovirus integration. J. Virol. **64:**907-912
45. **Milot, E., A. Belmaaza, E. Rassart, and P. Chartrand.** 1994. Association of a host DNA structure with retroviral integration sites in chromosomal DNA. Virology **201:**408-412

46. **Withers-Ward, E. S., Y. Kitamura, J. P. Barnes, and J. M. Coffin.** 1994. Distribution of targets for avian retrovirus DNA integration in vivo. Genes Dev. **8:**1473-1487
47. **Sels, F. T., S. Langer, A. S. Schulz, J. Silver, M. Sitbon, and R. W. Friedrich.** 1992. Friend murine leukaemia virus is integrated at a common site in most primary spleen tumours of erythroleukaemic animals. Oncogene **7:**643-652
48. **Stevens, S. W., and J. D. Griffith.** 1994. Human immunodeficiency virus type 1 may preferentially integrate into chromatin occupied by L1Hs repetitive elements. Proc. Natl. Acad. Sci. USA **91:**5557-5561
49. **Stevens, S. W., and J. D. Griffith.** 1996. Sequence analysis of the human DNA flanking sites of human immunodeficiency virus type 1 integration. J. Virol. **70:**6459-6462
50. **Howard, M. T., and J. D. Griffith.** 1993. A cluster of strong topoisomerase II cleavage sites is located near an integrated human immunodeficiency virus. J. Mol. Biol. **232:**1060-1068
51. **Bor, Y.-C., F. Bushman, and L. Orgel.** 1995. In vitro integration of human immunodeficiency virus type 1 cDNA into targets containing protein-induced bends. Proc. Natl. Acad. Sci. USA **92:**10334-10338
52. **Bushman, F. D.** 1994. Tethering human immunodeficiency virus 1 integrase to a DNA site directs integration to nearby sequences. Proc. Natl. Acad. Sci. USA **91:**9233-9237
53. **Pruss, D., F. D. Bushman, and A. P. Wolffe.** 1994. Human immunodeficiency virus integrase directs integration to sites of severe DNA distortion within the nucleosome core. Proc. Natl. Acad. Sci. USA **91:**5913-5917
54. **Pruss, D., R. Reeves, F. D. Bushman, and A. P. Wolffe.** 1994. The influence of DNA and nucleosome structure on integration events directed by HIV integrase. J. Biol. Chem. **269:**25031-25041
55. **Pryciak, P., H.-P. Muller, and H. E. Varmus.** 1992. Simian virus 40 minichromosomes as targets for retroviral integration in vivo. Proc. Natl. Acad. Sci. USA **89:**9237-9241
56. **Muller, H.-P., and H. E. Varmus.** 1994. DNA bending creates favored sites for retroviral integration: an explanation for preferred insertion sites in nucleosomes. EMBO J. **13:**4704-4714
57. **Bushman, F. D., and R. Craigie.** 1992. Integration of human immunodeficiency virus DNA: adduct interference analysis of required DNA sites. Proc. Natl. Acad. Sci. USA **89:**3458-3462
58. **Scottoline, B. P., S. Chow, V. Ellison, and P. O. Brown.** 1997. Disruption of the terminal base pairs of retroviral DNA during integration. Genes Dev. **11:**371-382
59. **Bhagirath Singh** CIHR Research: Taking Down Goliaths: New Vaccines May Spell the End for Ebola, Marburg and Lassa Virus Infections Healthcare Quarterly, 8(4) 2005: 20-22