African Journal of Biotechnology Vol. 8 (10), pp. 2125-2130, 18 May, 2009
Available online at http://www.academicjournals.org/AJB
ISSN 1684–5315 © 2009 Academic Journals

Full Length Research Paper

A model for mapping of Ebola and Marburg RNA integration sites in rhesus *Macaca mulatta* genome in silico: Ebola virus acceptors sites located on chromosomes 4, 6, 7, 8, 9, 14 and 15

Wayengera Misaki[1]*, Byarugaba Wilson[2], Kajjumbula Henry[3], J. Olobo[3], Kaddu Mulindwa[4]

[1]Faculty of Medicine, Makerere University, Uganda.
[2]Division of Human Genetics, Department of Pathology, Makerere University, Uganda.
[3]Department of Microbiology, Makerere University, Uganda.
[4]Division of Molecular Biology, Department of Microbiology, Makerere University, Uganda.

Accepted 25 September, 2007

Viral integration into the host genetic material is necessary for replication and survival, since viruses are obligate intracellular organisms. Understanding of the exact loci of integration may thus provide targets for future therapeutic and vaccine strategies, pathogenesis elucidation, as well as a model for the evolutionary trends of successful viral cross over. Although the exact natural reservoir for the filovirade family of viruses still remains elusive, most index cases in human outbreaks have been linked to contact with nonhuman primates (NHP). 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 genes of this family of viruses within NHP genomes, we queried both Ebola and Marburg Glycoprotein (GP) gene sequences against the whole genome of rhesus macaque using BLAST-N analysis. Of all the contigs length 2.87 Gb (2,863,665,185) bases in the genome of rhesus macaque, Marburg GP blast hits to rhesus genome nucleotide database were 6,451,736 compared to 4,012,901 for Ebola. 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 and 15. We also found an efficiency of 66.6% within Marburg GP alignments compared to 100% for 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.

Key words: Ebola, Marburg, *In-vivo* integration, rhesus macaca, line elements, Insilico genomics.

INTRODUCTION

All viruses are obligating intracellular organisms whose survival and replication depends on the successful integration into the host genome. For many viruses however, the exact host integration acceptor sites are ill defined, 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 (Coffin, 1996; Varmus and Brown, 1989). The determinants of integration efficiency *in vivo* thus remain incompletely defined, despite their predicted importance in defining targets for future therapeutic and

*Corresponding author. E-mail: wmisaki@yahoo.com.

Abbreviations: NHP, Non human primates; **EBOV**, Ebola virus; **MBGV**, Marburg virus; **GP**, Glycoprotein; **HGSC**, Human genome sequencing centre; **HSP**, High scoring Segment pair; **NHGRI**, National Human genome Research institute; **NIH**, National Institutes of Health; **NCBI**, National Centre for Biotechnology information; **PICs**, Pre integration complexes.

vaccine strategies, pathogenesis elucidation, as well as a modeling the evolutionary trends of successful viral cross over (Varmus and Brown, 1989).

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). 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 (Towner, 2006; Mayo and Pringle, 1998). 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., 1995; Dowell et al., 1995; Feldmann et al., 1993; Felsenstein, 1993; Baize, 1999; Leroy, 2000; Yang, 1998).

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, 2000) and to mediate entry into the host cell (Bowen, 1980). GP thus forms the model Pre-integration Complex (PIC) for this family of viruses. Although the exact natural reservoir of filovirade viruses is still debatable, most index cases of human infection with this family of viruses have been linked to contact with non human primates (NHP) (Wayengera, 2005). To this effect, NHPs genomes thus provide the only model for mapping of Filoviridae genomic RNA integration in their reservoir genomes.

We believe that proximity of homogeneity in nucleotide sequences between the host and the virus is a significant determinant of viral integration, since foreign DNA or RNA with the same homogeneity and alignment as a particular host's genomic constitution will more likely integrate into that host's genome relative to a host with distant homogeneity. Basing on this hypothesis, we present here results of the first ever attempt to map integration sites of filoviridae genomic RNA in the Rhesus macaca mulatta genome with a focus on (1) location of site of integration (2) efficiency of integration and (3) evolutionary, pathogenesis and future therapeutic strategy development implications for both Ebola and Marburg.

MATERIALS AND METHODS

Design

Homology Basic Local Alignment Survey

Materials

(1) 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 (Altschul, 1991; Altschul, 1993; Altschul et al., 1994; Altschul et al., 1990; Claverie and States, 1993; Gish and States, 1993; Henikoff and Henikoff, 1992; Karlin and Altschul; 1990; Karlin and Altschul; 1993; States and Gish, 1994; States et al., 1991; Wootton and Federhen, 1993). Blastn compares a nucleotide query sequence against a nucleotide sequence database;

(2) The Human genome sequencing centre (HGSC) has sequenced the genome of the rhesus monkey (rhesus macaque, *Macaca mulatta*). The sequencing and comparative analysis was funded by the National Human Genome Research Institute (NHGRI), National Institutes of Health (NIH), and led by the Baylor College of Medicine Human Genome Sequencing Center, and in collaboration with the J. Craig Venter Institute Joint Technology Center, and the Genome Sequencing Center at Washington University, St. Louis. The January 2006 rhesus macaque (*M. mulatta*) draft assembly — v.1.0, Mmul_051212 — used in this paper was obtained from the Baylor College of Medicine Human Genome Sequencing Center (BCM HGSC). Several whole genome shotgun (WGS) libraries, with inserts of 2 - 4 kb and 10 kb, fosmids with ~35 kb inserts, and BACs with 180 kb inserts were used to produce the clone-end sequence data. Approximately 20.1 million reads were used in the assembly, representing about 14.9 Gb of sequence and about 5.1X coverage of the (clonable) macaque genome. The assemblers produced a set of contigs and scaffolds. Scaffolds include sequence contigs that could be ordered and oriented with respect to each other as well as isolated contigs that could not be linked (single contig scaffolds or singletons). The N50 of the contigs is 25.7 kb and the N50 of the scaffolds is 5.87 Mb. (The N50 size is the length such that 50% of the assembled genome lies in blocks of the N50 size or longer). The total length of all contigs is 2.87 Gb. When the gaps between contigs in scaffolds are included, the total span of the assembly is 3.01 Gb (Michael, 2007; Rhesus Genome Project, 2006). This macaque genome is organized into 20 autosomes and the XY sex chromosomes.

(3) The NCBI Genome database provides views for a variety of genomes, complete chromosomes, sequence maps with contigs, and integrated genetic and physical maps. The database is organized in six major organism groups: Archaea, Bacteria, Eukaryotae, Viruses, Viroids, and Plasmids; and includes complete chromosomes, organelles and plasmids as well as draft genome assemblies. The Virus group was searched for Whole GP proteins of both Ebola and Marburg Viruses (Table 1 for accession and version) (Ryan et al., 2007; http://www.ncbi.nlm.nih.gov/genomes/static/vis.html).

Interventions

Both Ebola and Marburg GP gene sequences were queried against the whole genome of the January 2006 rhesus macaque (*M. mulatta*) draft assembly — v.1.0, Mmul_051212 (rheMac2) using BLAST-N analysis. Measured variables were number of sequences producing significant alignments as a measure of prevalence of integration, mapping of integration sites and efficiency of integration.

RESULTS

Of all the contigs length 2.87 Gb (2,863,665,185) bases

Table 1. ID Summary of Marburg and Ebola GP genes used in analysis.

Name of virus	Definition	Accession	Version
Marburg	Marburg Virus genomic RNA of GP gene.(bases 1 to 2948) Bukreyev, A., Volchkov, V.E., Blinov, V.M. and Netesov, S.V. The GP-protein of Marburg virus contains the region similar to the immunosuppressive domain' of oncogenic retrovirus P15E proteins FEBS Lett. 1993; 323 (1-2): 183-187	X68493	X68493.1 GI:296960
Ebola	Zaire Ebola virus envelope glycoprotein (GP) gene, complete cds, secreted small glycopro-tein (sGP) gene, complete cds. Volchkov, V.E., Becker, S., Volchkova, V.A., et al. GP mRNA of Ebola virus is edited by the Ebola virus polymerase and by T7 and vaccinia virus polymerases Virology 1995;214 (2): 421-430	U31033 S59050	U31033.1 GI:1141778

Source: National Center for Biotechnology information (NCBI): Genbank-viruses, Ebola and Marburg genes NCBI.

in the genome of rhesus macaque, Marburg GP blast hits to DB were 6,451,736 compared to 4,012,901 for Ebola. Prevalence of integration was measured as a direct representation by the number of significant alignments produced by the query viral GP gene against the rhesus macaca genome. Marburg GP genomic RNA exhibited 18 alignments relative to 7 of Ebola GP.

Efficiency of integration

All Ebola aligning Sequences had 100% efficiency compared to 66.6% of Marburg. We took this to mean that the Rhesus macaca is a more suited reservoir for EBOV than MBGV, although it is still premature at the moment to conclusively hold this view without *in vitro* molecular epidemiological assays, especially given that more alignments (despite their ill defined loci and imperfect alignment) were produced within the Rhesus macaca genome by the MBGV GP gene than Ebola.

Site mapping of loci of integration

Marburg GP genomic RNA had all its 18 alignments located on undefined scaffolds of the rhesus macaca genome compared to 7 of Ebola located on chromosomes 4, 6, 7, 8, 9, 14 and 15 (Tables 2 and 3). No *in vivo* or *in vitro* follow-up integration studies were done at this point, although in the past, such studies have employed host Cell culture and virus infection, Cloning of integration sites using ligated PCR methods, Mapping viral integration sites using the BLAT program, microarray analysis to determine the expression levels of integration sites, Analysis of base frequency around the integration sites and Bioinformatic analysis employing PERL (Practical Extraction and Report Language) to analyze the microarray data (Moalic et al., 2006).

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. A set of HSPs is thus defined by two sequences, a scoring system, and a cutoff score; this set may be empty if the cutoff scores insufficiently high. In the programmatic implementations of the BLAST algorithm described here, each HSP consisted of a segment from the query sequence and one from Rhesus genome database sequence, with 21 HSP's better >2.0; no gapping produced by Marburg GP RNA compared to only 7 for Ebola. The Effective length of HSP was 22. KARLIN-ALTSCHUL Statistical Constant values were Lambda = 1.37, H = 0.711, and K = 0.31.

E-values and Scores

The Score measures the sequence by sequence homogeneity within the query (filovirus GP) and the searched database (Rhesus genome). Ideally, a true homolog of, say of the Marburg GP used here of length 1 to 2948 bases should have at least above 80% score values or above, that's is about 2000bps. However, considering the wide evolutionary distance between the two taxa being compared (viruses and vertebrates), this is highly unlikely to occur. However, we hypothesize that it is likely that even the slight homogeneity in regions flanking (exons) the site of integration within the host may facilitate viral integration.

DISCUSSION

Our results serve to map the possible loci of integration sites for Ebola and Marburg viruses within the Rhesus macaca genome without necessary describing them, or their associated determinants. Moreover, because these results are based on a model of the integration of only the GP gene as the PIC, it would be more true that the integration loci discussed here are more likely to be of only the GP section of the entire filovirus genome, as its more likely that the actual *in-vivo* PC of the virus is the entire viral progeny (proviral RNA). If homogeneity betw-

Table 2. Summary of sequence alignments, Score, E-value and 100% identity for marburg GP BLAST against Rhesus Genome.

A. Marburg GP Queried Against Rhesus macaca Genome(18 Alignments)			
Sequence With Significant alignment	**Score (Bits)**	**E Value**	**(%) homogeinity**
gnl|Mmul_051212|Scaffold1099553000140	46	0.12	93%
gnl|Mmul_051212|Scaffold1099553000132	46	0.12	96%
gnl|Mmul_051212|Scaffold1099548049108	46	0.12	96%
gnl|Mmul_051212|Scaffold1099553000136	44	0.47	100%
gnl|Mmul_051212|Scaffold1099553000090	44	0.47	100%
gnl|Mmul_051212|Scaffold1099548049513	44	0.47	100%
gnl|Mmul_051212|Scaffold1099548049201	44	0.47	100%
gnl|Mmul_051212|Scaffold1099553000123	42	1.8	100%
gnl|Mmul_051212|Scaffold1099553000104_2_2	42	1.8	100%
gnl|Mmul_051212|Scaffold1099553000056	42	1.8	100%
gnl|Mmul_051212|Scaffold1099553000033	42	1.8	100%
gnl|Mmul_051212|Scaffold1099553000020	42	1.8	100%
gnl|Mmul_051212|Scaffold1099548049706	42	1.8	100%
gnl|Mmul_051212|Scaffold1099548049414	42	1.8	96%
gnl|Mmul_051212|Scaffold1099548049409	42	1.8	96%
gnl|Mmul_051212|Scaffold1099548049324	42	1.8	93%
gnl|Mmul_051212|Scaffold1099548049086	42	1.8	100%
gnl|Mmul_051212|Scaffold1099548049085	42	1.8	100%

Table 3. Summary of sequence alignments, score E-value and 100% identity for Ebola GP BLAST against Rhesus genome.

A. Marburg GP Queried Against Rhesus macaca Genome(7 Alignments)			
Sequences producing significant alignments	**Score (Bits)**	**E Value**	**(%) Homogeinity**
gnl|Mmul_051212|Chr9	48	0.023	100%
gnl|Mmul_051212|Chr14	44	0.36	100%
gnl|Mmul_051212|Chr8	44	0.36	100%
gnl|Mmul_051212|Chr15	42	1.4	100%
gnl|Mmul_051212|Chr7	42	1.4	100%
gnl|Mmul_051212|Chr6	42	1.4	100%
gnl|Mmul_051212|Chr4	42	1.4	100%

*Hits to DB 4,012,901.

een viral integration sites and the viral PIC is a determinant of viral integration as hypothesized, 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. 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. However, the exact effect of the imperfection or perfection of this flanking alignment on viral integration can only be defined by *in-vivo* experiments. If however, LINE elements favor viral integration as previous described (Stevens and Griffith, 1994; Pryciak et al., 1992; Vijaya et al., 1986), then it may be hypothesized that these imperfections may affect the site specificity of the endonuclease motif (C terminal Zn2+ fingerlike) within the LINE Reverse transcriptase consistent with all mobile Group II intronic elements, which is limited to 6-8 base pairs(Stevens and Griffith, 1994; Pryciak et al., 1992; Vijaya et al., 1986). In humans, who are close relatives of Rhesus- these ancestors of mobile Group II introns, also known as spliceosomal elements form about 25-35% of the entire Genome (International Human Genome Sequencing Consortium). More studies are however needed

both *in vitro* and *in vivo* to support these conclusions by clearly and finely demarcating the loci of both Ebola and Marburg integration sites on the gross picture provided by our results.

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 Ihypersensitive sites (Pryciak et al., 1992; Vijaya et al., 1986) or CpG islands (Scherdin et al., 1990) were apparently enriched near integration sites. Another study proposed that unusual host DNA structures were common near integration sites (Milot et al., 1994). A study of avian leukosis virus integration frequencies at several chromosomal sites failed to show any major differences among the regions studied (Withers-Ward et al., 1994), contrary to an earlier report (Sels et al., 1992). 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 (Stevens and Griffith, 1994) or Alu islands (Stevens and Griffith, 1996) or topoisomerase cleavage sites (Howard and Griffith, 1993). 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 (Pryciak et al., 1992; Bor et al., 1995; Bushman, 1994). In contrast, wrapping DNA on nucleosomes can create hot spots for integration at sites of probable DNA distortion (Pryciak et al., 1992; Pruss et al., 1994; Pryciak et al., 1992). Distortion of DNA in several other protein-DNA complexes can also favor integration (Muller and Varmus, 1994) consistent with the possibility that DNA distortion is involved in the integrase mechanism (Bushman and Craigie, 1992; Scottoline et al., 1997).

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 (Carteau et al., 1998). 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 (Moalic et al., 2006).

Using this model and the current literature, we conclude that these low homologous hits to filovirus GP may actually signify areas of “site specificity” of the endo- nuclaese activity of Rhesus ancestors of mobile Group II introns-LINE elements that have been previously been shown to be likely sites of viral integration *in-vivo*. Marburg may be better adapted to integrate into Rhesus than Ebola. However, if alignment is an inherent determinant of viral integration, one may otherwise say that either MBGV may have evolved to avoid integration in homogeneous chromatic host DNA to optimize gene expression, or it’s actually less adapted to tropism of rhesus tissue relative to Ebola with 100% alignment. Thus, 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 were Ebola glycoprotein gene complexes) than Marburg. Our results make future mapping for integrations sites for EBOV GP within rhesus macaca a more finely focused search, possibly limited to chromosomes 4, 6, 7, 8, 9, 14 and 15 of the Rhesus macaca mulatta genome.

REFERENCES

Altschul SF (1993). A protein alignment scoring system sensitive at all evolutionary distances. J. Mol. Evol.; 36: 290-300.

Altschul SF, Boguski MS, Gish W, Wootton JC (1994). Issues in searching molecular sequence databases. Nat. Genet. 6: 119-129.

Altschul, Stephen F (1991). Amino acid substitution matrices from an information theoretic perspective. J. Mol. Biol. 219: 555-65.

Altschul, Stephen F, Warren Gish, Webb Miller, Eugene Myers W, David Lipman J (1990). Basic local alignment search tool. J. Mol. Biol. 215: 403-10.

Baize S, Leroy EM, Georges-Courbot MC, Capron M, Lansoud-Soukate J, Debré P, Fisher-Hoch SP, McCormick JB, Georges AJ (1999). Defective humoral responses and extensive intravascular apoptosis are associated with fatal outcome in Ebola virus-infected patients. Nat. Med. 5: 423-426.

Bausch DG, Nichol ST, Muyembe-Tamfum JJ, Tugume B, Colebunders R, Kond ÃKM (2006). Marburg hemorrhagic fever associated with multiple genetic lineages of virus. N. Engl. J. Med., August 31; 355(9): 909 - 919.

Bor YC, Bushman F, Orgel L (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

Bowen ET, Platt GS, Lloyd G, Raymond RT, Simpson DI (1980). A comparative study of strains of Ebola virus isolated from southern Sudan and northern Zaire in 1976. J. Med. Virol. 6: 129-138.

Bushman FD (1994). Tethering human immunodeficiency virus 1 integrase to a DNA site directs integration to nearby sequences. Proc. Natl. Acad. Sci. USA 91: 9233-9237

Bushman FD, Craigie R (1992). Integration of human immunodeficiency virus DNA: adduct interference analysis of required DNA sites. Proc. Natl. Acad. Sci. USA 89: 3458-3462

Carteau S, Hoffmann C, Bushman F (1998). Chromosome structure and human immunodeficiency virus type 1 cDNA integration: centromeric alphoid repeats are a disfavored target. J. Virol. 72(5): 4005-14.

Claverie JM, States DJ (1993). Information enhancement methods for large scale sequence analysis. Comput. Chem. 17: 191-201.

Coffin JM (1996). Retroviridae: the viruses and their replication, *In* Fields BN, Knipe DM, Howley RM (ed.), Virology. Lippincott-Raven Publishers, Philadelphia, Pa. p. 1767-1848.

Feldmann H, Klenk HD, Sanchez A (1993). Molecular biology and evolution of filoviruses. Arch. Virol. 7: 81-100.

Felsenstein J (1993). PHYLIP: Phylogeny Inference Package,; version 3.5c. University of Washington, Seattle, WA, USA.

Gish W, States DJ (1993). Identification of protein coding regions by database similarity search. Nat. Genet. 3: 266-72.

Henikoff, Steven, Jorga G, Henikoff (1992). Amino acid substitution ma-

trices from protein blocks. Proc. Natl. Acad. Sci. USA1992; 89: 10915-19.
Howard MT, Griffith JD (1993). A cluster of strong topoisomerase II cleavage sites is located near an integrated human immunodeficiency virus. J. Mol. Biol. 232: 1060-1068
Karlin Samuel, Stephen Altschul F (1993). Applications and statistics for multiple high-scoring segments in molecular sequences. Proc. Natl. Acad. Sci. USA; 90: 5873-7.
Karlin, Samuel, Stephen Altschul F (1990). Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. Proc. Natl. Acad. Sci. USA; 87: 2264-68.
Leroy EM, Baize S, Volchkov VE, Fisher-Hoch SP, Georges-Courbot MC, Lansoud-Soukate J, Capron M, Debré P, McCormick JB, Georges AJ (2000). Human asymptomatic Ebola infection and strong inflammatory response. *Lancet*; 355: 2210-2215.
Mayo MA, Pringle CR (1998). Virus taxonomy – 1997. J. Gen. Virol. 79: 649-657.
Milot E, Belmaaza A, Rassart E, Chartrand P (1994). Association of a host DNA structure with retroviral integration sites in chromosomal DNA. Virology 201: 408-412
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.
Muller HP, Varmus HE (1994). DNA bending creates favored sites for retroviral integration: an explanation for preferred insertion sites in nucleosomes. EMBO J. 13: 4704-4714
National Center for Biotechnology information (NCBI): Genbank-viruses. Ebola and Marburg genes NCBI; Accessed May 2007 http://www.ncbi.nlm.nih.gov/genomes/static/vis.html
Pruss D, Bushman FD, Wolffe AP (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, Reeves R, Bushman FD, Wolffe AP (1994). The influence of DNA and nucleosome structure on integration events directed by HIV integrase. J. Biol. Chem. 269: 25031-25041
Pryciak P, Muller HP, Varmus HE (1992). Simian virus 40 minichromosomes as targets for retroviral integration *in vivo*. Proc. Natl. Acad. Sci. USA 89: 9237-9241
Pryciak PM, Sil A, Varmus HE (1992). Retroviral integration into minichromosomes *in vitro*. EMBO J. 11: 291-303
Rhesus Genome Project. (2006). http://www.hgsc.bcm.tmc.edu/projects/rmacaque/
Richard A, Gibbs, Jeffrey Rogers, Michael G, Katze, Roger Bumgarner, George M, Weinstock, Elaine R, Mardis, Karin A, Remington (2007). Rhesus Macaque Genome Sequencing and Analysis Consortium: Evolutionary and Biomedical Insights from the Rhesus Macaque. Genome Sci. 13 April;316(5822): 222 – 234
Ryan D, Hernandez, Melissa J, Hubisz, David A, Wheeler, David G, Smith, Betsy Ferguson, Jeffrey Rogers (2007). Demographic Histories and Patterns of Linkage Disequilibrium in Chinese and Indian Rhesus Macaques Science 13 April; 316(5822): 240 – 243.
Sanchez A, Kiley MP, Holloway BP, Auperin DD (1993). Sequence analysis of the Ebola virus genome: organization, genetic elements, and comparison with the genome of Marburg virus. Virus Res. 29: 215-240
Scherdin U, Rhodes K, Breindl M (1990). Transcriptionally active genome regions are preferred targets for retrovirus integration. J. Virol. 64: 907-912.
Scottoline BP, Chow S, Ellison V, Brown PO (1997). Disruption of the terminal base pairs of retroviral DNA during integration. Genes Dev. 11: 371-382.
Sels FT, Langer S, Schulz A. S, Silver J, Sitbon M, Friedrich RW (1992). Friend murine leukaemia virus is integrated at a common site in most primary spleen tumours of erythroleukaemic animals. Oncogene 7: 643-652.
States DJ, Gish W (1994). Combined use of sequence similarity and codon bias for coding region identification. J. Comput. Biol. 1: 39-50.
States DJ, Gish W, Altschul SF (1991). Improved sensitivity of nucleic acid database similarity searches using application specific scoring matrices. Methods: A companion to Methods in Enzymol. 3: 66-70.
Stevens SW, Griffith JD (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 SW, Griffith JD (1996). Sequence analysis of the human DNA flanking sites of human immunodeficiency virus type 1 integration. J. Virol. 70: 6459-6462.
Towner JS, Khristova ML, Sealy TK, Erickson BR, Bawiec D, Hartman AL (2006). Marburg virus genomics and association with a large hemorrhagic Fever outbreak in angola. J. Virol., 80(13): 6497 - 6516.
Varmus HE, Brown PO (1989). Retroviruses, *In* Berg DE, Howe MM (ed.), Mobile DNA. Am. Soc. Microbiol. Washington, D.C. pp. 53-108.
Vijaya S, Steffan DL, Robinson HL (1986). Acceptor sites for retroviral integrations map near DNase I-hypersensitive sites in chromatin. J. Virol. 60: 683-692
Watanabe S, Takada A, Watanabe T, Ito H, Kida H, Kawaoka Y (2000). Functional Importance of the coiled-coil of the Ebola virus glycolprotein. J. Virol. 74: 10194-10201.
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.
Withers-Ward ES, Kitamura Y, Barnes JP, Coffin JM (1994). Distribution of targets for avian retrovirus DNA integration *in vivo*. Genes Dev. 8: 1473-1487.
Wootton JC, Federhen S (1993). Statistics of local complexity in amino acid sequences and sequence databases. Comput. Chem. 17: 149-163.
Yang ZY, Delgado R, Xu L, Todd RF, Nabel EG, Sanchez A, Nabel GJ (1998). Distinct cellular interactions of secreted and transmembrane Ebola virus glycoproteins. Science 279: 1034-1037.