

**PENCIRIAN BERDASARKAN KOMPUTER RETROVIRUS  
ENDOGENOUS MANUSIA DARIPADA PANGKALAN DATA PROJEK  
GENOM MANUSIA**

**MOHD HAFIZ SHAUKAT BIN MOHD NEGUIB  
HS2003-2838**

**PERPUSTAKAAN  
UNIVERSITI MALAYSIA SABAH**

**DISERTASI INI DIKEMUKAKAN UNTUK MEMENUHI SEBAHAGIAN  
DARIPADA SYARAT UNTUK MEMPEROLEHI IJAZAH SARJANA  
MUDA SAINS DENGAN KEPUJIAN**

**PROGRAM BIOTEKNOLOGI  
SEKOLAH SAINS DAN TEKNOLOGI  
UNIVERSITI MALAYSIA SABAH**

**2006**



**UMS**  
UNIVERSITI MALAYSIA SABAH

## UNIVERSITI MALAYSIA SABAH

## BORANG PENGESAHAN STATUS TESIS@

UDUL: Pencirian Berdasarkan komputer terhadap  
endogenous moneta daripada portal data projek  
moneta  
 JAZAH: IJAZAH SARJANA MUDA MAINS DEN KEPUJIAN  
BIOTEKNOLOGI  
 SAYA MOHD HAFIZ PHAUKAT SESI PENGAJIAN: 2003-2006  
 (HURUF BESAR)

Mengaku membenarkan tesis (LPSM/Sarjana/Doktor Falsafah) ini disimpan di Perpustakaan Universiti Malaysia Sabah dengan syarat-syarat kegunaan seperti berikut:-

1. Tesis adalah hakmilik Universiti Malaysia Sabah.
2. Perpustakaan Universiti Malaysia Sabah dibenarkan membuat salinan untuk tujuan pengajian sahaja.
3. Perpustakaan dibenarkan membuat salinan tesis ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. Sila tandakan (/)

 SULIT

(Mengandungi maklumat yang berdarjah keselamatan atau Kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA RASMI 1972)

 TERHAD

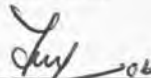
(Mengandungi maklumat TERHAD yang telah ditentukan oleh organisasi/badan di mana penyelidikan dijalankan)

 TIDAK TERHAD

Disahkan Oleh



(TANDATANGAN PENULIS)



(TANDATANGAN PUSTAKAWAN)

Alamat Tetap: 38, JLN INDAH 3,  
TAN CHERAS INDAH,  
56100. K.L.

DR. ROZIAH BT. KAMBOL  
 Nama Penyelia

Tarikh: 21-04-06Tarikh: 21-04-06

CATATAN:- \*Potong yang tidak berkenaan.

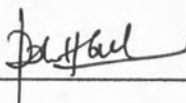
\*\*Jika tesis ini SULIT atau TERHAD, sila lampirkan surat daripada pihak berkuasa /organisasi berkenaan dengan menyatakan sekali sebab dan tempoh tesis ini perlu dikelaskan sebagai SULIT dan TERHAD.

@Tesis dimaksudkan sebagai tesis bagi Ijazah Doktor Falsafah dan Sarjana secara penyelidikan atau disertai bagi pengajian secara kerja kursus dan Laporan Projek Sarjana Muda (LPSM).



**PERAKUAN PEMERIKSA****DIPERAKUKAN OLEH**1. **PENYELIA**

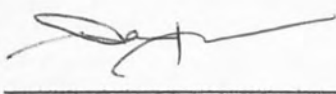
(DR. ROZIAH BINTI KAMBOL)

**Tandatangan**

---

2. **PEMERIKSA 1**

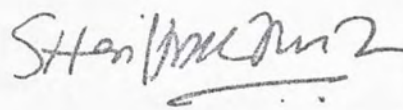
(CHANG PHEH PING)



---

3. **DEKAN**

(PROF. MADYA DR. SHARIFF A.K OMANG)



---



## PENAKUAN

Saya akui karya ini adalah hasil kerja saya sendiri nukilan dan ringkasan yang setiap satunya telah dijelaskan sumbernya.

MARCH 2006



---

MOHD HAFIZ SHAUKAT BIN MOHD NEGUIB  
HS2003-2838





## PENGHARGAAN

Assalamualaikumwarahmatullahiwabarakatuh Wamaghfiratuh Waridwanuh,

Syukur di atas rezeki nikmat pemberian Rabbi, akhirnya dengan kehendak-Nya saya berjaya menaip titik noktah pada mukasurat terakhir kajian saya ini. Tidak terkira bahagia bagai burung-burung berkicau merdu bak pesona seorang seniman memainkan irama dan alunan segar, semerdu dan selunak suara. Ucapan jutaan malah ratusan juga ribuan jika mungkin, terima kasih kepada penasihat projek saya Dr. Roziah binti Kambol kerana sudi mengajar saya yang sememangnya mentah dan kurang cerdik hendaknya. Juga kepada ibu dan ayah kerana memberi duit poket untuk membeli pena dan dakwat pencetak serta kepada Jamal Abdul Salam kerana sanggup menyediakan roti goreng tepat pada masanya, walaupun kadang-kadang memang lambat.

Seperti tidak percaya atau kaget malahan berada dalam alam mimpi apabila melihat projek telah selesai, bak kata rakan saya Syazwan, ‘aku nak lepaskan batu besar atas kepala aku ni..’. Sungguh sukar diduga pahit getir dalam mengharungi ujian sebagai pelajar tahun akhir dalam menyiapkan projek, di mana hari sebelum penghantaran saya sempat bermain pesta air untuk mewakili kolej, dan agak kecewa kerana penyertaan kolej telah dibatalkan akibat-akibat yang tertentu.

Semoga dengan kehadiran laporan projek ini, ia serba banyak atau semestinya dan tidak syak lagi dapat membantu warga pelajar dan pengkaji yang berminat dalam bidang retrovirus molekular untuk mendapat sedikit input daripada apa yang telah saya usahakan selama ini. Sekian Waalaikumussalam warahmatullahiwabarakatuh wamaghfiratuh waridwanuh. Hidup fungsi-!



## ABSTRAK

Dalam kajian ini, satu jujukan lengkap genom endogenous retrovirus telah diperolehi daripada pangkalan data genom manusia. Jujukan lengkap endogenous retrovirus ini diperolehi daripada jujukan dengan nombor akses gb|AF045450.1| yang terletak di kromosom 21q22.3 pada kosmid Q11M15. Saiz jujukan lengkap endogenous MSRV dianggarkan berjumlah 9.37k pasangan bes nukleotida. Jujukan endogenous retrovirus MSRV yang dijumpai dalam kajian ini mempunyai struktur organisasi seperti berikut; LTR-*gag-pol-env*-LTR di mana jujukan ini menyerupai jujukan kebanyakan retrovirus endogenous yang lain. LTR adalah sepanjang 355 pasangan bes nukleotida pada hujung 5' dan 367 pasangan bes nukleotida di hujung 3'. Dapat diperhatikan delesi sepanjang 12 pasangan bes di hujung 5' LTR. Keputusan BLAST2seq menunjukkan 79% homologi di antara 5' LTR dan 3' LTR. Beberapa motif terpelihara telah dijumpai di kesemua gen virus. Di kawasan *gag*, motif 'Major Homology Region' (MHR) dan motif 'zinc finger' telah dijumpai dengan jujukan QGKEENPTA dan CPLCQGNHWKAHC pada kedudukan 4386 dan 4869 pasangan bes nukleotida masing-masing. Di kawasan *pol*, enzim protease (LLDAGA), enzim transkriptase berbalik (CMDD) dan enzim RNaseH (YTDG) telah dijumpai pada kedudukan 5021, 5870 dan 6804 pasangan bes dalam genom retrovirus. Manakala di kawasan *env*, gen transmembran helices telah dijumpai dengan jujukan KPKPLFLCHSAVYCPLLKWYLSSQIEPLVWVSTFFEGLHAHKN pada lokasi 8769 pasangan bes nukleotida. Kesemua gen virus di atas telah disahkan motifnya melalui kaedah BLAST kedua, Pfam dan TM Pred, manakala LTR dicari menggunakan program BLAST2seq.





## ABSTRACT

In this project, the complete sequence of endogenous retrovirus' genome has been obtained from the human genome database. The complete genome sequence is acquired from sequence with the accession number of gb|AF045450.1| that is located on human chromosome 21q22.3 cosmid Q11M15. The size's of MSR'V endogenous retrovirus complete sequence is estimated about 9.37kb of nucleotides. The sequence of MSR'V endogenous retrovirus has the structure of LTR-*gag-pol-env*-LTR that is similar to many endogenous retroviruses. The 5' LTR has the length of 355 base pairs and the 3' end of LTR has 367 base pairs of nucleotides. 12 deletions can be observed on the 5' end of LTR, resulted to shorter LTR compared to 3' end of LTR. It is shown that 5' LTR and 3' LTR have 79% identity when they are put into BLAST2seq program in order to search for the designated LTR. A few conserved motifs have been found in the virus. In *gag* region, the 'Major Homology Region' and the zinc finger gene have been located at 4387 and 4870 base pairs with the amino acid sequence of QGKEENPTA and CPLCQGNHWKAHC. In *pol* region, protease enzyme (LLDAGA), reverse transcriptase enzyme (CMDD) and RNAseH (YTDG) have been observed on location 5022, 5871 and 6805 base pairs in the virus' genome. Meanwhile in the *env* region, the transmembrane gene has been observed on location 8770 base pairs with the following sequence; KPKPLFLCHSAVYCPLLKWYLSSQIEPLVWV STFFEGLHAHKN. Second BLAST searches, Pfam and TM Pred techniques confirmed all the genes that have been characterized, and the LTR sequences have been discovered using BLAST2seq program.



## KANDUNGAN

	Muka Surat
MUKA DEPAN	i
PENGAKUAN	ii
PERAKUAN PEMERIKSA	iii
PENGHARGAAN	iv
ABSTRAK	v
ABSTRACT	vi
KANDUNGAN	vii
SENARAI JADUAL	x
SENARAI RAJAH	xi
<b>BAB 1        PENDAHULUAN</b>	<b>1</b>
1.1 Pengenalan	1
1.2 Objektif Kajian	4
<b>BAB 2        ULASAN PERPUSTAKAAN</b>	<b>5</b>
2.1 Retrovirus	5
2.1.1 Pengelasan Tradisional Retrovirus	5
2.1.2 Pengelasan Moden Retrovirus	6
2.1.3 Retrovirus Endogenous Manusia	10
2.1.4 Struktur dan Komponen Retrovirus	16
2.1.5 Replikasi Retrovirus	20
2.2 Bioinformatik	22
2.2.1 Projek Pangkalan Data Genom Manusia	23
2.2.2 Pangkalan Data	24





2.2.3	Nilai E	25
2.2.4	Pfam	26
2.2.5	Rangka Bacaan Terbuka	28
2.2.6	BLAST	28
2.2.7	BLAST2seq	29
2.2.8	Transmembrane Prediction	30
<b>BAB 3</b>	<b>METODOLOGI</b>	<b>31</b>
3.1	Pembangunan Prob	31
3.2	Pencarian BLAST	32
3.2.1	Program BLAST	33
3.2.2	Pangkalan data dalam BLAST	36
3.2.3	Analisis Statistik	36
3.2.4	Kaedah BLAST	37
3.3	Penentuan Rangka Bacaan Terbuka (ORF)	38
3.4	Kaedah Carian BLAST kedua	39
3.5	Pfam	40
3.5.1	Struktur Kemasukan data dalam Pfam	41
3.5.2	Garis SA dan SS dalam susunan	42
3.5.3	Perbezaan antara karektor (-) dan (.)	43
3.6	BLAST2seq	44
3.7	Transmembrane Prediction	47
<b>BAB 4</b>	<b>KEPUTUSAN</b>	<b>48</b>
4.1	Jujukan Prob	48
4.2	Pencarian BLAST	49
4.3	Rangka Bacaan Terbuka	51



4.4	Pencirian keseluruhan genom retrovirus MSRV	52
4.5	Ringkasan bagi setiap gen	61
4.6	Carian kedua BLAST	62
4.7	Keputusan carian Pfam	64
4.8	Keputusan struktur transmembran dari TM Pred	66
4.9	Keputusan BLAST2seq	68
<b>BAB 5</b>	<b>PERBINCANGAN</b>	<b>70</b>
5.1	Analisis Kaedah	70
5.1.1	Prob untuk mencari gen yang dikehendaki	70
5.1.2	Pencarian BLAST	71
5.1.3	Penentuan Rangka Bacaan Terbuka	72
5.2	Analisis Jujukan	75
5.2.1	Kaedah carian BLAST kedua	75
5.2.2	Pencarian famili protein dalam Pfam	76
5.2.3	Meramal struktur transmembran	77
5.2.4	BLAST2seq untuk LTR	77
5.3	Analisis genom virus	78
5.3.1	LTR dan jujukan transkriptase kawalan	78
5.3.2	Gen <i>Gag</i>	79
5.3.3	Gen <i>Pro</i> dan <i>Pol</i>	81
5.3.4	Gen <i>Env</i>	83
<b>BAB 6</b>	<b>KESIMPULAN</b>	<b>85</b>
	<b>RUJUKAN</b>	<b>87</b>



## SENARAI JADUAL

	Muka Surat
2.1 Klasifikasi Terkini Retrovirus Endogenous Manusia	13
3.1 Jenis-Jenis Program dalam BLAST	34
3.2 Maksud disebalik kod-kod huruf yang digunakan dalam Pfam	42
4.1 Menunjukkan motif-motif terpelihara dan lokasinya bagi setiap gen	61





## SENARAI RAJAH

	Muka Surat
2.1 Struktur dan Komponen Retrovirus	17
2.2 Struktur genomik bagi retrovirus	19
2.3 Proses Replikasi dan Kitar Retrovirus	20
2.3 Organisasi dan Domain Protein	26
3.1 Program BLAST	36
3.2 Menunjukkan struktur kemasukan data dalam Pfam	42
3.3 Kemasukan data bagi HMM	44
3.4 Program BLAST2seq	45
3.5 Program TM Pred	47
4.1 Keputusan carian BLAST	49
4.2 Rangka bacaan terbuka retrovirus MSR/V	51
4.3 Struktur gen RT apabila BLAST ke dua dilakukan	62
4.4 Struktur gen RNaseH apabila BLAST ke dua dilakukan	63
4.5 Struktur gen <i>gag</i> dan zinc finger dalam carian Pfam	64
4.6 Struktur domain <i>gag</i> matriks dan RNaseH dalam Pfam	64
4.7 Nilai bacaan domain <i>gag</i> , <i>pro</i> dan RT dalam Pfam	65
4.8 Ramalan kehadiran struktur transmembran	66
4.8.1 Model topologi untuk struktur transmembran	66
4.9 Keputusan BLAST2seq untuk carian LTR	68



## BAB 1

### PENDAHULUAN

#### 1.1 Pengenalan

Retroviridae adalah nama taksonomi bagi keluarga besar virus RNA yang mengandungi virus-virus yang boleh menyebabkan tumor atau kanser darah putih (leukimia) dalam haiwan veterbrata. Virus-virus ini pada mulanya dikenali sebagai virus tumor RNA, leukovirus atau oncornavirus. Nama baru bagi virus-virus ini, iaitu retrovirus timbul daripada pengelasan kriteria yang mana virus-virus ini mengandungi enzim transkriptase berbalik. RNA ini membolehkan DNA polimerase menyalin terus DNA daripada genom virus, yang berbeza dengan proses genetik sebenar. Retrovirus-retrovirus berbeza daripada semua virus-virus RNA yang lain berdasarkan kepada keadaan genomnya. Struktur genetik retrovirus boleh mengakibatkan transformasi genetik kepada sel perumah dan boleh mengakibatkan penyakit tumor dan leukemia. Walaupun virus ini tidak patogenik, sitosidal atau boleh melakukan transformasi, kemungkinan untuk sel perumah mengalami perubahan dari segi genetik tetap wujud. Virus Tumor Tikus (MMTV) dan Virus Gross Leukimia Tikus (MuLV) adalah antara retrovirus onkogenik mamalia yang pertama ditemui (Gross, 1951). Retrovirus lain seterusnya telah ditemui daripada burung dan tikus



serta banyak virus telah didapati daripada beberapa perumah yang berbeza termasuklah manusia.

Bioinformatik merupakan satu aplikasi teknologi komputer terhadap pengurusan maklumat biologi. Komputer digunakan untuk mengumpul, menyimpan, menganalisis, dan mengintegrasikan maklumat biologi dan genetik yang mana seterusnya boleh diaplikasikan dalam penemuan dan perkembangan gen untuk menghasilkan dadah dan ubatan. Keperluan terhadap bioinformatik telah mendapat perhatian meluas melalui pendedahan yang disediakan daripada maklumat genomik terutamanya dari Projek Genom Manusia. Sains Bioinformatik, yang merupakan campuran biologi molekular bersama sains komputer perlu digunakan dalam maklumat genomik untuk lebih memahami tentang penyakit manusia dan dalam pengenalpastian molekul sasaran bagi penemuan dadah. Dalam usaha ini, kebanyakan universiti, institusi kerajaan dan firma farmasi telah menubuhkan kumpulan bioinformatik, yang terdiri daripada tokoh biologi komputer dan saintis bioinformatik komputer.

Keperluan utama dalam bioinformatik melibatkan analisis terhadap maklumat jujukan. Biologi Perkomputeran adalah nama yang diberikan terhadap proses ini, dan ia melibatkan penemuan gen-gen dalam jujukan DNA pelbagai jenis organisma. Proses selanjutnya melibatkan perkembangan kaedah-kaedah untuk mengenalpasti struktur dan fungsi protein-protein yang baru ditemui dan struktur jujukan-jujukan RNA. Biologi Perkomputeran juga melibatkan pemotongan jujukan-jujukan protein kepada beberapa famili jujukan yang berkait, dan pembangunan model-model protein dan seterusnya penyusunan protein-protein yang sama dan penubuhan pokok filogenetik untuk mengkaji hubungan evolusi antara organisma.





Bermula secara rasmi pada tahun 1990, Projek Genom Manusia Amerika Syarikat merupakan usaha 13 tahun Kementerian Tenaga Amerika Syarikat dan Institusi Kesihatan Antarabangsa. Projek asal telah dirancang semenjak 15 tahun yang lepas, tetapi kemajuan teknologi yang pesat telah mempercepatkan tarikh penyelesaian pada tahun 2003. Dalam beberapa dekad yang lalu, kemajuan dalam biologi molekular dan kelengkapan peralatan untuk tujuan kajian dalam bidang ini telah menyebabkan peningkatan yang cepat dalam penjujukan bahagian-bahagian besar genom bagi beberapa spesis. Fakta menunjukkan beberapa genom bakteria, sepertimana eukaryot-eukaryot ringkas dan eukaryot kompleks telah dijujukan dengan lengkap. Projek Genom Manusia ditubuhkan untuk menjujukan 24 kromosom manusia dan seterusnya melakarkan satu draf kasar yang selesai pada tahun 2003. Matlamat projek ini adalah untuk mengenalpasti jujukan bagi seluruh genom manusia (lebih kurang tiga billion pasangan bes).

Objektif penubuhan Projek Genom Manusia adalah untuk mengenalpasti kesemua lebih kurang 20,000-25,000 gen dalam DNA manusia dan untuk mengenalpasti jujukan-jujukan pasangan bes DNA. Selepas proses pengenalpastian jujukan-jujukan, maklumat yang diperolehi kemudiannya disimpan dalam pangkalan data di seluruh dunia dan dapat diakses di internet. Projek Genom Manusia turut memajukan perkakas data analisis yang berkaitan dengan kesemua kerja bioinformatik dan perkomputeran biologi. Ia turut membenarkan pemindahan teknologi yang berkaitan kepada sektor swasta dan memperkenalkan etika, undang-undang, dan isu sosial yang mungkin timbul daripada projek supaya projek ini dapat diterima oleh orang ramai.



## 1.2 Objektif Kajian ini adalah:

- 1 Untuk mendapatkan gen *gag*, *pol*, *env* dan LTR dalam retrovirus endogenous manusia daripada pangkalan data projek genom manusia
- 2 Untuk mencirikan keseluruhan retrovirus endogenous manusia yang merangkumi *gag*, *pol*, *env* dan LTR daripada pangkalan data projek genom manusia



## BAB 2

### ULASAN PERPUSTAKAAN

#### 2.1 Retrovirus

##### 2.1.1 Pengkelasan tradisional Retrovirus

Retoviridae adalah satu-satunya keluarga retrovirus yang boleh mengakibatkan transformasi pada sel perumah dan seterusnya menghasilkan virus daripada sel perumah. Tiga subfamili telah dikenalpasti berdasarkan aspek morfologi dan siri biologikal iaitu oncoviridae, spumavirinae dan lentivirinae. Kebanyakan oncoviridae telah dikaitkan dengan penyakit sarkoma dan leukemia dalam banyak spesis haiwan. Spumarivinae telah mendapat namanya daripada bentuk vakuol yang berbuih yang boleh dilihat pada sel perumah yang telah dijangkiti oleh subfamili retrovirus ini. Walaupun dikaitkan dengan perubahan bentuk pada perumah dalam kultur sel, ia tidak pernah dikaitkan dengan mana-mana penyakit. Lentiviridae pula mendapat namanya daripada tempoh jangkamasa inkubasi yang panjang akibat daripada jangkitan penyakit. Secara umum, kumpulan virus ini dikaitkan dengan penyakit kronik seperti encephalopathies, pneumonitis, artritis, kelemahan sistem pertahanan dan hemolitik anaemia. Ia telah dikenalpasti dalam banyak haiwan veterbrata seperti kuda, kambing, biri-biri, primat dan manusia (Heinz, 1988).





Subfamili oncovirinae seterusnya dikelaskan berdasarkan kriteria morfologikal. Virus jenis A selalunya terletak pada sitoplasma dan terdiri daripada dua jenis yang utama iaitu intrasitoplasmik dan intersisternal. Virus jenis intersisternal selalunya dijumpai pada embrio yang sedang berkembang dan selalunya virus jenis ini bertindak dengan virus tikus jenis D. Peranan virus ini masih lagi tidak ditemui. Virus jenis intrasitoplasmik merupakan pemangkin kepada partikel jenis B dan disisipkan ke dalam sampul virus apabila proses pertunasan virion dengan membran plasma berakhir. Dalam hal yang berlainan, partikel nukleoid jenis C wujud semasa proses pertunasan. Partikel jenis B boleh dibezakan dengan kehadiran duri-duri tajam dan nukleoid yang hadir pada partikel yang matang, berbeza dengan nukleoid sentral pada virus jenis C. Seterusnya, partikel jenis D wujud dalam bentuk nukleoid sempurna, yang mirip bentuk silinder sebelum proses pertunasan. Proses ini membezakan partikel tersebut daripada kumpulan famili yang lain. Spumavirus dan Lentivirus boleh dikenalpasti daripada perbezaan morfologi mereka apabila dilihat di bawah mikroskop elektron, tetapi mereka boleh berkongsi aspek morfologikal dan biologikal dengan sub-jenis oncovirus yang lain (Heinz, 1998).

### 2.1.2 Pengkelasan secara moden Retrovirus

#### a) Alpharetrovirus

Virus ini selalunya dikaitkan dengan banyak penyakit yang merbahaya seperti osteoporosis. Saiz genom retrovirus ini adalah sebanyak 7.2kb dan mempunyai gen (*gag*, *pro*, *pol* dan *env*) dengan tiada gen tambahan. Gen LTR adalah sepanjang 350 pasangan bes (U3, U5 dan kawasan R adalah 250 pasangan bes, 80 pasangan bes dan



20 pasangan bes). tRNA primer selalunya adalah tRNA<sup>trp</sup>. Struktur genomik bagi virus ini adalah gen *gag* dan *pro* berada dalam rangka bacaan terbuka (open reading frame) manakala *pol* dan *env* berada pada rangka bacaan terbuka yang berbeza (Van Regenmortel *et al.*, 2000).

b) Betaretrovirus

Saiz genom virus ini selalunya sepanjang 8-10 kb. Gen tambahan iaitu *sag*, telah wujud pada hujung 3' pada genom MMTV bersama dengan *gag-pol-env* dan bertindak sebagai jujukan superantigen semasa proses percantuman virus dengan membran perumah.

tRNA primer yang digunakan oleh virus ini adalah tRNA<sup>Lys</sup>, dengan tRNA<sup>Lys-3</sup> untuk MMTV dan untuk ahli kumpulan lain di bawah genus yang sama. LTR untuk MMTV adalah sepanjang 1300 pasangan bes dengan 1200 pasangan bes pada kawasan U3, 15 pasangan bes pada kawasan R dan 95 hingga 120 pasangan bes pada kawasan U5 (Heinz, 1998). Struktur genomik bagi genus ini adalah gen *gag*, *pro*, *pol* dan *env* yang wujud dalam rangka bacaan terbuka (ORF) yang berasingan (Van Regenmortel *et al.*, 2000).

c) Gammaretrovirus

Saiz genom bagi gammaretrovirus adalah sepanjang 8.3kb, dengan tiada gen tambahan. *Gag*, *pro* dan *pol* terletak di dalam rangka bacaan terbuka yang sama manakala gen *env* terletak pada rangka bacaan terbuka yang berbeza. tRNA primer



yang digunakan bagi genus kumpulan ini adalah tRNA<sup>Pro</sup> dan tRNA<sup>Glu</sup> (pada virus endogenous tertentu bagi tikus). LTR pula adalah sepanjang 600 pasangan bes, di mana kawasan U3 adalah sepanjang 500 pasangan bes, kawasan R pula adalah 60 pasangan bes dan kawasan U5 sepanjang 75 pasangan bes (Van Regenmortel *et al.*, 2000).

d) Deltaretrovirus

Saiz genom ini adalah sepanjang 8.3kb dengan gen virus yang umum iaitu *gag*, *pro*, *pol* dan *env* dalam rangka bacaan terbuka yang berbeza. Dua gen tambahan iaitu *tax* dan *rex* wujud dalam genom deltaretrovirus dan telah terbukti bahawa gen tambahan tersebut mempunyai peranan yang signifikan semasa proses replikasi dan transkripsi. tRNA primer yang digunakan adalah tRNA<sup>Pro</sup> dan LTR adalah sepanjang 550-750 pasangan bes (200-300 pasangan bes pada kawasan U3, 135-235 pasangan bes pada kawasan R dan 100-200 pasangan bes pada kawasan U5) (Van Regenmortel *et al.*, 2000).

e) Epsilonretrovirus

Dalam klasifikasi terkini, virus daripada kumpulan ini telah dikelaskan kepada satu genus yang baru berdasarkan kepada struktur genomiknya yang kompleks, di mana beberapa gen tambahan dan satu tRNA primer telah diperhatikan. Sebagai contoh, Virus Sarkoma Walleye Dermal (WDSV) mempunyai tiga gen tambahan berfungsi yang dikenali sebagai *orf-a*, *orf-b* dan *orf-c* dengan tambahan daripada tiga gen utama





iaitu *gag*, *pol* dan *env* (Van Regenmortel *et al*, 2000). Pada masa *orf-a* dan *orf-b* terletak pada hujung 5' padan *env*, *orf-c* pula mendahului gen *gag*. Tiada daripada ketiga-tiga gen tersebut menunjukkan jujukan yang sama kepada gen yang telahpun diketahui ataupun telah wujud sebelum ini. WDSV juga menggunakan tRNA<sup>His</sup> primer baru yang tidak diperhatikan pada mana-mana retrovirus yang lain sebelum ini. Saiz genom pada WDSV adalah sepanjang 12.8 kb dengan LTR menduduki kawasan penghujung genom yang bersaiz 90 pasangan bes. Daripada 590 pasangan bes pada jujukan LTR, 440 daripada pasangan bes adalah dalam kawasan U3, 77 dan 73 pasangan bes pada kawasan R dan kawasan U5 (Van Regenmortel *et al.*, 2000).

f) Spumaretrovirus

Saiz genom spumaretrovirus adalah sepanjang 11.6 kb dengan gen umum dalam retrovirus seperti gen *gag* dan *env* yang menduduki rangka bacaan terbuka yang berbeza manakala *pro* dan *pol* berada dalam rangka bacaan terbuka yang sama. Dua gen tambahan iaitu *tas* dan *bet* juga hadir dalam genom spumaretrovirus (Van Regenmortel *et al*, 2000). Gen *tas* telah dipercayai menjadi protein pencantuman DNA dengan fungsi transaktif semasa transkripsi virus, fungsi gen *bet* masih lagi tidak diketahui. tRNA primer yang digunakan adalah tRNA<sup>Lys1,2</sup> dan LTR pada haiwan primat dalam spumaretrovirus adalah sepanjang 1770 pasangan bes (1400 pasangan bes pada kawasan U3, 200 pasangan bes dalam kawasan R dan 150 pasangan bes pada kawasan U5) (Van Regenmortel *et al.*, 2000).



### g) Lentivirus

Saiz genom bagi lentivirus adalah sepanjang 9.3 kb dengan beberapa gen tambahan wujud seperti *vif*, *vpr*, *vpx*, *tat*, *rev*, *rex*, *vpu* dan *nef* bergantung kepada jenis virus itu. Gen tambahan yang terlibat dalam regulasi sintesis/proses pembentukan viral RNA memainkan peranan penting dalam replikasi virus. Kebanyakan gen tambahan terletak pada hujung 3' dalam *gag-pro-pol*, kebanyakannya menjadi hujung 5' untuk gen *env* dan hujung 3' untuk *env*. tRNA primer yang digunakan adalah tRNA<sup>Lys1,2</sup>. LTR pula adalah sepanjang 600 pasangan bes (450 dalam kawasan U3, 100 pasangan bes dalam kawasan R dan 80 pasangan bes dalam kawasan U5) dan *gag*, *pro*, *pol* dan *env* semuanya menduduki rangka bacaan terbuka yang berasingan (Van Regenmortel *et al.*, 2000).

#### 2.1.3 Pengkelasan Retrovirus Endogenous pada manusia

Retrovirus yang berasal daripada manusia merupakan retrovirus yang paling baru dijumpai berbanding dengan famili-famili retrovirus yang lain. Virus yang pertama merupakan virus 'foamy' yang telah dijumpai pada tahun 1971 pada kultur sel kanser pada manusia. Walaubagaimanapun, kaitan virus tersebut dengan penyakit yang dihadapi oleh manusia masih lagi tidak dapat dipastikan. Virus kedua ialah oncovirus yang berkaitan dengan sel-T leukemia yang dijumpai di utara Kepulauan Jepun, dan di Selatan Amerika Syarikat serta di kepulauan Caribbean. Virus tersebut dikenali sebagai sel-T leukemia virus (HTLV-1) ataupun virus sel-T dewasa, yang menjangkiti sel-sel tertentu, khususnya subset pembantu sel-T dan boleh mengabadikan virus itu sendiri di dalam 'cell line' yang telah terbentuk (Heinz, 1988).



Virus ini walaubagaimanapun boleh menjangkiti fibroblast di dalam kultur. Selain daripada itu, HTLV-1 telah dikenalpasti dalam jangkitan pada arnab dan tikus. HTLV-1 jenis lain yang dipanggil sebagai HTLV-II telah diperolehi daripada sel leukemia manusia, tetapi kaitan virus tersebut dengan penyakit masih lagi tidak dicatat sehingga kini (Heinz, 1988).

Lentivirus merupakan jenis virus yang paling baru ditemui pada manusia. Ia telah dikaitkan dengan virus AIDS. Lentivirus pada manusia amat mudah untuk dibezakan dengan retrovirus yang lain berdasarkan daripada struktur morfologi dan biologikalnya. Tiga kelas berbeza daripada lentivirus yang dijumpai pada manusia telah dilaporkan dan setiap kelas dipanggil sebagai 'Lymphenopathy associated virus' ataupun LAV, 'Human T sel lymphotropic virus type III' (HTLVIII) dan virus yang berkaitan dengan AIDS iaitu ARV.

Lebih daripada 20 famili HERV telah dikenalpasti lebih daripada dua dekad yang lalu. Tristem dalam kerjanya telah mengklasifikasi HERV kepada 22 kumpulan berdasarkan data yang telah diperolehi daripada kajian Projek Genom Manusia (Tristem, 2000). Berdasarkan kepada kajian terkini, HERV telah diklasifikasikan kepada 3 kelas utama. Tiga kelas utama tersebut adalah kumpulan I, II dan III (Nelson *et al.*, 2003).

HERV kelas I boleh dibahagikan lagi kepada enam kumpulan yang berkongsi aspek homologi dengan gammaretrovirus. 3 famili daripada kumpulan ini berhomologi dengan Virus leukemia Murine (MuLV) dan Virus Endogenous Babun (BaEV) berdasarkan kepada analisis molekular dan enzim transkriptase berbalik. Ahli





daripada kumpulan HERV kelas I ini adalah HERV-H, HERV-I dan HERV-R (ERV-9).

HERV kelas II berhomologi dengan virus daripada genera betaretrovirus dan deltaretrovirus. Kelas ini boleh dibahagikan lagi kepada 10 kumpulan kecil. Virus yang mempunyai kaitan dengan kumpulan ini adalah HERV-K (C4). Semua daripada HERV kelas II menunjukkan tRNA lisin berasal daripada virus kumpulan betaretrovirus dan deltaretrovirus. HERV-K adalah virus yang paling aktif secara biologikal. Kumpulan ini seterusnya dipecahkan lagi kepada jenis I dan jenis II, berdasarkan kepada kewujudan 292 segmen pasangan bes pada sempadan *pol-env* (Nelson *et al.*, 2003).

Kumpulan ini mempunyai persamaan dengan spumavirus dan dua jenis daripada HERV adalah termasuk daripada kumpulan ini iaitu HERV- dan HERV-S. Sistem klasifikasi ini menyatukan sistem klasifikasi HERV tetapi retrovirus endogenous yang berkesinambungan dengan sel-T leukemia manusia, HRES-1 tidak termasuk di dalam klasifikasi ini kerana kumpulan tersebut menunjukkan homologi yang sedikit dengan HTLV-1 dalam kawasan LTR. Jadual 2.1 di bawah merupakan ringkasan kepada pengkelasan terkini retrovirus endogenous pada manusia.



## RUJUKAN

- Attwood, T. K. dan Parry-Smith, D. J. 1999. *Introduction to Bioinformatics*. Pearson Education Limited, England.
- Altschul, S. F., W. Gish, W. Miller, E. W. Myers, dan D. L. Lipman. 1990. *Basic alignment search tool*. J. Mol. Biol. 215:403-410
- Baltimore, D. (1985) *Retroviruses and retrotransposons: the role of reverse transcriptase in shaping the eukaryotic genome*. Cell 40:481
- Benit, L., Dessen, P., dan Heidmann, T. 2001. *Identification, Phylogeny, and Evolution of Retroviral Elements Based on Their Envelope Genes*. J. Virol. p. 11709-11719, Vol. 75, No. 23
- Blond, J. L., Besème, F., Duret, L., Bouton, O., Bedin, F., Perron, H., Mandrand, B., dan Mallet, F. 1999. *Molecular Characterization and Placental Expression of HERV-W, a New Human Endogenous Retrovirus Family*. J. Virol. p. 1175-1185, Vol. 73, No. 2
- Bruce, A. V, 2002, *The Biology of Viruses 2<sup>nd</sup> Edition*, New York
- Boyer, P. L., Ferris, A. L., dan Hughes, S. H., 1992, *Mutational Analysis Of The Fingers Domain Of Human Immunodeficiency Virus Type I Reverse Transcriptase*. J. Virol. 66:7533-7537
- Burke, C. J., Sanyal, G., Burner, M. W., Ryan, J. A., LaFemina, R. L., Robbins, H. L., Zeff, A. S., Middaugh, C. R., dan Cordingley, M. G. 1992. *Structural Implication of Spectroscopic Characterisation Of A Putative Zinc Finger Peptide From HIV-1 Integrase*. J. Biol. Chem 267:9639-9644.
- Cooper, G. M. dan Hausman, R. E. 2004. *The Cell a Molecular Approach*. 3<sup>rd</sup> Edition. ASM Press, Washington.



- Franke, E. K., Yuan, H. E. H., Bossolt, K. L., Goff, S. P., dan Luban, S. 1994. *Specificity and Sequence Requirements for Interactions Between Various Retroviral gag Proteins*. J. Virol. 68: 5300-5305.
- Gross, L. 1951. "Spontaneous" leukemia developing in CH3 mice following inoculation, in infancy, with AK leukemia extracts for AK embryos. Proc. Soc. Exp. Biol. Med. 76:27
- Heinz, F. C. 1988. *Virology Second Edition*. Prentice Hall, New Jersey.
- Krane, D. E. dan Raymer, M. L. 2003. *Fundamental Concepts of Bioinformatics*. Pearson Education, Inc., San Francisco.
- Karp, G. 2002. *Cell and Molecular Biolog. Concepts and Experiments*. 3<sup>rd</sup> Edition. John Wiley & Sons Inc., New York.
- Kambol, R. 2003. *Evolution and Distribution of Endogenous Retrovirus within Amphibian and Piscine Hosts*. Unpublished PhD Thesis. Department of Biological Sciences, Imperial College London, United Kingdom.
- Kambol, R. 2005. *Xenopus Endogenous Retrovirus (XEN) Possesses a Relatively Complex Retrovirus Sequence In Its Genome*. Borneo Science. In Press.
- Lewin, B. 2004. *Genes VIII International Edition*. Pearson Prentice Hall, America.
- Lesk, M. L. 2002. *Introduction to Bioinformatics*. Oxford University Press Inc., New York.
- Madigan, M. T., Martinko, J. M., dan Parker, J. 2003. *Brock Biology of Microorganisms*. 10<sup>th</sup> Edition. Pearson Education Inc., United States of America.
- Nelson, C., Shen, L. P., Meister, A., Fodor, E., dan Rutter, W. J. 1990. *Pan: A Transcriptional Regulator That Binds Chymotrypsin, Insulin And AP-4 Enhancer Motif*. Genes Dev. 4: 1035-1043.





- Nelson, P.N, Carnegie, P.R., Martin, J., Ejtehadi, H.D., P Hooley, P., Roden, D., Rowland-Jones, S., P Warren, P., Astley, J., and Murray, P.G., 2003. *Demystified: Human endogenous retrovirus*. *Molecular Pathology* 56, 11-18.
- Tizard, I. R. 1984. *Immunology An Introduction*. Thomson Learning, Inc., America.
- Tristem. M. 2000. *Identification and Characterisation Of Novel Human Endogenous Retrovirus Families by Phylogenetic Screening of Human Mapping Project Databes*. *J. Virol.* 74:3715-3730.
- Van Regenmortel, Fauquet, M. H. V., Bishop, C. M., Carsten, D. H. L., Estes, E.B., Lemon, S. M., Maniloff, J., Mayo, M. A., McGeoch, D. J., Pringle, C. R., Wickner, R. B. 2000. *Virus Taxonomy: The Classification And Nomenclature of viruses*. The seventh report of the International Committee on Taxonomy of Viruses. Academic Press. San Diego.
- Wilkinson, D. A., Mager, D. L., dan Leong, J. A. C. 1994. *Endogenous Human Retroviruses. Eds A Levy In The Retroviridae. Vol 3*. Plenum Press, New York.
- Anon,  
[//www.ch.embnet.org/TmPred.html](http://www.ch.embnet.org/TmPred.html)
- Anon,  
[//www.expasy.org](http://www.expasy.org)
- Anon,  
[//www.ncbi.nlm.nih.gov/BLAST/BLAST2seq.html](http://www.ncbi.nlm.nih.gov/BLAST/BLAST2seq.html)
- Anon,  
[//www.ncbi.nlm.nih.gov/Education/BLASTinfo/rules.html](http://www.ncbi.nlm.nih.gov/Education/BLASTinfo/rules.html)
- Anon,  
<http://www.sanger.ac.uk/Software/Pfam/>



Anon,

[www.tulane.edu/dmsander/WWW/335/Retrovirus.html](http://www.tulane.edu/dmsander/WWW/335/Retrovirus.html)

