

PENYARINGAN SECARA BERKOMPUTER ELEMEN
LTR RETROTRANSPOSON DI DALAM
GENOM MANUSIA

NURUL HIDAYAH MOHAMED ISA

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

PROGRAM BIOTEKNOLOGI
SEKOLAH SAINS DAN TEKNOLOGI
UNIVERSITI MALAYSIA SABAH

Mac 2005



UMS
UNIVERSITI MALAYSIA SABAH

UNIVERSITI MALAYSIA SABAH

BORANG PENGESAHAN STATUS TESIS@

JUDUL: PENYARINGAN SECARA BERKOMPUTER ELEMEN LTR-RETROTRANSPONSON
DI DALAM PANGKALAN DATA PROJEK GENOM MANUSIA

Ijazah: IJAZAH SARJANA MUDA SAINS DENGAN KEPUJIAN (BIOTERNOLOGI)

SESI PENGAJIAN: 2002/2005

Saya NURUL HIDAYAH MOHAMED ISA

(HURUF BESAR)

mengaku membenarkan tesis (LPS/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

H. Hidayah

(TANDATANGAN PENULIS)

(TANDATANGAN PUSTAKAWAN)

Alamat Tetap: LOT 85, TAMAN LIMAN,
KAITI, 33020 KUALA KANGSAR

Nama Penyelia

Tarikh: 31 MAC 2005

Tarikh: _____

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.

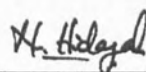
@ Tesis dimaksudkan sebagai tesis bagi Ijazah Doktor Falsafah dan Sarjana secara penyelidikan, atau disertasi bagi pengajian secara kerja kursus dan penyelidikan, atau Laporan Projek Sarjana Muda (LPSM).



PENGAKUAN

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

31 Mac 2005



NURUL HIDAYAH MOHAMED ISA

HS 2002-3101

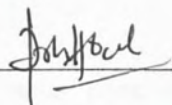


DIPERAKUKAN OLEH

Tandatangan

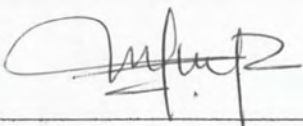
1. PENYELIA

(Dr Roziah Hj Kambol)



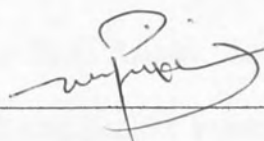
2. PEMERIKSA 1

(Dr Vijay Kumar)



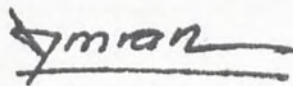
3. PEMERIKSA 2

(Puan Teoh Peik Lin)



4. DEKAN

(Prof. Madya Dr. Amran Ahmed)





PENGHARGAAN

Syukur ke hadrat ilahi kerana dengan limpah dan kurnianya, saya dapat menyiapkan kajian ini dengan sempurna. Di kesempatan ini, saya ingin mengucapkan jutaan terima kasih kepada penyelia saya, Dr Roziah Hj Kambol kerana banyak menolong saya dalam menyiapkan kajian ini. Segala tunjuk ajar, idea dan teguran amat saya hargai. Tidak lupa kepada pembantu makmal, Cik Nor Baizura dan En. Noel yang banyak membantu dalam kajian ini.

Saya juga ingin merakamkan penghargaan buat keluarga saya yang sentiasa memahami dan memberi sokongan samada secara material ataupun emosi. Sokongan ini, amat saya hargai.

Sekalung penghargaan buat teman-teman bioinformatik, Roziah, Robert, Sabrina, Saratha, Malaty dan Anusuya. Tak lupa pada sahabat-sahabat lain, Azmira, Ismak, Shima, Julaiha dan kepada semua yang membantu saya dalam menyiapkan kajian ini samada secara langsung atau tidak.



ABSTRAK

Tujuan kajian ini adalah untuk menyaring retroelemen *Gypsy* dan *Del* daripada serangga *Drosophila Melanogaster* di dalam genom manusia. Retroelemen adalah elemen bergerak yang mengandungi gen transkriptase berbalik. Elemen yang menggunakan perantara RNA ini boleh menyebabkan ketakstabilan genetik melalui mutasi penyisipan dan mutasi lain. Kaedah penyaringan yang digunakan adalah secara berkomputer. Retroelemen tersebut disaring di dalam pangkalan data manusia. Retroelemen yang dipilih disusun dengan menggunakan dua cara; cara manual (penyusunan mata kasar) dan cara berkomputer (program clustalw). Pokok filogenetik khusus digunakan untuk melihat hubungan evolusi antara prob yang digunakan dengan retroelemen lain. Pokok filogeni universal pula digunakan untuk melihat hubungan antara retroelemen dan retrovirus. Daripada kajian sebanyak 773 elemen *Gypsy* dan 773 elemen *Del* berjaya dipancing keluar daripada bank gen. Daripada jumlah ini hanya 100 elemen *Gypsy* dan 100 elemen *Del* dipertimbangkan. 28 elemen *Gypsy* dan 21 elemen *Del* disusun secara manual. Hanya 13 elemen *Gypsy* dan 8 elemen *Del* disusun dengan menggunakan program clustalw dan pembinaan pokok filogenetik. Keputusan yang diperolehi daripada kajian ini menunjukkan retroelemen *Gypsy* dan *Del* pada serangga *Drosophila Melanogaster* ini telah dijumpai dalam genom manusia. Kedua-dua elemen ini paling banyak didapati pada kromosom seks,X. Selain itu elemen-elemen ini turut didapati dalam bilangan yang tinggi dalam kromosom autosom. Elemen *Gypsy* dijumpai di dalam kromosom pertama dan kedua. Elemen *Del* pula banyak terdapat pada kromosom kelima dan ketujuh.



ABSTRACT

The purpose of this study was to screen retroelements of *Gypsy* and *Del* from insect *Drosophila Melanogaster* in human genome. Retroelements are mobile elements containing reverse transcriptase that used RNA intermediate to jump among chromosome. It can make genetic instability through insertion and other mutation. Computer based screening was used as screening method. Retroelements was screened in human database. The selected sequence had been aligned with two methods; manual (eyeball alignment) and with computer program (using clustalw program). Specific phylogenetic tree was used to determine evolutionary relationship among prob that been used with other prob. Universal phylogenetic tree was used to determine the relationship between retroelements and 7 genera of retrovirus. 773 of *Gypsy* elements and 773 of *Del* elements were fishing out from human genome database. From this search, only 100 of *Gypsy* elements and 100 of *Del* elements have chosen. Only 28 *Gypsy* elements and 21 *Del* elements were aligning by manual alignment. However, only 13 elements *Gypsy* and 5 *Del* elements were chosen for sequence alignment with clustalw program and for phylogenetic tree construction. The result for this study showed that retroelements of *Gypsy* and *Del* from *Drosophila Melanogaster* are found in human genome. These two elements frequently seen in X chromosome. They also seen in autosome chromosomes like in *Gypsy* is found in first and second chromosome while *Del* is found in fifth and seventh chromosome.



KANDUNGAN

	Muka surat
PENGAKUAN	ii
PENGESAHAN	iii
PENGHARGAAN	iv
ABSTRAK	v
ABSTRACT	vi
SENARAI KANDUNGAN	vii
SENARAI JADUAL	ix
SENARAI RAJAH	x
BAB 1 PENDAHULUAN	1
1.1 Pengenalan	1
1.2 Objektif Kajian	2
BAB 2 ULASAN PERPUSTAKAAN	3
2.1 Bioinformatik	3
2.1.1 Penyaringan Bank gen	4
2.1.2 Pembinaan pokok filogenetik	5
2.2 Projek Genom Manusia	6
2.3 Penemuan Elemen Bergerak	7
2.4 Pengkelasan retroelemen	9
BAB 3 KAEDAH	13
3.1 Mendapatkan jujukan gen transkriptase berbalik elemen LTR Retrotransposon yang bertindak sebagai primer	13
3.2 Penyaringan Bank Gen	14
3.3 Penyusunan jujukan	18
3.4 Pembinaan pokok filogenetik	20



BAB 4 KEPUTUSAN DAN ANALISIS DATA	22
4.1 Mendapatkan jujukan gen transkriptase berbalik elemen LTR Retrotransposon yang bertindak sebagai primer	22
4.2 Penyaringan Bank Gen	23
4.3 Penyusunan jujukan	32
4.4 Pembinaan pokok filogenetik	39
BAB 5 PERBINCANGAN	42
5.1 Mendapatkan jujukan gen transkriptase berbalik elemen LTR Retrotransposon yang bertindak sebagai primer	42
5.2 Penyaringan Bank Gen	44
5.3 Penyusunan jujukan	45
5.4 Pembinaan pokok filogenetik	47
BAB 6 KESIMPULAN	50
RUJUKAN	51



SENARAI JADUAL

No. Rajah	Muka surat
3.1 Program-program di dalam BLAST	15
4.1 Motif terpelihara	23
4.2 (a) Data BLAST bagi Gypsy	26
4.2 (b) Data BLAST bagi Del	30



SENARAI RAJAH

No. Rajah	Muka surat
2.1 Pengelasan retroelemen	11
2.2 Evolusi retroelemen	12
3.1 Kotak carian BLAST	16
3.2 Contoh rajah keputusan carian BLAST	17
3.3 Format clustalw	19
3.4 Kotak penyusunan clustalw	19
3.5 Contoh fail dnd	21
4.1 (a) Keputusan pencarian BLAST bagi Gypsy	24
(b) Graf elemen melawan jenis kromosom bagi Gypsy	27
4.2 (a) Keputusan pencarian BLAST bagi Del	28
(b) Graf elemen melawan jenis kromosom bagi Del	31
4.3 Carta pai domain kelima gen transkriptase berbalik	32
4.4 Penyusunan secara manual bagi Gypsy	34
4.5 Penyusunan secara manual bagi Del	36
4.6 Penyusunan jujukan dengan menggunakan clustalw	38
4.7 Pokok filogeni spesifik	39
4.8 Pokok filogeni universal	41
5.1 Hubungan antara famili Ty3/Gypsy dan retrovirus	43



BAB 1

PENDAHULUAN

1.1 Pengenalan

Elemen bergerak memberi impak yang besar pada semua genom eukariot walaupun pada asalnya ia lebih dikenali sebagai 'junk' DNA. Elemen yang menggunakan perantara RNA ini dikenali sebagai retroelemen dan mengkodkan enzim transkriptase berbalik. Retroelemen boleh menyebabkan ketakstabilan genetik melalui mutasi penyisipan dan mutasi lain.

Kajian yang terperinci telah dijalankan untuk mengetahui bagaimana fungsi dan pergerakan elemen ini di dalam sesuatu genom organisma. Selain itu, kita juga boleh mengetahui kesan penyisipan elemen ini pada kepelbagaian diversiti dan penyakit manusia.

Terdapat pelbagai hipotesis yang menyatakan retroelemen terlibat dalam proses evolusi. Terdapat bukti yang menyatakan retroelemen mengadaptasi perumahannya. Walaupun ia selalunya tidak memberi kesan pada perumah tetapi terdapat banyak contoh kes yang menunjukkan mutasi yang disebabkan oleh elemen ini. DNA transposon, LTR-



retrotransposon, *long interspersed elements* (LINEs) dan *short interspersed elements* (SINEs) membentuk sekurang-kurang 45% daripada DNA manusia dan 8% dari nilai itu disumbangkan oleh LTR-retrotransposon.

Projek Genom Manusia adalah projek yang mengenalpasti sehingga 20 000 – 25 000 gen di dalam DNA manusia. Gen memberi maklumat untuk membina protein yang diperlukan oleh manusia. Projek Genom Manusia ini juga telah mengenalpasti 3 bilion jujukan pasangan bes kimia yang membentuk DNA manusia. Semua maklumat ini tersimpan dalam pangkalan data manusia.

1.2 Objektif

Objektif bagi kajian ini adalah

- 1) Penyaringan secara berkomputer retroelemen daripada pangkalan data genom manusia
- 2) Penyusunan jujukan retroelemen yang boleh didapati dalam genom manusia
- 3) Pembinaan pokok filogenetik untuk mengenalpasti hubungan evolusi antara primer yang digunakan dan retroelemen lain



BAB 2

ULASAN PERPUSTAKAAN

2.1 Bioinformatik

Kemajuan dalam bidang biologi molekul dan kemudahan yang terkini telah menggalakkan penyelidikan di dalam bidang ini pada kebanyakan spesies. Terdapat banyak pangkalan data yang popular seperti GenBank dan EMBL yang telah dibangunkan dan semua data yang diperolehi akan disimpan, disusun dan didaftarkan dalam bentuk jujukan DNA. Semua maklumat sains yang digunakan dalam bidang biologi dikenali sebagai bioinformatik. Bioinformatik melibatkan penciptaan dan penyelenggaraan pangkalan data maklumat biologi. Ia menggabungkan proses penjujukan DNA dan aplikasi komputer dalam sains biologikal (Attwood dan Parry-Smith, 1999).

Dalam kata umum, bioinformatik merujuk pada teknologi maklumat yang digunakan untuk menganalisa samada dalam bentuk jujukan nukleotida dan protein hinggalah ke peringkat analisa genom. Ia juga boleh ditafsirkan sebagai manipulasi komputer dan analisis jujukan data biologikal (samada DNA atau protein) (Attwood dan Parry-Smith, 1999).



2.1.1 Penyaringan Bank Gen

Langkah penyaringan bank gen dilakukan dengan menggunakan kaedah BLAST (Basic Local Alignment Search Tool) di laman web <http://www.ncbi.nlm.nih.gov/>. BLAST adalah satu kaedah untuk mencari persamaan jujukan. Program ini memerlukan satu jujukan yang ingin dibezakan (*query sequence*) yang digunakan sebagai prob. Ia akan menyusun jujukan prob tersebut dan membandingkannya dengan setiap jujukan subjek (*subject sequences*) yang terdapat dalam pangkalan data (Altschul *et al.*, 1990).

Keputusan BLAST memaparkan bit skor, nilai E, nombor akses dan susunan jujukan. Nilai E ialah nilai jangkaan daripada skor yang diperolehi dalam pangkalan data berdasarkan peluang. Semakin rendah nilai E, semakin nyata nilai itu. Dalam erti kata lain, apabila nilai E semakin menghampiri sifar, semakin banyak persamaan antara jujukan yang ingin dibezakan dengan jujukan prob yang digunakan. Skor bit pula merujuk pada skor yang diperolehi daripada darjah homolog antara jujukan yang ingin dibezakan dengan jujukan asal. Semakin tinggi nilai skor, semakin banyak persamaan antara jujukan. Di dalam kajian ini, pemilihan jujukan adalah berdasarkan nilai E yang rendah dan bit skor yang tinggi (Altschul *et al.*, 1990). Nombor akses pula merujuk pada nombor rujukan jujukan tersebut pada pangkalan data genom manusia. Terdapat dua nombor akses yang berlainan digunakan. Nombor akses yang bermula daripada NT merujuk pada nombor akses yang digunakan dalam pangkalan data NCBI manakala nombor akses yang bermula daripada Hs merujuk pada nombor akses yang digunakan dalam pangkalan data genom manusia (Altschul *et al.*, 1990).



2.1.2 Pembinaan Pokok Filogenetik

Filogeni bermaksud sejarah evolusi sesuatu spesies atau sekumpulan spesies organisma. Pengelasan filogeni merupakan satu sistem pengelasan semulajadi yang menunjukkan hubungan evolusi dan juga mencerminkan sejarah evolusi bagi organisma yang dikaji. Filogeni berasal dari perkataan *phylon* dan *geny* daripada bahasa Greek. *Phylon* bermakna bangsa, keluarga ataupun kelas. *-geny* pula merupakan singkatan daripada *geneia* bermaksud titik mula ataupun sumber (origin).

Terdapat beberapa andaian yang diguna pakai oleh sistem ini. Andaian pertama menunjukkan organisma-organisma yang mempunyai ciri-ciri homolog mempunyai hubungan evolusi yang rapat. Andaian kedua pula menunjukkan ciri homolog merupakan perkembangan daripada leluhur yang sama tetapi perbezaan berlaku disebabkan penyesuaian terhadap persekitaran. Ini menyebabkan terdapat organisma yang mempunyai struktur homolog yang sama meskipun melakukan aktiviti yang berbeza (Woese, 2000).

Pokok filogenetik adalah cara termudah untuk membuktikan sejarah evolusi kehidupan. Pokok filogenetik adalah carta bercabang-cabang seperti cabang pokok tumbuhan. Pengelasan filogeni menunjukkan sistem yang mencari ciri-ciri homolog yang menyokong hubungan evolusi sesuatu spesies atau sekumpulan spesies organisma.



Dahan (*branch*) menunjukkan hubungan di antara taksa asal dengan pecahannya. Panjang dahan (*branch length*) menunjukkan perubahan yang muncul pada dahan. Semakin panjang dahan ini, semakin lama tempoh evolusi bagi organisma tersebut. Skala jarak (*distance scale*) pula menunjukkan skala yang menunjukkan bilangan perubahan antara spesies dengan jujukan. *Clade* pula menunjukkan sekumpulan dua atau lebih taksa atau jujukan DNA yang melibatkan leluhur sama (Brown, 1999).

2.2 Projek Genom Manusia

Projek Genom Manusia telah dimulakan pada tahun 1990 dan siap sepenuhnya pada 14 April 2003 (Initial Human Genome Sequencing Consortium, 2001). Projek telah dijalankan selama 13 tahun ini dikoordinasikan oleh Institut Kesihatan Kebangsaan dan Jabatan Tenaga Amerika Syarikat. Projek Genom Manusia turut mendapat kerjasama daripada negara-negara lain seperti Negara Jepun, Perancis, Jerman, China dan United Kingdom (Initial Human Genome Sequencing Consortium, 2001).

Ianya bertujuan untuk mendapatkan sepenuhnya jujukan DNA di dalam genom manusia. Dianggarkan terdapat 3 ribu juta pasangan bes dan di antara 20 000 hingga 25 000 gen yang tersimpan di dalam genom ini. Genom merujuk pada maklumat yang tersimpan dalam semua set kromosom DNA dan secara amnya merujuk kepada semua gen yang membentuk manusia (Karp, 2002). Selain itu, maklumat - maklumat yang diperolehi akan disimpan di dalam pangkalan data dan seterusnya dapat meningkatkan kemudahan untuk analisis data.



Teknologi-teknologi baru turut dikaji bagi penskrinan polimorfisme dalam gen-gen yang berlainan. Ini boleh membantu menolong individu-individu yang berisiko tinggi yang penyakit dimana gen-gen yang terlibat adalah lebih kompleks. Berdasarkan maklumat yang tersimpan dalam genom, ianya boleh membantu individu ini seperti mengikut diet yang telah ditetapkan dan latihan bagi mengurangkan risiko penyakit (Karp, 2002).

2.3 Penemuan elemen pindah (TE)

Elemen boleh pindah (*Transposable element* atau lebih dikenali sebagai TE) ditemui pada pertama kalinya oleh Barbara McClintock, seorang pakar genetik yang mengkaji penyakit jagung, pada tahun 1961. Beliau menggelarkan TE sebagai elemen pengawal kerana beliau mendapati elemen ini boleh muncul dan hilang daripada lokasi genetik individu dan mengubah kapasiti untuk pengekspresan bagi lokasi genetik yang terlibat. Beliau merumuskan bahawa terdapat beberapa elemen genetik tertentu yang boleh berpindah di dalam kromosom pada tapak-tapak yang berlainan. Beliau menamakan mekanisme penyusunan genetik ini sebagai transposisi (Shapiro, 1999).

Pada lewat tahun 1960, terdapat beberapa makmal yang menjumpai bahawa terdapat jujukan DNA tertentu di dalam bakteria yang mempunyai keupayaan untuk berpindah daripada satu kromosom kepada kromosom lain di dalam sesuatu genom



organisma. Elemen ini dikenali sebagai ‘transposons’. Seperti yang telah dibuktikan oleh McClintock, genom eukariotik mengandungi sejumlah besar TE (Karp, 2002).

Sebelum 1970, semua maklumat ujikaji genetik dipercayai konsep ‘central dogma’ iaitu transkripsi bahan genetik dari DNA ke RNA. Walaubagaimanapun, pada tahun 1970, kajian ke atas virus tumor RNA menunjukkan ‘central dogma’ tidak menerangkan keseluruhan mekanisma transkripsi tersebut. Eksperimen yang dilakukan secara berasingan oleh Howard Temin dan David Baltimore secara berasingan menunjukkan enzim transkriptase berbalik boleh mensintesis DNA daripada RNA. Pada peringkat awalnya, para saintis tidak menerima mekanisma ini kerana bertentangan dengan teori ‘central dogma’ (Karp, 2002).

Penemuan enzim transkriptase berbalik memberi impak sains yang besar dalam perbagai cara. Kebolehan menukarkan mRNA pada DNA membenarkan pembinaan perpustakaan cDNA. cDNA ialah DNA pelengkap yang tidak mengandungi introns. Ini memudahkan pengklonan dan kajian ke atas gen dalam pelbagai aspek biologi. Selain turut merangsang kajian terhadap retrovirus, virus RNA yang bereplikasi melalui transkripsi berbalik, gen transkriptase berbalik juga diperlukan untuk replikasi beberapa kelas TE dan beberapa virus tumbuhan dan haiwan (Dej *et al.*, 1998).

Semenjak penemuan TE dalam jagung oleh Barbara McClintock pada tahun 1956, terdapat banyak kajian dan debat mengenainya. Pandangan awal mengenai TE ialah TE adalah DNA parasit yang bersifat ‘*selfish*’ dengan sedikit fungsi yang relevan pada



perumahnya (Medstrand *et al.* 2002). TE adalah satu contoh DNA yang tidak mengkodkan sebarang gen berfungsi. Sebanyak 97% daripada genom vertebrata telah dilaporkan terdiri daripada genom yang tidak mengkod. TE memainkan peranan yang penting dalam membentuk genom mamalia. Teori ini disokong penuh oleh projek genom manusia apabila ia lengkap sepenuhnya pada tahun 2003. Ia membentuk 45% daripada DNA manusia (Deininger *et al.*, 2003).

Ia boleh dikelaskan kepada dua kelas berdasarkan struktur jujukan DNA dan mekanisme transposisinya (Deininger *et al.*, 2003). Kelas pertama TE terdiri daripada retroposons dan retrotransposons. Ia juga dikenali sebagai retroelemen. Mekanisma kelas ini memerlukan perantara RNA. Kelas kedua TE pula terdiri daripada jujukan penyisipan (elemen IS) dan transposons. Ia dikenali sebagai elemen DNA dan mekanisme kelas ini memerlukan perantara DNA (Deininger *et al.*, 2003).

2.4 Pengkelasan Retroelemen

'Retroelement' adalah elemen genetik boleh pindah yang mengandungi gen transkriptase berbalik. Elemen ini boleh wujud sama ada sebagai molekul DNA atau RNA. Ia boleh dipindahkan berganda melalui proses yang dikenali sebagai retrotransposisi atau retroposisi. Elemen ini mengalami transkripsi berbalik daripada RNA kepada DNA dan dipindahkan pada kromosom lain di dalam genom sesuatu organisma.



Retroelemen boleh ditemui dengan mudah daripada yis kepada haiwan. Ia boleh dibahagikan kepada dua kelas iaitu kelas virus dan bukan virus. Pengelasan ini bergantung pada kehadiran atau ketidakhadiran terminal ulangan panjang (*Long Terminal Repeat* atau lebih dikenali sebagai LTR). Kelas virus boleh dibahagikan kepada dua kelas iaitu retrovirus dan peraretrovirus manakala bukan virus boleh dibahagikan kepada tiga kumpulan iaitu LTR retrotransposons, bukan LTR transposons dan elemen primitif seperti retron, plasmid mitokondria dan kumpulan II introns (Deininger dan Batzer, 2002).

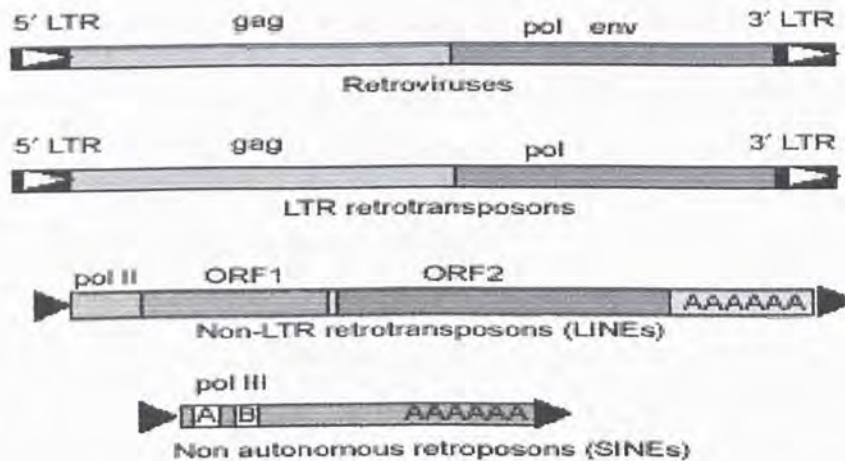
Struktur LTR retrotransposons adalah sama seperti retrovirus cuma ia tidak mempunyai gen bersampul (*env*) berfungsi yang diperlukan untuk membina kapsul virus. Ia masih mempunyai jujukan transkripsi yang teratur dan tapak prob untuk memulakan transkripsi berbalik dan beberapa rangka bacaan terbuka (ORF) yang mengkod protein untuk retrotransposisi. Protein ini termasuklah domain untuk enzim endonuklease yang bertanggungjawab untuk memotong tapak integrasi genom dan memulakan transkripsi berbalik (Deininger dan Batzer, 2002).

Elemen bukan LTR retrotransposons pula dicirikan berdasarkan struktur teras protein (*gag*) dan dua enzim; enzim transkriptase berbalik (*rt*) dan enzim integrase (*int*). Ciri lain bagi elemen bukan LTR ialah ia tidak mempunyai jujukan LTR dan mempunyai satu ekor A (poli A) pada hujung 3' sebagai tapak prob untuk transkripsi. Elemen bukan LTR boleh dibahagikan kepada dua kelas iaitu Long Interspersed Nuclear Element (LINE) dan Short Interspersed Nuclear Element (SINE) (Deininger dan Batzer, 2002). Pengelasan retroelemen boleh dilihat dengan jelas pada rajah 2.1.



(LINE) dan Short Interspersed Nuclear Element (SINE) (Deininger dan Batzer, 2002).

Pengelasan retroelemen boleh dilihat dengan jelas pada rajah 2.1.

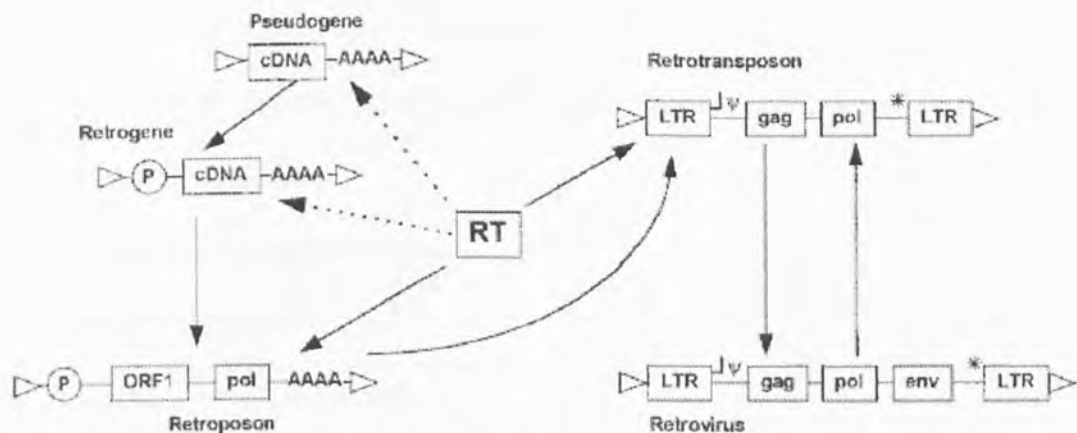


Rajah 2.1 : Pengelasan retroelemen. LTR retrotransposon tidak mempunyai pembalut (env). LINE mempunyai ORF1 dan ORF 2. SINE dan LINE mempunyai poli A pada hujung 3'. (Sumber : Deininger *et al.*,2003)

Temin mencadangkan idea yang menyatakan retroelemen berubah seiring dengan perubahan gen transkriptase berbalik. Komposisi retroelemen yang hadir dalam eukariot yang berlainan, termasuklah manusia, dipengaruhi oleh aktiviti-aktiviti enzim seperti RNase H, Integrase dan Protease (Lower *et al.*,1996).

Retrotransposons berkembang dalam pelbagai organisma daripada protozoa kepada manusia. Di dalam elemen ini, gen transkriptase berbalik dihubungkan dengan gen yang mengkodkan poliprotein yang boleh bergabung sesama sendiri untuk membentuk teras partikel. Protein ini sama dengan kapsid protein retroviral yang

biasanya dibina oleh kumpulan spesifik antigen (*'group-specific antigen'* atau lebih dikenali sebagai Gag). Elemen ini diapit oleh terminal ulangan panjang (*'Long Terminal Repeats'* atau LTR), yang bertindak sebagai tapak jujukan promoter. Retrotransposons juga digandakan pada jumlah yang banyak dan kebanyakan daripada elemen ini tidak berfungsi. Kajian retroposons telah banyak dijalankan terutama di dalam yis, *Drosophila* dan juga tikus. Retrotransposons boleh muncul sebagai terbitan ataupun leluhur kepada retrovirus (Rajah 2.2)(Lower *et al.*, 1996).



Rajah 2.2 : Evolusi retroelemen. Anak panah putus-putus menunjukkan penerbitan elemen genetik melalui transkripsi berbalik. Anak panah penuh pula menunjukkan penerbitan elemen baru dengan kombinasi gen transkripsi berbalik, ORF, gag, pol, env, LTR, promoter (P) dan poly (A) (Lower *et al.*, 1996)

BAB 3

KAEDAH

Terdapat 4 langkah utama dalam kajian ini

- 1) Mendapatkan jujukan transkripsi berbalik elemen LTR retrotransposons yang bertindak sebagai prob
- 2) Penyaringan daripada bank gen
- 3) Penyusunan jujukan
- 4) Pembinaan pokok filogenetik

3.1 Mendapatkan jujukan transkriptase berbalik elemen LTR retrotransposons yang bertindak sebagai prob

Langkah pertama di dalam kajian ini, adalah mendapatkan jujukan transkriptase berbalik elemen LTR retrotransposons yang bertindak sebagai prob. Prob adalah jujukan pendek yang digunakan untuk memancing keluar jujukan retroelemen daripada pangkalan data projek genom manusia. Prob digunakan untuk mengesan kehadiran jujukan transkriptase berbalik LTR retrotransposons dalam genom manusia. Prob yang



Rujukan

- Alberola, T. M. and Frutos R. D., 1996. Molecular structure of a gypsy element of *Drosophila subobscura* (gypsyDs) constituting a degenerate form of insect retroviruses. *Nucleic Acid Research* **24** : 914 – 923
- Altschul, S. F., Gish, W., Miller, W., Meyers, E. W. dan Lipman, D. J., 1990. Basic Local Alignment Search Tool. *J.Mol Bio* **215** : 403 – 410
- Attwood, T. K. dan Parry-Smith, D.J., 1999. *Introduction to bioinformatics*, Prentice Hall
- Barner, R. M. dan Ian C., 1999. *Bioinformatics for Geneticists*. Glaxomithklie Pharmaceuticals. United Kingdom.
- Brown, T. A., 1999. *Genomes*. John Wiley & Sons (Asia) Pte Ltd.
- Campbell, N. A., Reece, J. B. dan Mitchell, L. G., 1999. *Biology*. Addison Wesley Longman, Inc
- Deininger, P.L. dan Batzer, M.A, 2002. Mammalian retroelements. *Genome Research* **12** : 1455– 1465
- Deininger, P.L, Moran, J. V., Batzer, M.A, Kazazian H. H. Jr., 2003. Mobile elements and mammalian genome evolution. *Current Opinion in Genetics & Development* **13** : 651 = 658
- Dej, K. J., Gerasimova, T., Corces, V. G. dan BoekeI, J. D., 1998. A hotspot for the *Drosophila* gypsy retroelement in the ovo locus. *Nucleic Acids Research*. **26** : 4019–4024



- Higgins, D.G., Thompson, J.D. dan Gibson, T.J, 1996. Using CLUSTAL for multiple sequence alignments. *Method Enzymol* **266** : 383 - 402
- Initial Human Genome Sequencing Consortium, 2001. Initial sequencing and analysis of the human genome. *Nature* **409**
- Kambol, R., 2003. Distribution and evolution of retrovirus within amphibian and piscines hosts. Unpublished PhD thesis. Department of Biological Sciences, Imperial College, London, United Kingdom
- Karp, G., 2002. *Cell and molecular biology. Concepts and experiments*. 3rd edition. John Wiley & Sons, Inc, New York, 397 – 493.
- Kim, H. S., Hyun, B. H., Choi, J. Y. dan Crow, T. J., 2000. Phylogenetic analysis of a retroposon family as represent on a human X chromosome *Syst.Genes Genet.* **75** : 197 – 202
- Llorens, C. Dan Marin, I., 2001. A mammalian gene involved from the integrase domain of an LTR-retrotransposon. *Mol.Biol.Evol* **18** : 1597 – 1600
- Lower, R., Lower, J. dan Kurth, R.,1996. The viruses in all of us: Characteristics and biological significance of human endogenous retrovirus sequences. *Proc. Natl. Acad. Sci.* **93**: 5177–5184
- Majerus, M., Amos, W. dan Hurst, 1996. *Evolution. The Four Billion Year War*. Addison Wesley Longman Limited.
- Malik, H. S., dan Eickbush, T. H., 1999. Modular Evolution of the Integrase Domain in the Ty3/Gypsy Class of LTR Retrotransposons. *Journal of Virology* **73**: 5186–5190



- Medstrand, P., de Lagemaat, L. N. V. dan Mager, D. L., 2002. Retroelement distributions in the human genome : variations associated with age and proximity to genes. *Genome research* **12** : 1483 – 1495
- Morgenstern, B., Dress, A. dan Werner, T., 1996. Multiple DNA and protein dequence alignment based on segment-to-segment comparison. *Applied Mathematics* **93** : 12098-12103
- Plant, E. P., Wang, P., Jacobs P. L. dan Dinman, J. D., 2004. A programmed ± 1 ribosomal frameshift signal can function as a cis-acting mRNA destabilizing element. *Nucleic Acids Research* **Vol. 32**: 784 - 790
- Shapiro, J. A., 1999. Transposable elements as the key to a 21st century view of Evolution. *Genetica* **107** : 171 – 1799
- Springer, M. S. dan Britten R. J., 1993. All Phylogenetic Relationships of Reverse Transcriptase and RNase H Sequences and Aspects of Genome Structure in the Gypsy Group of Retrotransposons. *Mol. Biol Ed.* **10** : 1370- 1379.
- Thompson, J. D., Higgins, D. G. dan Gibson, T. J., Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**: 4673-80
- Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W. dan Mural, R. J., 2001. The Sequence of the Human Genome. *Science* **291**: 1304-1351
- Woese, C. R., 2000. Interpreting the universal phylogenetic tree. *PNAS* **97**, 8392 – 8396

