Viruses within animal genomes

A. de Brogniez & L. Willems

Molecular Biology (GxABT) and Molecular and Cellular Epigenetics (GIGA), University of Liege, Belgium
E-mail: Alix.deBrogniez@ulg.ac.be; Luc.willems@ulg.ac.be

Summary
Viruses and their hosts can co-evolve to reach a fragile equilibrium that allows the survival of both. An excess of pathogenicity in the absence of a reservoir would be detrimental to virus survival. A significant proportion of all animal genomes has been shaped by the insertion of viruses that subsequently became ‘fossilised’. Most endogenous viruses have lost the capacity to replicate via an infectious cycle and now replicate passively. The insertion of endogenous viruses has contributed to the evolution of animal genomes, for example in the reproductive biology of mammals. However, spontaneous viral integration still occasionally occurs in a number of virus–host systems. This constitutes a potential risk to host survival but also provides an opportunity for diversification and evolution.

Keywords

Introduction
Viruses are biological entities that use the host cell machinery to replicate. Viruses can either integrate their genetic material into the host cell genome or replicate autonomously within cellular compartments such as the cytoplasm. For one family of RNA viruses, the Retroviridae (or retroviruses), chromosomal integration is an essential step of the viral replication cycle (1). Endogenous viral elements (EVEs) are evolutionary remnants of viruses that have stably integrated into the host genome and are now transmitted through the germ line (2). DNA viruses of the Paroviridae (3), Herpesviridae (4), Hepadnaviridae (5) and Adenoviridae families (6) also occasionally undergo chromosomal integration. This review will describe the different groups of integrated viruses and the possible implications of virus integration for human and animal health.

Retroviruses
Retroviruses encode a reverse transcriptase (RT) that polymerises DNA using the viral genomic RNA as a template and an integrase (IN) that inserts the proviral DNA into the host cell genome (1). Infection by retroviruses can be asymptomatic (e.g. human foamy virus) or can cause diseases such as cancer and immunodeficiency (7). Retroviruses can infect a broad variety of vertebrates from fish to humans, leading to disease, for example leukemia in cattle (8). In humans, retroviruses such as human immunodeficiency virus (HIV) induce immunosuppression (e.g. acquired immunodeficiency syndrome). Another retrovirus, human T-lymphotropic virus type 1 (HTLV), causes adult T-cell leukaemia and neurodegenerative diseases (e.g. HTLV-associated myelopathy/tropical spastic paraparesis) (9).

Retroviruses comprise a large and diverse family of enveloped viruses, characterised by a viral particle of 80–100 nm in diameter and an RNA genome of 7–12 kilobases. The retrovirus genome is linear, unsegmented and single-stranded and has positive polarity. Retroviruses were originally classified as simple or complex based on their genomic organisation. Simple viruses contain four major coding regions: gag, pro, pol and env (Fig. 1). The gag gene directs the synthesis of the matrix (MA), capsid (CA) and nucleocapsid (NC) proteins that form the virion core. These proteins are cleaved from a precursor by the viral protease, encoded by the pro gene. Pol contains the genetic information for two replication enzymes: RT and IN. The viral envelope is formed by components of the host cell membrane in association with surface (SU) and transmembrane (TM) proteins encoded by the env gene. Complex retroviruses additionally encode a series of regulatory proteins involved in transcriptional activation (Tax), RNA processing (Rex, R3) or cellular transformation (Tax and G4) (Fig. 1).

More recently, retroviruses were subdivided into seven genera based on their evolutionary relatedness (Table I) (1).
Retroviral replication cycle

A key step in the retroviral replication cycle is the integration of viral DNA into the host cell genome. There is some evidence that genomic integration is an essential stage in the retroviral life cycle. Firstly, non-integrated retroviral DNAs do not replicate autonomously, for example as episomes (extrachromosomal circular DNA). Viral DNA integration is therefore necessary to stably maintain the virus in dividing cells. Secondly, integration is important for the efficient transcription of viral messenger and genomic RNAs.

The retroviral replication cycle can be divided into several steps (shown in Fig. 2). Firstly, specific interactions between envelope glycoproteins and host cell receptors determine viral tropism. Through these interactions, the retroviral envelope attaches and then fuses with the host cell membrane, releasing the viral core into the host cell cytoplasm. After viral entry and viral particle uncoating, the RT enzyme becomes active. The primers for reverse transcription are then used to synthesize full-length viral DNA, which is then integrated into the host cell genome. After integration, the provirus is transcribed into genomic and mRNA, which are then translated into viral proteins. These proteins are then packaged with genomic RNA into new viral particles that 'bud' off from the cell.

Table I

<table>
<thead>
<tr>
<th>Genus</th>
<th>Examples</th>
<th>Viral particle morphology</th>
<th>Genome type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alpharetrovirus</td>
<td>Rous sarcoma virus, Avian leukosis virus, Mouse mammary tumor virus</td>
<td>Central, spherical core</td>
<td>Simple</td>
</tr>
<tr>
<td>Betaretrovirus</td>
<td>Mason-Pfizer monkey virus, Jaagsiekte sheep viruses</td>
<td>Eccentric, spherical core</td>
<td>Simple</td>
</tr>
<tr>
<td>Gammaretrovirus</td>
<td>Murine leukaemia virus, Feline leukaemia virus, Gibbon ape leukaemia virus</td>
<td>Central, spherical core</td>
<td>Simple</td>
</tr>
<tr>
<td>Deltaretrovirus</td>
<td>Bovine leukaemia virus, Human T-lymphotropic virus, Simian T-lymphotropic virus</td>
<td>Central, spherical core</td>
<td>Complex</td>
</tr>
<tr>
<td>Epsilonretrovirus</td>
<td>Walleye dermal sarcoma virus</td>
<td>Cylindrical core</td>
<td>Complex</td>
</tr>
<tr>
<td>Lentivirus</td>
<td>Human immunodeficiency virus</td>
<td>Cylindrical core</td>
<td>Complex</td>
</tr>
<tr>
<td>Spumavirus</td>
<td>Human foamy virus</td>
<td>Central, spherical core</td>
<td>Complex</td>
</tr>
</tbody>
</table>
transcription of viral genomic RNA are cellular transfer RNAs (tRNAs) that hybridise with the primer-binding site of the 5′ long terminal repeat (5′LTR). The RT enzyme also has RNase H activity, which removes the parental RNA strand from the complementary DNA (cDNA). After strand transfer to the 3′LTR, strand synthesis of the proviral genome is completed. Strand synthesis is then initiated at the polypurine tract using an RNA primer that is not digested by RNase H. Double-stranded proviral DNA then forms a complex with viral (IN) and cellular proteins, enters the nucleus and integrates into the host chromosome.

Mechanism of proviral integration into the chromosome

After reverse transcription is completed, linear double-stranded viral DNA (flanked by two LTRs) is incorporated into a preintegration complex (PIC) that migrates to the nucleus. Some retroviruses only infect dividing cells because their cDNA products can only interact with chromosomal DNA during mitosis when the nuclear membrane is disrupted (12, 13). In contrast, spumaviruses and lentiviruses can infect both dividing and resting cells because they can enter the nucleus through an active yet poorly understood mechanism (14) involving the PIC and the nuclear pore complex (15, 16).

The retroviral integration mechanism has three main steps (Fig. 3):

i) After the completion of viral DNA synthesis, both proviral ends contain 5–15 base pair extensions that are recognised by the IN protein complex. The IN enzyme cleaves the 3′ termini of both proviral DNA ends, thus eliminating the two terminal bases. The resulting 3′ hydroxyl (OH) groups are involved in proviral attachment to the host DNA. This first step is the 3′ processing reaction that forms a chemically activated viral DNA with 3′ OH radicals at the terminal ends of the viral DNA.

ii) The second step is a transfer reaction involving transesterification. Firstly, the IN protein–viral DNA complex binds to the host cell DNA and introduces an asymmetrical nick of 4–6 nucleotides in length, depending on the retroviral protein structure. In the transfer reaction, the energy released by breaking phosphodiester bonds in target DNA is used to form new bonds joining the viral 3′ ends to the target DNA.

iii) Finally, the host DNA repair machinery cleaves the protruding viral 5′ nucleotides and fills in the 4–6 base pair gap, leading to duplication of the gap nucleotide sequence surrounding the provirus. A final ligation step completes the proviral integration process.

Preferred integration sites

The site of proviral insertion into the host chromosome can determine the outcome of the virus infection cycle. Integration into a transcriptionally active region favours viral gene expression and exposure to the host immune response. In contrast, integration into repressed chromatin facilitates viral latency. However, retroviral integration at a site nearby or inside cellular genes can affect host transcription, leading to tumorigenesis via a mechanism called ‘insertional mutagenesis’. Differences in their integration site preferences cause retroviruses to have different genotoxic potentials (e.g. gammaretroviruses are more prone to insertional mutagenesis than lentiviruses).

Sequencing the sites flanking proviral DNA insertion has shown that retroviral integration is not random (17, 18, 19). For example, HIV integration occurs within active transcription units characterised by high G+C content, high CpG island density, high gene density, short introns and high frequencies of Alu repeats (20, 21, 22, 23). Gammaretroviruses, spumaviruses and endogenous retroviruses are preferentially integrated around transcription start sites and within CpG islands (24, 25, 26). Alpharetroviruses display only a weak preference for insertion into CpG islands and transcription units (20, 27), while betaretrovirus integration is random (28). Deltertroviruses, which include HTLVs, simian T-cell lymphotropic viruses and bovine leukaemia virus (BLV), initially target actively transcribed regions of the genome. Deltertroviruses are preferentially inserted into a palindromic consensus sequence with transcriptional units (29, 30). During primary infection, most infected cell clones are depleted by the host immune response (27). At this stage, the proviral load is maintained by a change in
the mode of viral propagation that favours the proliferation of pre-existing clones of infected cells. The site of viral integration thus influences the proliferative potential of infected cells. The proviral loads are therefore determined by clone abundance and limited by the host immune response.

Accurate quantification by high-throughput sequencing has demonstrated that integration site preference is neither specific to the host cell nor dependent on the viral entry route. Different models have been proposed to explain the integration site preference of retroviruses (31). It is possible that the chromatin structure (relaxed or condensed) influences the accessibility of target DNA sequences to the PIC. The cell cycle phase may also modulate site-specific integration; moreover, it is possible that for each retroviral genus cellular proteins act as tethering factors that simultaneously interact with specific chromatin sites and PICs.

Consequences of viral integration
Viral infection can have opposing outcomes on the cell fate, i.e. proliferation or death (32). Indeed, strand breaks
generated during viral integration activate the DNA repair machinery, which can be overloaded by an excess of infection events. In contrast, apoptosis may also result from the deregulation of host cell metabolism (33).

If the cell survives infection, the virus can be either actively expressed and modulate the cell transcriptome in trans or silenced and thus enter latency. Some retroviruses encode viral oncoproteins that directly inhibit cell death or stimulate cell proliferation (34). Cell fate therefore depends on the viral activation state. For example, HIV infection of resting or activated lymphocytes leads to pyroptosis or caspase-9 dependent apoptosis, respectively. Integration can also lead to insertional mutagenesis that disrupts genes or alters cellular transcription in cis. Viral integration at sites near to oncogenes or tumour suppressor genes can also lead to oncogenesis (35).

Another outcome of viral integration is transcriptional silencing and latency to form ‘viral reservoirs’ that escape host immunity. Several factors influence entry into latency: the availability of cellular transcription factors, the expression of viral proteins that activate transcription and the integration site in the host cell genome. Integration into transcriptionally inactive heterochromatin is reported to favour viral latency (36). Thus, the formation of viral reservoirs can be controlled by epigenetic mechanisms such as lysine acetylation of histones (4, 37).

Endogenous viral elements

Viruses other than retroviruses have also become integrated into host cell genomes. EVEs are retrotransposons that evolved from viruses and are vertically transmitted through germ cells (38). Although they are mostly latent, EVEs can be reactivated by specific environmental conditions (e.g. irradiation), leading to the production of new infectious viral particles. The uncontrolled spread of these viruses can be hazardous to health (39). Nevertheless, EVEs are essential drivers of evolution that act by promoting genomic diversity.

The RepBase database has compiled EVEs into six families containing more than 200 consensus sequences (40, 41). LTR-containing members include simple retrotransposons and endogenous retroviruses. Non-LTR retroelements comprise long and short interspersed nuclear elements (known as LINEs and SINEs, respectively) (42). These retrotransposons and their evolutionary remnants constitute a large proportion (45%) of the human genome (43). Besides retroviruses (44), EVEs are also derived from RNA viruses; for example, the flavivirus-related endogenous element has been integrated into the Aedes mosquito genome (45). Other examples include EVEs derived from bornaviruses (46), filoviruses (47), orthomyxoviruses (48), reoviruses and rhabdoviruses (48). Even small DNA viruses such as circoviruses and paroviruses have entered animal genomes during evolution (3, 49). In paleovirology, EVEs derived from germ-line integration events that occurred millions of years ago can provide historical information about ancient viruses (48). The analysis of EVEs gives an estimate of the time since divergence of orthologue-containing host species groups and thus provides insight into long-term viral evolution.

Like retroviruses, EVEs are generally latent but can be reactivated by epigenetic mechanisms such as cytosine methylation at CpG dinucleotides or acetylation of histone lysines (50). Although reintegration of activated EVEs can pose a threat to the host genome, the resulting mutations can provide a selective advantage during host evolution. For example, the syncytin protein essential for placental morphogenesis in mammals is derived from a retroviral envelope protein (51, 52, 53). In humans, syncytin is encoded by an endogenous retrovirus ERVWE1 on chromosome 7.

The mechanism for stable integration of a non-retroviral RNA element into chromosomal DNA has been demonstrated for lymphocytic choriomeningitis virus (LCMV) (31, 54). LCMV is transcribed in the cytoplasm by the virally encoded RNA-dependent RNA polymerase (RdRp). During reverse transcription, LCMV RNA recombines with intracisternal A-type particle (IAP) elements, and the resulting cDNA undergoes random integration (55).

Integration of DNA viruses

DNA viruses can sometimes be integrated into the host genome without the need for a processing step such as reverse transcription. The genome of DNA viruses can translocate into the nucleus and persist in an episomal form. Integration events into host chromosomes have occasionally been reported for viruses of different families: Paroviridae, Herpesviridae, Hepadnaviridae and Adenoviridae.

Adeno-associated virus type 2 (AAV2) is a widespread non-pathogenic virus of the Paroviridae family that requires a helper virus, such as adenovirus, human papillomavirus or herpes simplex virus, to complete its replicative cycle. In the absence of a helper virus, AAV integrates at specific regions of the host genome known as AAV integration site (AASVs) loci, e.g. AAVS1 on chromosome 19, AAVS2 on chromosome 5 and AAVS3 on chromosome 3 (56). Site-specific AAV-2 integration into non-repetitive elements located in gene-dense regions involves two specific host genome sequences: the terminal resolution site (TRS) and the Rep-binding site (RBS) (57, 58). TRS and RBS are
predicted to recombine with homologous sequences of the viral genome. Thus, AAVS1 provides a safe site for AAV integration.

Although herpesviruses persist as double-stranded DNA episomes in the nucleus, they are occasionally integrated into the host cell genome. Stable integration of Epstein-Barr virus (EBV) ensures its long-term persistence (59, 60, 61). EBV integration is non-random: heterochromatin regions are preferentially targeted (59, 62). Herpesvirus 6 (HHV-6) also undergoes chromosomal integration (ciHHV-6), probably by homologous recombination between viral and cellular telomeric sequences (4, 63).

About one million people die each year from hepatitis B virus (HBV)-induced hepatocellular carcinoma, liver cirrhosis and other complications (64). Although usually present in an episomal form, HBV can integrate into the host chromosome during acute infection by a mechanism similar to retroviral insertion. HBV integration preferentially targets genes involved in cell survival, proliferation and oncogenesis in transcriptionally active regions (65, 66, 67).

Although adenoviruses replicate extrachromosomally in the nucleus, integration occasionally occurs by homologous recombination and non-homologous end joining (6, 68).

Conclusion

Animal genomes contain a large proportion of endogenous viruses that are generally unable to undergo an infectious cycle but replicate passively via mitotic division. Of the exogenous viruses, retroviruses are designed to integrate into the host genome and therefore constitute a potential risk of mutation. Other types of viruses with either RNA or DNA genomes can also occasionally be integrated into the host genome.

Acknowledgements

The work was supported by the ‘Fonds National de la Recherche Scientifique’ (FNRS), the Interuniversity Attraction Poles BELVIR programme (initiated by the Belgian Science Policy Office), the Belgian Foundation against Cancer, the ‘Centre anticancéreux près l’Université de Liège’ (ULg), the ‘AgricultureIsLife’ project of Gembloux Agro-Bio Tech, and the ULg ‘Fonds Spéciaux pour la Recherche’. AdB is currently a postdoctoral researcher at the Télévie. LW is research director of the FNRS. The authors thank Arsène Burny for his helpful comments.

Les virus présents dans les génomes animaux

A. de Brogniez & L. Willems

Résumé

Les virus et leurs hôtes ont démontré leur capacité à co-évoluer pour atteindre le fragile équilibre qui assure leur survie mutuelle. Une pathogénicité excessive sans réservoir disponible peut compromettre la survie d’un virus. Un nombre proportionnellement significatif de génomes animaux ont vu leur structure modifiée par l’insertion de virus qui se sont par la suite « fossilisés ». La plupart des virus endogènes ayant perdu leur aptitude à se répliquer via le déclenchement d’un cycle infectieux, leur réplication s’effectue désormais de manière passive. L’insertion de virus endogènes dans les génomes animaux a contribué à les faire évoluer, comme l’illustre la biologie de la reproduction des mammifères. Néanmoins, des intégrations spontanées de virus continuent de se produire ponctuellement dans certains systèmes virus-hôtes. Elles représentent un risque potentiel pour la survie de l’hôte mais ouvrent également de nouvelles perspectives de diversification et d’évolution.

Mots-clés

Virus dentro de genomas animales

A. de Brogniez & L. Willems

Resumen

Los virus y sus anfitriones pueden coevolucionar hasta alcanzar un frágil equilibrio que permite la supervivencia de ambos. A falta de un reservorio, una patogenicidad excesiva resultaría perjudicial para la supervivencia del virus. Hay una proporción importante de todos los genomas animales en cuya configuración ha intervenido la inserción de virus, ulteriormente «fosilizados» en el genoma. La mayoría de los virus endógenos han perdido la capacidad de replicarse por medio de un ciclo infeccioso y se replican ahora de forma pasiva. La inserción de virus endógenos ha contribuido a la evolución de los genomas animales, por ejemplo en la biología reproductiva de los mamíferos. No obstante, en muchos sistemas virus-anfitrión se sigue dando ocasionalmente una integración vírica espontánea, lo que supone a la vez un posible riesgo para la supervivencia del anfitrión y una oportunidad de diversificación y evolución.

Palabras clave


References


