Integration of the viral genome into the host cell genome: a double-edged sword

Mauro Pistello (1) and Guido Antonelli (2)

(1) Retrovirus Centre and Virology Section, Department of Translational Research, University of Pisa, Pisa, Italy
(2) Department of Molecular Medicine and Pasteur Institute - Cenci Bolognetti Foundation, "Sapienza" University of Rome, Italy

Historically, integration of viral genomes has been considered a clever way for viruses to perpetuate their life and hitchhike cells to disseminate infection within and between hosts. A wealth of new evidences show that this is only part of the story and that integration of the viral genetic material is integral part of virus-host interplay and has important, and unexpected, consequences for both players.

Until a few years ago, integration was considered a must only for retroviruses. Integration is indeed an obligatory step of retroviral replication in which the viral RNA genome is first converted to double stranded DNA by the viral-encoded reverse transcriptase, then travels across the cell cytoplasm to enters the nucleus, and is at last incorporated into the host cell genome. Integrase, a viral encoded enzyme that accompanies the genome along its travel into the cell, operates the nicks and sealing between host cell and retrovirus DNAs in host genome sites that were believed to be picked at random up to few years ago but, as well described herein by Dr. Ciuffi (1), it turned out to be directed by a number of cell and viral factors. Integration accounts for the nearly ubiquitous distribution of retroviruses and the existence of many endogenous retrovirus and retrovirus-like elements to the point that "retrovirus

signatures" are remarkably abundant in mammalian genomes, up to 8% in human genome (2).

When integration of retroviral genome into the host cell genome was first hypothesized, it was assumed to be a mechanism possibly driving to viral persistence and passively accepted by the cell. Further, depending on the site of integration and interference with flanking genes, it could also lead to detrimental effects (i.e. death or neoplastic transformation) for the cell itself (1). Subsequent studies demonstrated that this mechanism has also downsides for the retroviruses. Most proviruses (i.e. the integrated retroviral genomes) underwent progressive rearrangement and lost of genetic material that ultimately lead to the inability of these viruses to replicate (2). Indeed, most retroviral genomes in mammalian and non-mammalian lineages are relics of ancient integration events, have irremediably lost their capacity to generate infectious particles and become permanent part of host cell genome. Irreversible integration of endogenized retroviruses and endogenous retroviral elements is used to analyze evolution and speciation of vertebrates, define genetic links between animal species, and identify events of cross-specie transmission (3, 4). The initial advantage for retroviruses to ensure persistent infection of host cell is then, in the long term, shifted in favor of the infected cell that essentially neutralizes virus ability to generate progeny particles and acquires resistance to superinfection by same or similar viruses. The latter mechanism was thought to be similar to what occurs to a lysogenic bacterial cell that incorporates a phage DNA genome (prophage) into its bacterial chromosome. The prophage confers resistance to superinfection by similar phages by expressing a few proteins that repress phage promoters driving lytic replication.

Next generation sequencing and various cutting-edge biomolecular techniques have made massive DNA sequencing routine allowing to gain further details on the

intertwined cascade of events following integration of the viral genome. First and foremost, mass sequencing unveiled that integration is not limited to retroviruses but occurs for a broad spectrum of viruses and, therefore, undermined the idea that integration is an event secondary or necessary for replication of some viruses. It has long been established, for instance, that hepatitis B virus (HBV) and papillomavirus (HPV) occasionally integrate their genomes into the genome of the target cells. As for retroviruses, integration of HBV and HPV is a net lost in term of dissemination of infection; both viruses loose conspicuous parts of genetic material and become incompetent for replication. In turn, this event creates conditions for cell transformation. More recently and, to our opinion, quite unexpectedly, it has been shown that integration takes place for many viruses, even those RNA viruses that entirely replicate in the cytoplasm and do not convert their RNA genome into DNA. The scientific community is struggling to find a mechanistic model that may explain how integration occurs for, for instance, bornaviruses, filoviruses, flaviviruses, picornaviruses, rabdoviruses, etc., which are single-stranded RNA (either negative or positive polarity) genome and whose replication is entirely cytoplasmatic (3). Less surprising seen their nuclear replication, nonetheless difficult to explain, is the case of the members of the Orthomixoviridae family, whose segmented single-stranded negative polarity RNA genome, has been found integrated in insect cells (4). The same consideration also applies to various single stranded DNA viruses (circoviruses and parvoviruses) (for a comprehensive list see references 3 and 4). According to various molecular clocks, integration occurred several millions years ago and, as happened for endogenized retroviruses, the viral genomes underwent gross rearrangement with conspicuous loss of genetic material.

Analyses beyond mere pinpointing of integration sites and genome sequencing reinforced the idea that integration bears more advantages to the cell rather than the virus. As likely result of host pressure, some endogenous retroviruses were eventually lost with a decline rate that differs among mammalian lineage and is particularly fast in humans (2, 4). Among the factors that contributed to extinction of various retroviruses, cellular restriction factors are thought to have played a major role and, particularly, TRIM5 that specifically targets retroviruses via direct binding to the viral capsid after entry into the cytoplasm of the infected cell (1). This constant pressure causes progressive accumulation of adaptive changes in the viral genome that blunted, in the long run, the replicative capability of retroviruses. Recent evidences demonstrated that this phenomenon took place in various circumstances and resulted, for instance, in the ability to restrict non-human primate lentiviruses, but had no effect (positive or negative) on restriction of other retroviruses (5). In this context, particularly interesting and suggestive is the theory put forward by Raoult and colleagues and described herein (6) that a few HIV infections occurred in humans resulted in the endogenization of the viral genome that progressively reduced its fitness for replication and survival. This fascinating theory, which main outline may be arisen from the koala model described in 2006 (7), is supported by the fact that patients in whom endogenization has believed to occur have benign clinical course and bear HIV strains that are either totally unable or replicate very slowly with little consequences for the cells. It is not known whether impairment of virus ability occurred by chance or has been actively pursued by host factors (cohort analyses of genetic and innate and adaptative immune response favor the second hypothesis) and other studies are warranted to understanding the mechanistic models. Whatever the case, if molecular studies of elite controllers and long-term non-progressor patients

will confirm this theory, integration and the resulting cascade events need to be re-evaluated. Further elements suggesting that integration is actually an advantage for the cells are the discovery that integration of viral genomes provokes reshuffling of methylation pattern and chromatin of cellular DNA and that most retroviral genome remnants are involved in the regulation of essential immune functions (8) and cellular metabolic pathways. As described herein by Naville and colleagues (9), many retrovirus and transposable elements have repeatedly been used as a source of novel protein coding genes during the evolution of most eukaryotic lineages and underwent phenomenon called “molecular domestication”. In particular, cellular genes derived from *gag* and *env*, as well as from the integrase- and protease-coding sequences, intervenes in many important biological processes including placenta formation, cognitive functions in the brain and immunity against retroelements, as well as in cell proliferation, apoptosis and cancer. Finally, it was known that one of the mechanisms used by the cells to silence integrated viral genomes is methylation. What is emerging from recent studies is that methylation also extends to the recipient genome and in remote sites from the site of foreign DNA insertion. It is unclear how this phenomenon occurs and its impact on the transcription profiles of the cellular genomes (10). For some viruses, e.g. herpesviruses, it has been observed that genome modifications also include chromatin assembly, histone binding and higher-order chromosome structures. Whether this is necessary to enable gammaherpesvirus to establish stable latent infections and mediate viral pathogenesis is not clear at the moment. Within the context of this rapidly evolving area, Duncan present in this “theme section” a significant revision of human herpes virus type 6 latency and persistence issues which takes into account recent studies characterizing its chromosomal integration (11).

In conclusion, integration of the viral genome is a long-established virological fact that, as thoroughly described herein by Dr. Ciuffi, it was believed to have a precise significance for retroviruses as a whole. Recent data demonstrated that the picture is much more complicated and have unexpected and unpredictable outcomes. Many issues are still unclear and need to be addressed to understand whether recent revolutionary and, in the case of HIV, challenging theories holds true. One of the factors that certainly complicate a comprehensive grasp is timing. Integration and rearrangement of retrovirus genomes are thought to have happened in ages (million years). Similarly, molecular clocks date the "accidental" integration of single-strand RNA viruses and other viruses to thousand years ago. We don't actually know whether integration and rearrangement occurs sequentially and progressively as demonstrated by sequencing studies of integrated retroviruses in different mammalian species (2-4), or can be a concomitant event for, for instance, viruses for which integration is an accidental event rather than a prerequisite. Finally, if endogenization and domestication take ages, can this be considered viral persistence and/or host resistance mechanisms or is mere a natural evolving process which consequence is determined by casual drifting rather than an ordinate scheme? Whatever the mechanism, to study in full and without bias this old and still puzzling process we need an open mind and to think high. Thankfully, now we have at disposal such a number of techniques and technologies that allows pinpointing the various facets with an unprecedented level of detail.

Potential conflicts of interest: None

References

1. Ciuffi A. The benefits of integration. *CMI* 2016 (in press).
2. Magiorkinis G, Blanco-Melo D, Belshaw R. The decline of human endogenous retroviruses: extinction and survival. Retrovirology 2015; 12: 8.
3. Holmes EC. The evolution of endogenous viral elements. Cell Host Microbe 2011;10: 368-377.
4. Aiewsakun P, Katzourakis A. Endogenous viruses: Connecting recent and ancient viral evolution. Virology 2015; 479-480:26-37.
5. McCarthy KR, Kirmaier A, Autissier P, Johnson WE. Evolutionary and functional analysis of old world primate TRIM5 reveals the ancient emergence of primate lentiviruses and convergent evolution targeting a conserved capsid interface. PLoS Pathogens 2015; 11(8): e1005085.
6. Tamalet C, Colson P, Decroly E, Dhiver C, Ravaux I, Stein AS, Raoult D. Re-evaluation of possible outcomes of infections with human immunodeficiency virus. *CMI* 2016 (in press).
7. Tarlinton RE, Meers J, Young PR. Retroviral invasion of the koala genome. Nature. 2006; 442:79-81.
8. Chuong EB, Elde NC, Feschotte C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. Science 2016; 351:1083-1087.
9. Naville M, Warren IA, Haftek-Terreau Z, Chalopin D, Brunet F, Levin P, Galiana D, Volff JN. Not so bad after all: retroviruses and LTR retrotransposons as a source of new genes in vertebrates. *CMI* 2016 (in press).
10. Doerfler W. Epigenetic consequences of foreign DNA insertions: de novo methylation and global alterations of methylation patterns in recipient genomes. Rev Med Virol. 2011; 21: 336-346.
11. Duncan AC. Clinical and laboratory features of human herpesvirus 6 chromosomal integration. *CMI* 2016 (in press).