

Abridged version (outline) of the doctoral thesis

Name in full Wanjing Zheng

Title A Comparative Study of Host Genome Evolution in Relation to Endogenous Retrovirus Load

It is known that endogenous retroviruses (ERV) are present in all vertebrates investigated and that retrovirus infection in vertebrates has a history spanning hundreds of million years. The unique type of relationship between hosts and ERVs/retroviruses throughout the long history, which includes both conflict and co-option, may have shaped the host-parasite evolutionary interaction in vertebrates and this evolutionary interaction may differ between vertebrate groups. Herein, evolutionary interaction refers to any kind of association between changes in two subjects during evolution, including: (1) the one-way association, whereby the evolution of subject A is the cause or driving force of the evolution of subject B, which is to say, the evolution of subject B is a consequence or response to the evolution of subject A; and (2) the two-way association, which is equivalent to co-evolution. Mammals and birds differed largely in their ERV load, which is defined herein as the ERV copy number per giga base pairs (Gb) of the host genome, and host-ERV relationship may be related to this difference. This thesis will report a study aimed at contributing to understanding the host-ERV relationship during long-term evolution. This study consists of two parts.

Since the host immune system can take an important part in host-ERV evolutionary interaction, especially some innate immune receptors that have potential for recognizing retroviruses, the first part of this study is a case study of the functional evolution of innate immune receptors using the RIG-I-like receptors (RLRs) in birds. RLRs are pattern-recognition receptors for viral RNA and one of them, the retinoic acid-inducible gene I (RIG-I), is a potential sensor for retroviruses. Modes and intensity of natural selection of the coding genes of avian RLRs were examined to understand the roles of RLRs in bird evolution and bird-ERV evolutionary interaction. This part of my study provides results and discussion about the evolution of RLR genes in birds from aspects of conservation levels, positive selection modes, changes in selection intensity, and association between evolutionary rate of RLR genes and endogenous retrovirus load; many of these results will be shown and discussed in comparison with those of mammals. In brief, the three RLR genes show distinct patterns of functional evolution but with possible influences to the evolution of each other and the gene encoding RIG-I evolved in correlation with endogenous retrovirus load in bird genomes. These findings suggest the possibility of interaction between

host immunity and endogenous retroviruses in bird evolution.

The second part of my study takes a broader investigation at genome-wide scales on the evolutionary interaction between hosts and ERVs/retroviruses in mammals and birds. Phylogenetic gene-phenotype association analyses were applied to the gene evolutionary rate and ERV load, and gene set enrichment analyses (GSEA) based on the association results were performed to provide information about the relative weight of biological process in the evolutionary interaction of hosts with ERVs/retroviruses. From this study, I detected genes that evolved in association with ERV load in mammals and birds, separately, and revealed that the distribution of degrees of association between gene evolutionary rate and ERV load show a difference between mammals and birds, which indicate different levels of evolutionary interaction between mammals and birds. The genes that evolved in association with ERV load in both mammals and birds, as well as genes evolved in only one of the two groups, are reported. This part of my study also provides comparative insights into the evolutionary interaction between host genes and ERV loads in mammals and birds, with particular attention to the biological processes that have the highest potential for being host restrictions on ERV load. Such biological processes involve immune responses, gene silencing and DNA deletion. Genes showing high degrees of association between gene evolutionary rate and ERV load and involved in these biological processes are also reported and discussed. Results of this part of my study suggest that gene silencing may play an important role in host-ERV evolutionary interaction, and that mammals and birds might evolve different strategies in immune responses to ERVs/retroviruses.

More detailed abstracts for the two parts of this study are present at the beginning of Chapter 2 and 3 of this thesis, respectively. Overall, this thesis provides evidence of host-ERV evolutionary interaction in mammals and birds, proposes explanations to the ERV load difference between mammals and birds, and supports the long history of host-ERV relationships comprising of a balance between host-parasite conflict, tolerance and co-option.