Pararetroviruses (hepadnaviruses and caulimoviruses) are one of the major health threats worldwide to humans, animals and plants. Endogenous pararetrovirus sequences (EPRVs) in plant genomes, such as endogenous rice tungro bacilliform virus-like sequences (eRTBVLs) in rice genomes, are viral genomic fossil records and can help us to understand ancient virus activities and host histories.

Firstly, he collected about 300 eRTBVL segments from the genomic databases of Nipponbare (Oryza sativa japonica), 93-11 (O. sativa indica), and W1943 (O. rufipogon, eastern China wild accession). Virus sequences were reconstructed from eRTBVL segments in the three rice genomes and grouped into four distinct major families. Three of the families, eRTBVL-A, -B and -C, began endogenization during differentiation of O. rufipogon. Different proportions of these eRTBVL families were detected in the three genomes, suggesting that they had independent geographical origins. Screening of orthologous eRTBVL loci in 65 cultivated and 30 wild accessions demonstrated that the flow of eRTBVLs reflected the processes proposed previously for rice domestication. The genealogy of the putative viruses corresponding to the eRTBVL families was inferred from the results. The virus lineage of the eRTBVL-B family could be a parent for viruses of the eRTBVL-A and -C families. Further divergence might have resulted in the virus lineages corresponding to the eRTBVL-A1, -A2, and -X subfamilies; however, no more subfamilies were detected in the three rice genomes. The step-by-step virus lineages can be explained, for the most part, by recombination between the viruses corresponding to the eRTBVL families.

Secondly, orthologous eRTBVL screening in the cultivated and wild rice collection suggested that the oldest family, eRTBVL-D, had been endogenized before differentiation of the rice progenitor, O. rufipogon. Further orthologous screening in the genomic databases and DNA samples of other Oryza species revealed that the endogenization of eRTBVL-D sequences initially occurred about 5 million years ago, before the divergence of Oryza AA-genome and BB-genome species. Together with phylogenetic analysis, these results indicated that the divergence between rice tungro bacilliform virus (RTBV) and the virus of eRTBVLs could occur much earlier than the differentiation of the current major natural host of RTBV. Besides the vertical inheritance of the eRTBVL-D family in a part of Oryza species, he also identified multiple divergent eRTBVL-similar sequences (eRTBVL superfamily) in all the other grass genomes examined. The corresponding viruses of eRTBVL superfamily could be a sister genus of Tungrovirus and possessed distinct types of genomic structures. Further sequence analysis implied independent integration events of most sequences of eRTBVL superfamily in different grass genomes.
Thirdly, although all the known pararetroviruses are autonomous for the life circle in host cells, he identified unexpected multiple EPRVs derived from nonautonomous pararetrovirus species in grass genomes. The reconstructed viral genomes and endogenization profiles of these EPRVs, eGPV1s and eGPV2s, revealed that two distinct nonautonomous pararetroviruses losing part essential functions had horizontally transmitted to a wide range of grass hosts and developed into various groups and lineages during the long-term grass evolution. Comparative molecular biology and comparative genomics revealed the coding capacity complementarity and noncoding cis-regulation sequence identity between the viruses of eGPV1s and eGPV2s. Strikingly, the paleogenomic evidence support that the two virus species kept a co-evolution to develop a mutualistic symbiosis through noncoding region swapping. He further demonstrated that region-specific recombination had driven the homogenization of the noncoding regions required for the mutualism between nonautonomous pararetroviruses. The previously unknown genomic and ecological diversities of pararetroviruses revealed here supply fossil record evidence to support that not Darwinian model but symbiosis model contributes the viral rapid evolutionary events.

Finally, he identified a family of EPRVs, eRPVs, derived from a recombinant pararetrovirus in rice genomes. Comprehensive in silico sequence analyses and orthologous locus screening revealed that the illegitimate recombination between two distinct pararetrovirus species, the viruses of eRTBVLs and eGPV2s, generated the viruses of eRPVs possibly after the divergence of host rice subspecies. The viral genomic structures of eRPVs reassorted from recombination events are consistent with the modular evolution theory. The essential RT/RH (reverse transcriptase/RNase H) domain for viral replication was replaced by an inessential domain in the modular reassortment, which led the virus of eRPVs to be a nonautonomous pararetrovirus. Sequence analysis supported that this nonautonomous virus was able to replicate and even have undergone a persistent evolution during host divergence. The evidence of intergenic region sequence homogenization and co-infection suggested the parasitism of the virus of eRPVs on the virus of eRTBVLs, which proposed a symbiosis switch between pararetroviruses induced by interspecies recombination-mediated modular reassortment.

Chen’s study provides significant implications for the rice plants and virus evolutions through the endogenous virus sequences, particularly following points are highly-regarded, 1) he showed that the rice genomes have recorded only a few recombination events that nevertheless have contributed to major evolutionary lineages of the ancient pararetroviruses during rice speciation; 2) his results uncovered the extremely widespread “horizontal genome transfers” of a previously uncharacterized genus of pararetroviruses into grass genomes, of which independent viral integrations could have been shaping the genomic evolution of host grass species; 3) his study provides paleobiological insights into the evolutionary origin of complex viruses consisting of multiple sequences; 4) his results provide unique genomic fossil evidence supporting that interspecies illegitimate recombination is one of the mechanisms underlying viral modular evolution.

Taken together his study reveals the organization, endogenization, distribution and evolution of eRTBVLs, as well as of three newly identified EPRVs. These results provide new insights into pararetrovirus evolution and host grass divergence. Therefore, we acknowledge that Mr. Chen is qualified to be granted the Degree of Doctor of Philosophy in Agriculture from Hokkaido University.