Supplemental Figure Legend

Figure S1. Venn diagrams present the intersections of ERV insertion sites among different chicken breeds (Broiler, Tosa-Jidori, Yakido) and wild species (*Gallus Gallus*, *Gallus sonneratii*, *Gallus lafayettii*, and *Gallus varius*). Each circle represents one breed or species, with the numbers within the overlapping regions indicating the count of shared ERV insertions. Non-overlapping regions denote the number of unique insertions specific to each breed or species.

Figure S2. Mapping of contig sequences derived from extracted reads to known endogenous retrovirus (ERV) sequences in chicken genomes using IGV (Integrative Genomics Viewer). The contigs were mapped against the following Accession IDs: Chicken endogenous proviral avian retroviral_LTR (M31063, M31064, M31065, M31066), EAV-0 (AM418554), EAV-51 (AM418553), EAV-0/E51 (AM418555), EAV-HP (NC_005947.1, AJ623291, AF125529), ALV-E (MT263508), MLV-related endogenous retrovirus (DQ280312). M31064 was omitted because there was no mapped contig.

Figure S3. Pipeline for the detection of non-reference chicken endogenous retroviruses in whole-genome sequencing (WGS) read data. This figure is based on content taken from Ishihara 2023 [8] with modifications. In the right-upper panel, the black line denotes the chicken reference genome sequence. Blue and red boxes connected with lines denote the 5′ and 3′ ends of a paired-end sequencing read. Most paired-end reads were identified as proper mapping, whereas a small percentage of them were improper mapping. Proper pair: both ends of the paired-end sequence mapped accurately (**a**). Discordant read and split read: one end of the paired-end sequence mapped accurately, whereas the other end was only partially identified at the expected locus on the reference genome. The unidentified sequence could be mapped anywhere else on the reference genome (**b**, **c**). Singleton: one end of the paired-end sequence mapped accurately, whereas the other end did not map on the reference genome (**d**). Unmapped read pairs: neither read mapped to the reference genome (**e**). Discordant reads, split reads, and singletons were used for RetroSeq analysis. In the right-middle panel, a representative view of the Integrative Genomics Viewer version 2.8.9 (IGV) is used to confirm the presence of target site duplications (TSDs) at each locus, detected using RetroSeq, extract support reads from the TSD loci, perform local assembly, and analyze the contigs for the presence of endogenous retroviruses (ERV) - genome junctions from both sides. “A” denotes the

individuals that have the TSD, and “B” denotes the individuals that did not have the TSD. The right-lower panel shows a conceptual diagram of local de novo assembly using CAP3. Sequences in red indicate sequences not present on the reference genome, and sequences in purple indicate TSDs. The panels labeled "Homozygosity" and "Heterozygosity" illustrate the differences in read mapping patterns when an ERV insertion is homozygous versus heterozygous. Homozygosity: All reads break at the TSD. Heterozygosity: Some reads break at the TSD, but there are also reads that span across the TSD.