*Generation of Pleolipovirus database*

First, we generated a trusted database for each of the proteins conserved in all the available sequences of 16 previously isolated pleolipoviruses (ORFS 4,6, 7 and 8 in HRPV-1) (14, 17). Then, we used the sequences corresponding to each gene to query the IMG/VR database version 5.1 (27, 28) for related viruses using blastp implemented in Diamond (29). Hits with an e-value < 10-5 and a score > 50 were considered significant and added to the database of each core gene. This process was recursively iterated adding the new significant hits from each iteration to the respective database until exhaustion. Then, Hidden Markov model (HMM) were generated for each gene, for which alignments were performed with MAFFT v7.407 (30) and the --localpair and --reorder flags. Subsequently, HMM profiles were generated with HMMER v. 3.2.1 (31). HMM models for each gene were then used to further identify distant homologs for each gene using the same recursive approach described above. In order to reduce false positive results and to enhance specificity, minimum scores were set by using the PFAM database v.35.0 (32) as a training dataset. A minimum e-value of 10-5 and a score threshold at least 10 units higher than the highest? score obtained for each model against the training database were considered as true positives.

*Identification of genomes in metagenomic data*

Resultant models were then used to screen binned and non-binned contigs, generated in this study, to obtained potential virus genomes belonging to the *Pleolipoviridae* family. Contigs of at least 4.000 bp and with significant hits against at least four of the previously generated models of conserved genes among pleolipoviruses were considered for further analyses. Quality control to identify and exclude non-viral regions was performed with CheckV (33), database version 1.4 (Aug 27, 2022). Due to the limited representation of pleolipoviruses in CheckV database, and in order to properly asses the quality of potential pleolipoviruses-like genomes, the HMM models from conserved proteins previously generated in this work were added to the database and accounted as viral proteins. CheckV-trimmed sequences were clustered at 95% nucleotide identity with the software NUCmer (NUCleotide MUMmer) version 3.1 (34), which corresponds approximately to the species level (35, 36).