

A70 Viromes of two migratory birds as support for identifying the influence of molecular biology and bioinformatics settings

S. Tirera,¹ A. Salmier,¹ D. Donato,¹ J. M. Frigerio,² S. Therond,³ V. Lacoste,¹ B. de Thoisy,¹ A. Franc,² and A. Lavergne¹

¹LIVH, Institut Pasteur de la Guyane, Cayenne, French Guiana, ²INRA, UMR BioGeCo, Cestas, France and ³CNRS - IDRIS, Orsay, France

Thanks to high-throughput sequencing (HTS) and dedicated bioinformatics tools, it is possible to study all viruses present in a sample with an unprecedented detection of viral diversity. Applied to wildlife and humans, HTS becomes a powerful means to understand viral emerging infectious diseases. BLAST searches to find remote homolog sequences after an assembly of shotgun-generated sequences are often used to analyze virome datasets. However, this process might be biased by molecular biology approaches (use of whole genome/transcriptome amplification) and bioinformatics assembly that can induce chimeric sequences formation. Our objective was to identify confounding factors (molecular biology and bioinformatics) that can bias virome composition. Hence, we studied the viromes from cloacal swab samples of two migratory birds. Viral DNA and RNA were extracted using three different methods (easymag[®], phenol, and

trizol), followed by whole genome amplification (WGA)/WTA amplifications, before Illumina MiSeq sequencing. After trimming and assembly, sequences underwent BLAST (BLASTn and BLASTx). At the end of the first BLAST (n/x), some portions of sequences were uncovered and unassigned. We, then, implemented a novel recursive split-resubmit python program that searched for homologs of > 50 base pairs (bp), uncovering parts for a better exploitation of the datasets. Thirty-eight known viral families were detected in our samples. Circoviridae, Parvoviridae, and Microviridae were found in the majority of the results from easymag[®] and Phenol extracted samples, while trizol samples resulted in a majority of Picornaviridae and Coronaviridae. These results show that the viromes are of different composition depending on extraction methods. Most of the viral sequences were predominantly identified at the blastx step reflecting a high level of divergence with known viral sequences. Compared to a unique BLAST search, our resubmission pipeline allowed the assignment of up to 23 per cent of the viral sequences. Most of these sequences belonged to taxa already observed at the first blast step and few viral sequences of different families were detected. Knowing the complementarities between extraction methods and a better identification of viral sequences using our pipeline, the next step will be to get a precise position of each viral genomic segment in phylogenetic distance-based tree that will give a more accurate representation of viral diversity.