1   **Description of Additional Supplementary Files**

2   File Name: Supplementary Data 1

3   Description: Data showing virus-host associations, citations and GenBank accessions numbers.

4   File Name: Supplementary Data 2

5   Description: The relative importance of variables for different models.

6   File Name: Supplementary Data 3

7   Description: Species predictions generated by all stratified generalized boosted regression models for
8   virus groups and the labels used for training data.