

(a) Data

Total Viral Samples: 58314

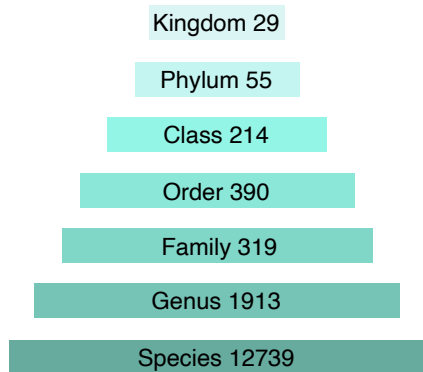
Data Source



Host Type

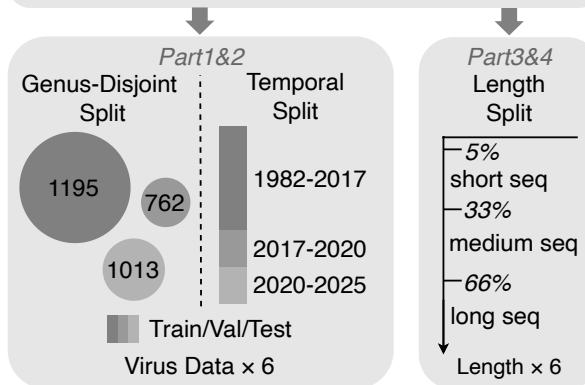


Virus Type



(b) Data Splitting

DNA/RNA/ALL Virus



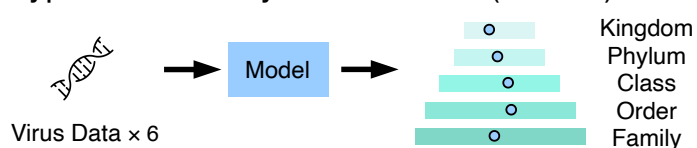
(c) 67 Baselines

- AIDO.DNA
- AIDO.RNA
- BiRNA-BERT
- Caduceus
- CNN
- DNABERT
- Genos
- Grover
- GENA-LM
- GENERator
- GenomeOcean
- HyenaDNA
- LucaOne
- LucaVirus
- MP-RNA
- Nucleotide Transformer
- OmniReg-GPT
- RNA-FM
- RiNALMo
- RNABERT

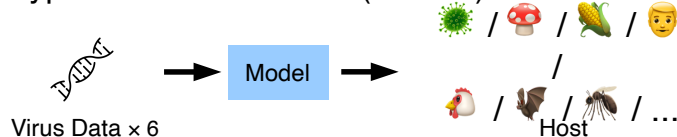
(d) 18 Benchmark Tasks

Understanding Axis: Classification Tasks

Type 1. Taxonomy Classification (6 tasks)

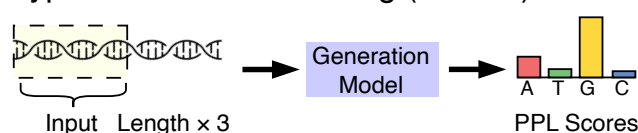


Type 2. Host Prediction (6 tasks)

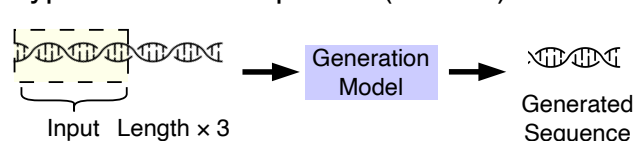


Biosecurity Axis: Generation Tasks

Type 3. Genome Modeling (3 tasks)



Type 4. CDS Completion (3 tasks)



(e) 11 Performance Metrics

Classification Tasks



Generation Tasks

