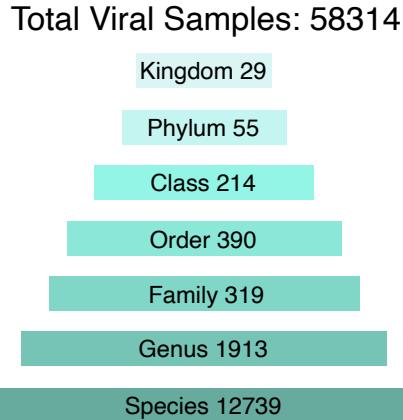


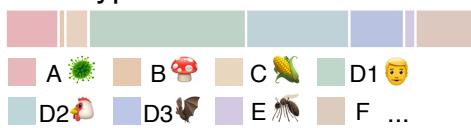
(a) Data



Data Source



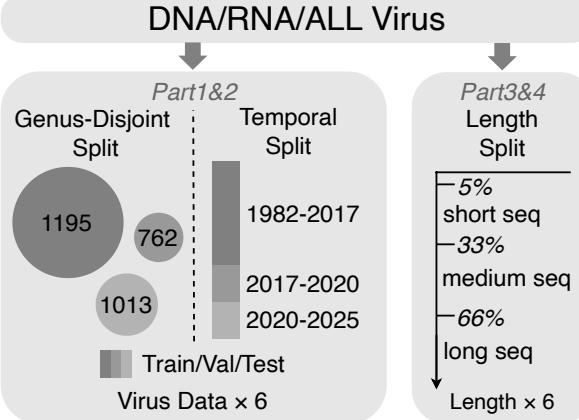
Host Type



Virus Type



(b) Data Splitting

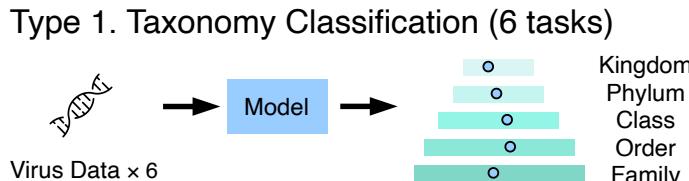


(c) 67 Baselines

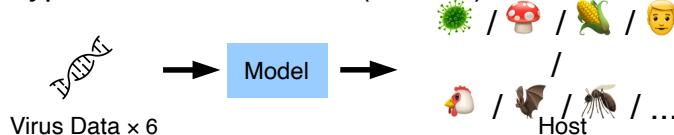
- AIDO.DNA
- HyenaDNA
- AIDO.RNA
- LucaOne
- BiRNA-BERT
- LucaVirus
- Caduceus
- MP-RNA
- CNN
- Nucleotide
- DNABERT
- Transformer
- Genos
- OmniReg-GPT
- Grover
- RNA-FM
- GENA-LM
- RiNALMo
- GENERator
- RNABERT
- GenomeOcean

(d) 18 Benchmark Tasks

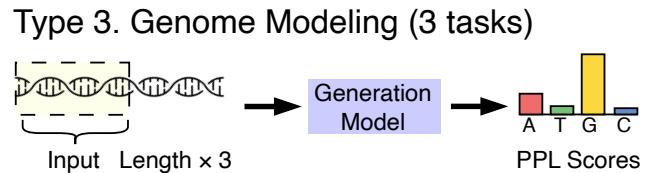
Understanding Axis: Classification Tasks



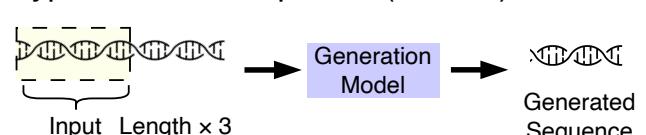
Type 2. Host Prediction (6 tasks)



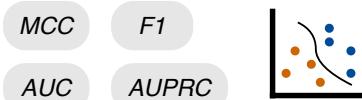
Biosecurity Axis: Generation Tasks



Type 4. CDS Completion (3 tasks)



Classification Tasks



Generation Tasks

- Bits-per-Base(BPB)
- Exact Match
- Alignment Identity
- Edit Distance
- k-mer JSD
- k-mer EMD
- k-mer KS