Download PPRV sequences and extract the following fields:
Accession no.; Country; Location; collection date; host;
isolate; strain; gene; length; sequence

Remove primers, patents,
unidentifiable sequences and short
sequences (<200bases)

Full-length
genome? —— Yes → Remove M-F intergenic
region

No

N, M, F or
H gene? —— No → Discard P and L
sequences

Identify and correct simple
sequence errors

Yes

Filter for contamination
errors using RDP5

Remove sequences requiring internal
gaps for alignment or containing
internal stop codons

Remove vaccine sequences and
other heavily passaged viruses

Remove sequences showing
excessive divergence as shown by
TreeTime

'Unique'
sequence sets ← Remove duplicate sequences (same
sequence, location and year)

'NoDups'
sequence sets ← Remove remaining duplicate
sequences (identical sequences)

Apply *Treemmer* filtering → Reduced size
sequence sets

S1 Fig Sequence screening and filtering workflow
The various stages of filtering applied to the collection of PPRV sequences in the database is shown
in this figure. The final output datasets are shown as Saved Data shapes (blue).