

**Table S4:** Accession numbers of sequences making up the negative training set for sub-type D when the *gag-pol* region is used for classification.

<b>Subtype</b>	<b>Accession Number</b>
A1	AB253422
A2	AF286238
B	AB565496
B	AF042102
B	AB480696
B	AF538305
B	AB485642
B	AF049495
C	AB254141
C	AB485645
F1	AB485656
F2	AJ249237
G	AB485662
G	AY586548
H	AF005496
H	FJ711703
J	AF082394
J	GU237072
K	AJ249235
K	AJ249239