

**Table S3:** Accession numbers of sequences making up the negative training set for all subtypes except D when the *gag-pol* region is used for classification.

<b>Subtype</b>	<b>Accession Number</b>
A1	AB253422
A2	AF286238
B	A04321
B	AB287372
C	AB254141
C	AB485645
D	A34828
D	AY773340
F1	AB485656
F2	AJ249237
G	AB485662
G	AY586548
H	AF005496
H	FJ711703
J	AF082394
J	GU237072
K	AJ249235
K	AJ249239