

**Table S1:** Accession numbers of sequences making up the positive training set for all subtypes except H & J when the *gag-pol* region is used for classification.

<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>F</b>	<b>G</b>
AM000053	AB097870	AF110963	AY773340	GQ290462	AB287003
AM000054	AB286956	AF286228	DQ054367	DQ979025	FJ389364
AB098330	AB287370	AF110974	AB485650	AB485659	AB485663
AB253421	AB289589	AB254150	AJ519489	AJ249238	AY586548
AF286237	AB428560	AB097871	A14116	AB480300	AF423760
AF286238	AB480698	AB485645	U88822	DQ189088	AY612637