

**Table S7:** Accession numbers of sequences making up the positive training set for all subtypes, except H and J, when the *env* region is used for classification.

<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>F</b>	<b>G</b>
AM000053	AB428560	AF286228	AY773340	GQ290462	AB287003
AM000054	AB287370	AB254150	AJ519489	AF075703	AF423760
AF286237	AB286956	AF110963	AB485650	DQ189088	AY612637
AF286238	AB480698	AB097871	DQ054367	AB485659	AY586548
AB098330	AB097870	AB485645	A14116	AJ249238	FJ389364
AB253421	AB289589	AF110974	U88822	DQ979025	AB485663