

Table S13: Number of sequences making up the positive training set used for determining whether a given sub-type (X) is present in the *env* region of a CRF.

Subtype(X)	Number of Sequences
A	12
B	30
C	10
D	20
F	12
G	12
H	4
J	4