

Figure S1. GenBank accession numbers and sequence alignment of U5-PBS region fragment nt 109 to 261 (MAL coordinates) of CRF01_AE (shadowed by yellow), subtype G (shadowed by blue) and CRF02_AG (shadowed by pink) studied in the present work*. Base changes are given as compared to the HXB2 isolate (K03455): transitions, transversions and insertions** are marked by green, red and blue, respectively, deletions – by hyphen. Non-identified nucleotides are in lilac. The functional motifs discussed in the text are shadowed: 5' and 3' strands of D-duplex by pink, I-duplex by green, U-duplex by blue and GAAA apical loops by turquoise. A portion of 5' strand of U5/AUG duplex and 5'-terminal CU of DIS hairpin as they are presented in the BMH structure model of the complete HIV-1 5'-UTR (Abbink & Berkhout, 2003) are shadowed by yellow. The PBS motif **tggcgcccgaaacagggac***** is omitted. Accession numbers of the sequences annotated as subtype reference sequences in the Los Alamos HIV sequence database are in bold.

K03455	cgctctg	ttgtg	tgactctggttaactagagatccctcag	accctttt	agtcagtgtg	gaaaatctctagcag	ctgaaagcgaaaggg	aaacagagaggagctctctc	gacgca	ggact	
AF070521	cgctctg	ttgtg	tgactctggttaactagagatccctcag	accctttt	agtcagtgtg	gaaaatctctagcag	ttgaaagcgaaagta	aagccagagaggagattctctc	gacgca	ggact	
X04415	catctg	ttgtg	tgactctggttaactagagatccctcag	accactct	agacggtgta	-aaaatctctagcag	tttaagagtgaagtt-aaacagggactcg	aaagcggaag	tt---ccagagaagttctctc	gacgca	ggact
AB032740	cgctctg	tggtta	agactctggttaactagagatccctcag	accactct	agactgagta	-aaaatctctagcag	ttgaaagcgaaagttaaatagggactcg	aaagcggaag	tt---ccagagaagttctctc	gacgca	ggact
AB032741	cgctctg	tggtta	ggactctggttaactagagatccctcag	accactct	agactgagta	-aaaatctctagcag	ttgaaagtgaagttaaatagggactcg	aaagcggaag	tt---ccagagaagttctctc	gacgca	ggact
AB052995	cgctctg	tggtta	ggactctggttaactagagatccctcag	accactct	agactgaata	-aaaatctctagcag	ttgaaagcgaaagttaaatagggactcg	aaagcggaag	tt---ccagagaagttctctc	gacgca	ggact
AB070352	cgctctg	tggtta	ggactctggttaactagagatccctcag	accactct	agactgagta	-aaaatctctagcag	ttgaaagcgaaagttaaatagggactcg	aaagcggaag	tt---ccagagaagttctctc	gacgca	ggact
AB220944	cgctctg	tggtta	ggactctggttaactagagatccctcag	accactct	agactgagta	-aaaatctctagcag	ttgaaagtgaagttaaatagggactcg	aaagcggaag	tt---ccagagaagttctctc	gacgca	ggact
AB220945	cgctctg	tggtta	ggactctggttaactagagatccctcag	accactct	agattgagtg	-aaaatctctagcag	ttgaaagcgaaagttaaatagggactcg	aaagcggaag	tt---ccagagaagttctctc	gacgca	ggact
AB220946	cgctctg	tggtta	ggactctggttaactagagatccctcag	accactct	agactgagta	-aaaatctctagcag	ttgaaagcgaaagttaaatagggactcg	aaagcggaag	tt---ccagagaagttctctc	gacgca	ggact
AB220947	cgctctg	tggtta	ggactctggttaactagagatccctcag	accactct	agactgagta	-aaaatctctagcag	ttgaaagcgaaagttaaatagggactcg	aaagcggaag	tt---ccagagaagttctctc	gacgca	ggact
AB220948	cgctctg	tggtta	ggactctggttaactagagatccctcag	accactct	agactgagt-	-aaaatctctagcag	ttgaaagcgaaagttaaatagggactcg	aaagcggaag	tt---ccagagaagttctctc	gacgca	ggact
AB253423	cgctctg	tggtta	ggactctggttaactagagatccctcag	accactct	agactgagta	-aaaatctctagcag	ttgaaagtgaagttaaatagggactcg	aaagcggaag	tt---ccagagaagttctctc	gacgca	ggact
AB253424	cgctctg	tggtta	ggactctggttaactagagatccctcag	accactct	agattgagt-	-aaaatctctagcag	ctgaaagcgaaagttaaatagggactcg	aaagcggaag	tt---ccagagaagttctctc	gacgca	ggact
AB253426	catctg	tggtta	ggactctggttaactagagatccctcag	accactct	agactgagta	-aaaatctctagcag	ttgaaagtgaagttaaatagggactcg	aaagcggaag	tt---ccagagaagttctctc	gacgca	ggact
AB253635	catctg	tggtta	ggactctggttaactagagatccctcag	accactct	agactgagta	-aaaatctctagcag	ttgaaagcgaaagttaaatagggactcg	aaagcggaag	tt---ccagagaagttctctc	gacgca	ggact
AB253647	catctg	tggtta	ggactctggttaactagagatccctcag	accactct	agactgagta	-aaaatctctagcag	ctgaaagcgaaagttaaatagggactcg	aaagcggaag	tt---ccagagaagttctctc	gacgca	ggact
AB253659	catctg	tggtta	ggactctggttaactagagatccctcag	accactct	agactgagta	-aaaatctctagcag	ttgaaagcgaaagttaaatagggactcg	aaagcggaag	tt---ccagagaagttctctc	gacgca	ggact
AB253669	catctg	tggtta	ggactctggttaactagagatccctcag	accactct	agactgagta	-aaaatctctagcag	ttgaaagcgaaagttaaatagggactcg	aaagcggaag	tt---ccagagaagttctctc	gacgca	ggact
AB253692	cgctctg	tggtta	ggactctggttaactagagatccctcag	accactct	agactgagta	-aaaatctctagcag	ttgaaagtgaagttaaatagggactcg	aaagcggaag	tt---ccagagaagttctctc	gacgca	ggact
AB253703	cgctctg	tggtta	ggactctggttaactagagatccctcag	accactct	agactgagta	-aaaatctctagcag	ttgaaagtgaagttaaatagggactcg	aaagcggaag	tt---ccagagaagttctctc	gacgca	ggact
AB253717	catctg	tggtta	ggactctggttaactagagatccctcag	accactct	agactgagta	-aaaatctctagcag	tcgaaagtgaagttaaatagggactcg	aaagcggaag	tt---ccagagaagttctctc	gacgca	ggact
AB485652	cgctctg	tggtta	ggactctggttaactagagatccctcag	accactct	agactgagt-	-aaaatctctagcag	ttgaaagcgaaagttaaatagggactcg	aaagcggaag	tt---ccagagaagttctctc	gacgca	ggact
AB485653	cgctctg	tggtta	ggactctggttaactagagatccctcag	accactct	agactgagt-	-aaaatctctagcag	ttgaaagcgaaagttaaatagggactcg	aaagcggaag	tt---ccagagaagttctctc	gacgca	ggact
AB485654	cgctctg	tggttt	ggactctggttaactagagatccctcag	accactct	agactgagta	-aaaatctctagcag	ttgaaagtgaagttaaatagggactcg	aaagcggaag	tt---ccagagaagttctctc	gacgca	ggact
AB485655	catctg	tggtta	ggactctggttaactagagatccctcag	accactct	agactgagta	-aaaatctctagcag	ttgaaagtgaagttaaatagggactcg	aaagcggaag	tt---ccagagaagttctctc	gacgca	ggact
AB565503	cgctctg	tggtta	ggactctggttaactagagatccctcag	accactct	agactgagta	-aaaatctctagcag	ttgaaagcgaaagttaaatagggactcg	aaagcggaag	tt---ccagagaagttctctc	gacgca	ggact
AF164485	cgctctg	tggtta	ggactctggttaacta ^{aa} atccctcag	accactct	agactgagta	-aaaatctctagcag	ttgaaagcgaaagtctaatagggactcg	aaagcggaag	tt---ccagagaagttctctc	gacgca	ggact
AF197339	cgctctg	tggtta	ggactctggttaactagagatccctcag	accactct	agactgagt-	-aaaatctctagcag	tagaaagcgaaagttaaatagggactcg	aaagcggaag	tt---ccagagaagttctctc	gacgca	ggact
AF197340	cgctctg	tgata	ggactctggttaactagagatccctcag	accactct	agactgagta	-aaaatctctagcag	ttgaaagcgaaagttaaatagggactcg	aaagcggaag	tt---ccagagaagttctctc	gacgca	ggact
AF197341	cgctctg	tgata	ggactctggttaactagagatccctcag	accactca	agactgagta	-aaaatctctagcag	ttgaaagcgaaagttaaacagggactcg	aaagcggaag	tt---ccagagaagttctctc	gacgca	ggact
AF259954	cgctctg	tggtg	ggactctggttaactagagatccctcag	accactct	agactgagt-	-aaaatctctagcag	ttgaaagcgaaagttaaatagggactcg	aaagcggaag	tt---ccagagaagttctctc	gacgca	ggact
DQ314731	cgctctg	tggttt	ggactctggttaactagagatccctcag	accactct	agactgagta	-aaaatctctagcag	ttgaaagcgaaagttaaatagggactcg	aaagcggaag	tt---ccagagaagttctctc	gacgca	ggact
DQ314732	cgctctg	tggttt	ggactctggttaactagagatccctcag	accactct	agactgagta	-aaaatctctagcag	ttgaaagcgaaagttaaatagggactcg	aaagcggaag	tt---ccagagaagttctctc	gacgca	ggact
DQ859178	catctg	tggtta	ggactctggttaactagagatccctcag	accactct	agactgagta	-aaaatctctagcag	ttgaaagtgaagttaaatagggactcg	aaagcggaag	tt---ccagagaagattctctc	gacgca	ggact
DQ859179	cgctctg	tggtta	ggactctggttaactagagatccctcag	accactct	agattgagta	-aaaatctctagcag	ttgaaagcgaaagttaaatagggactcg	aaagcggaag	tt---ccagagaagttctctc	gacgca	ggact
DQ859180	cgctctg	tggtta	ggactctggttaactagagatccctcag	ac-actgc	agttgagta	-aaaatctctagcag	ttgaaagcgaaagttaaatagggactcg	aaagcggaag	tt---ccagagaagttctctc	gacgca	ggact
EF036527	catctg	tggtta	ggactctggttaactagagatccctcag	accactct	agactgagta	-aaaatctctagcag	ttgaaagcgaaagttaaatagggactcg	aaagcggaag	tt---ccagagaagttctctc	gacgca	ggact
EF036528	cgctctg	tggttt	ggactctggttaactagagatccctcag	accactct	agactaagta	-aaaatctctagcag	ttgaaagcgaaagttaaatagggactcg	aaagcggaag	tt---ccagagaagttctctc	gacgca	ggact
EF036529	catctg	tggtta	ggactctggttaactagagatccctcag	accactcc	agactgagta	-aaaatctctagcag	ttgaaagcgaaagttaaatagggactcg	aaagcggaag	tt---ccagagaagttctctc	gacgca	ggact
EF036530	catctg	tggtta	ggactctggttaactagagatccctcag	accactct	agactgagta	-aaaatctctagcag	ttgaaagcgaaagttaaatagggactcg	aaagcggaag	tt---ccagagaagttctctc	gacgca	ggact
EF036531	catctg	tggtta	ggactctggttaactagagatccctcag	accactct	agactgagta	-aaaatctctagcag	ttgaaagcgaaagttaaatagggactcg	aaagcggaag	tt---ccagagaagttctctc	gacgca	agact
EF036532	cgctctg	tggtta	ggactctggttaactagagatccctcag	accactct	agactaagta	-aaaatctctagcag	tcgaaagcgaaagttaaatagggactcg	aaagcggaag	tt---ccagagaagttctctc	gacgca	ggact
EF036533	cgctctg	tggtta	ggactctggttaactagagatccctcag	accactct	agactaagta	-aaaatctctagcag	ttgaaagcgaaagttaaatagggactcg	aaagcggaag	tt---ccagagaagttctctc	gacgct	ggact
EF036534	cgctctg	tggtta	ggactctggttaactagagatccctcag	accactct	agactgagta	-aaaatctctagcag	cttgaagcggaagttaaatagggactcg	aaagcggaag	tt---ccagagaagttctctc	gacgca	ggact
EF036535	catctg	tggtta	ggactctggttaactagagatccctcag	accactct	agactgagta	-aaaatctctagcag	ttga--gtgacagttaaatagggactcg	aaagcggaag	tt---ccagagaagttctctc	gacgca	ggact
EF036536	catctg	tggttaa	ggactctggttaactagagatccctcag	accactct	agactgagta	-aaaatctctagcag	ttgaaagcgaaagttaaatagggactcg	aaagcggaag	tt---ccagagaagttctctc	gacgca	ggact
JX446666	cgctctg	tggtta	ggactctggttaactagagatccctcag	accactct	agactgagta	-aaaatctctagcag	ttgaaagcgaaagt-taaagggactcg	aaagcggaag	tt---ccagagaagttctctc	gacgca	ggact
JX447127	cgctctg	tggtta	ggactctggttaactagagatccctcag	accactct	agactgagt-	-aaaatctctagcag	ttgaaagtgaagttaaatagggactcg	aaagcggaag	tt---ccagagaagttctctc	gacgca	ggact
JX447349	-gtctg	tggtta	ggactctggttaactagagatccctcag	accactct	agactgagta	-aaaatctctagcag	ttgaaagcgaaagttaaatagggactcg	aaagcggaag	tt---ccagagaagttctctc	gacgca	ggact

JX447539	cgtctg	tg	gga	ccctcag	accactct	agactgagt-	-aaaatctctagcag	ttgaaagcgaag	tt	aaagcgaag	tt	ccagagaag	ttctctc	gacgca	ggact	
JX447638	-gtctg	tg	gga	ccctcag	accactct	agactgagta	-aaaatctctagcag	ttgaaagcgaag	tt	aaagcgaag	tt	ccagagaag	ttctctc	gacgca	ggact	
U51188	cgtctg	tg	gga	ccctcag	accactct	agactgagta	-aaaatctctagcag	ttgaaagcgaag	tt	aaagcgaag	tt	ccagagaag	ttctctc	gacgca	ggact	
U51189	catctg	tg	gga	ccctcag	atcactct	agactgagta	-aaaatctctagcag	ttgaaagcgaag	tt	aaagcgaag	tt	ccagagaag	ttctctc	gacgca	ggact	
U54771	cgtctg	tg	gga	ccctcag	accactct	agactgagta	-aaaatctctacag	tcgaaagcgaag	tt	aaagcgaag	tt	ccagagaag	ttctctc	gacgca	ggact	
AB231893	cgtctg	ttgtg	tg	ccctcag	accactct	agacagtgtgta	-aaaatctctagcag	ccgaaagtgaag	tt	aaagcgaag	tt	ccagagaag	ttctctc	gacgca	ggact	
AB485662	cgtctg	ttgtg	tg	ccctcag	accactct	agatagtgta	-aaaatctctagcag	ttgaaagtga	g	aaagcgaag	tt	ccagagaag	atctctc	gacgca	ggact	
AF084936	cgtctg	ttgtg	tg	ccctcag	accactct	agatagtgta	-aaaatctctagcag	ttgaaaacgaag	tt	aaagcgaag	tt	ccagagaag	ttctctc	gacgca	ggact	
AY586547	cgtctg	ttgtg	tg	ccctcag	accactct	agacgggtgta	-aaaatctctagcag	ttgaaagcgaag	tt	aaagcgaag	tt	ccagagaag	ttctctc	gacgca	ggact	
AY586549	cgtctg	ttgtg	tg	ccctcag	accactct	agacagtgtgta	-aaaatctctagcag	ttgaaagcgaag	tt	aaagcgaag	tt	ccagagaag	ttctctc	gacgca	ggact	
AY612637	cgtctg	ttgtg	tg	ccctcag	accactct	agatagtgta	-aaaatctctagaa	ttgaaagcgaag	tt	aaagcgaag	tt	ccagagaag	ttctctc	gacgca	ggact	
AY772535	cgtctg	ttgtg	tg	ccctcag	accactct	agacagtgtgta	-aaaatctctagcag	ttgaaagcgaag	tt	aaagcgaag	tt	ccagagaag	ttctctc	gacgca	ggact	
EU786670	cgtctg	ttgtg	tg	ccctcag	accactct	agacagata	-aaaatctctagcag	ttgaaaacgaag	tt	aaagcgaag	tt	ccagagaag	ttctctc	gacgca	ggact	
										aa	aaagcgaag	tt	ccagagaag	ttctctc	gacgca	ggact
FJ389363	cgtctg	ttgtg	cg	ccctcag	accactct	agatagtgta	-aaaatctctagcag	ttgaaaacgaag	tt	aaagcgaag	tt	ccagagaag	ttctctc	gacgca	ggact	
FJ389364	cgtctg	tt	tg	ccctcag	accactct	a	acagtgtgta	-aaaatctctagcag	ttgaaaacgaag	tt	aaagcgaag	tt	ccagagaag	atctctc	gacgca	ggact
FJ389366	cgtctg	ttgtg	tg	ccctcag	accactc	agacagtgtgta	-aaaatctctagcag	ttgaaaacgaag	tt	aaagcgaag	tt	ccagagaag	ttctctc	gacgca	ggact	
FJ389367	cgtctg	ttgtg	tg	ccctcag	accactct	agacagtgtgta	-aaaatctctagcag	ttgaaaacgaag	tt	aaagcgaag	tt	ccagagaag	ttctctc	gacgca	ggact	
FJ670520	cgtctg	ttgtg	tg	ccctcag	accactaa	aagaggtgtgta	-aaaatctctagcag	ttgaaagtgaag	tt	aaagcgaag	tt	ccagagaag	ttctctc	gacgca	ggact	
FJ670530	cgtctg	ttgtg	tg	ccctcag	accactct	aggtagtgta	-aaaatctctagcag	ttgaaagagaa	tt	aaagcgaag	tt	ccagagaag	ttctctc	gacgca	ggact	
FR846408	cgtctg	ttgtg	tg	ccctcag	accactct	agacagtgtgta	-aaaatctctagcag	ctgaaagcgaag	tt	aaagcgaag	tt	ccagagaag	ttctctc	gacgca	ggact	
FR846409	cgtctg	ttgtg	tg	ccctcag	accactct	aaacagtgtgta	-aaaatctctagcag	ttggaagcgaag	tt	aaagcgaag	tt	ccagagaag	ttctctc	gacgca	ggact	
GU362882	cgtctg	ttgtg	tg	ccctcag	accactct	agttagtgta	-aaaatctctagcag	ttgaaagcgaag	tt	aaagcgaag	tt	ccagagaag	ttctctc	gacgca	ggact	
GU362884	cgtctg	ttgt	tg	ccctcag	accactct	agatagtgta	-aaaatctctagcag	ttgaaaacgaag	tt	aaagcgaag	tt	ccagagaag	ttctctc	gacgca	ggact	
AB049811	cgtctg	ttgtg	tg	ccctcag	actactct	agactgtgtgta	-aaaatctctagcag	ttgaa	g	aaagcgaag	tt	ccagagaag	ttctctc	gacgca	ggact	
AB231895	cgtctg	ttgtg	tg	ccctcag	accactct	agactgtgtgta	gaaaatctctagcag	ttgaa	g	aaagcgaag	tt	ccagagaag	ttctctc	cgacgca	g	
										tcg	aaagcgaag	tt	ccagagaag	ctctctc	gacgcag	ggact
AB231896	cgtctg	ttgtg	tg	ccctcag	actacttt	agtcagtgtgta	-aaaatctctagcag	cggaa	g	aaagcgaag	tt	ccagagaag	atctctc	gacgcac	ggact	
AB231898	cgtctg	ttgtg	tg	ccctcag	accattt	agtc	ttgtgta	-aaaatctctagcag	cggaa	g	aaagcgaag	tt	ccagagaag	-atctc	gacgcac	ggact
AB286855	cgtctg	ttgtg	tg	ccctcag	accactgt	agactgtgtgta	aaaaatctctagcag	ctgaa	g	aaagcgaag	tt	ccagagaag	ttctctc	gacgcag	ggact	
AB286857	cgtctg	ttgtg	tg	ccctcag	accacttt	actcagtgtgta	-aaaatctctagcag	cggaa	g	aaagcgaag	tt	ccagagaag	ttctctc	gacgcac	ggact	
AB286859	cgtctg	ttgtg	tg	ccctcag	accactct	agactgtgtgta	gaaaatctctagcag	ttgaa	g	aaagcgaag	tt	ccagagaag	ctctctc	gacgcag	ggact	
AB286861	cgtctg	ttgtg	tg	ccctcag	accacttt	cagttgtgtgta	-aaaatctctagcag	cggaa	g	aaagcgaag	tt	ccagagaag	ctctctc	gacgcag	ggact	
AB286863	cgtctg	ttgtg	tg	ccctcag	actactct	agacgggtg-	-aaaatctctagcag	ttgga	g	aaagcgaag	tg	ccagagaag	atctctc	gacgcag	ggact	
AB286864	cgtctg	ttgtg	cg	ccctcag	actactct	agacgggtgta	-aaaatctctagcag	ttgga	g	aaagcgaag	tg	ccagagaag	atctctc	gacgcag	ggact	
AB485633	cgtcta	ttgtg	tg	ccctcag	accactct	agactgtgtgta	gaaaatctctagcag	cggaa	g	aaagcgaag	tt	ccagagaag	ctctctc	gacgcag	ggact	
AB485634	cgtctg	ttgtg	tg	ccctcag	accactct	aggttgtgtgta	aaaaatctctagcag	ttgaa	g	aaagcgaag	tt	ccagagaag	atctctc	gacgcag	ggact	
AB485635	cgtctg	ttgtg	tg	ccctcag	accactct	aggttgtgtgta	aaaaatctctagcag	cggaa	g	aaagcgaag	tt	ccagagaag	atctctc	gacgcag	ggact	
AB485636	cgtctg	ttgtg	tg	ccctcag	accactct	agactgtata	-aaaatctctagcag	ttgaa	g	aaagcgaag	tt	ccagg	-agctctc	gacgcag	ggact	
AY271690	cgtctg	ttgtg	tg	ccctcag	accacttt	agactgtgtgta	-aaaatctctagcag	ttgaa	g	aaagcgaag	tt	ccagagaag	atctctc	gacgcac	ggact	
EU786671	cgtctg	ttgtg	tg	ccctcag	accactct	agacttagtg	aaaaatctctagcag	cggaa	g	aaagcgaag	tt	ccagagaag	ctctctc	gacgcag	ggact	
EU884501	cgtctg	ttgtg	tg	ccctcag	accactcg	agacgggtgta	gaaaatctctagcag	cggaa	g	aaagcgaag	tt	ccagagaag	atctctc	gacgcag	ggact	
JQ316136	cgtctg	ttgtg	tg	ccctcag	accatttt	agctt	ggta	g	-aaaatctctagcag	ttgaa	g	aaagcgaag	ctctctt	gacgcag	ggact	
JQ316137	cgtctg	ttgtg	tg	ccctcag	accattct	agctt	ggta	-aaaatctctagcag	ttgaa	g	aaagcgaag	tt	ccagagaag	ctctctt	gacgcag	ggact
JQ316138	cgtctg	ttgtg	tg	ccctcag	accattat	agctt	ggta	-aaaatctctagcag	ttgag	g	aaagcgaag	tt	ccagagaag	ctctctt	gacgcag	ggact
L39106	cgtctg	ttgtg	tg	ccctcag	accactct	agactgtgtgta	-aaaatctctagcag	ttgac	g	aaagcgaag	tt	ccagagaag	atctctc	gacgcag	ggact	
K03455	cgtctg	ttgtg	tg	ccctcag	accctttt	agtcagtgtg	gaaaatctctagcag	ctgaaagcgaaggg				aaac	ccagaggagctctctc	gacgca	ggact	

- * The 52 isolates of CRF01_AE used in the present work have been sampled from at least 43 patients residing in 5 countries (CAR, Thailand, Indonesia, Japan and China); the 18 isolates of subtype G - from 18 different patients residing in 7 countries (Kenya, Cameroon, Ghana, Belgium, Spain, Portugal and Cuba); the 21 isolates of CRF02_AG - from 17 different patients residing in 7 countries (Liberia, Cameroon, Nigeria, Ghana, Spain, France and South Korea).
- ** As compared to the MAL isolate and some isolates of A1/A2 subgroup, for example, AB253421, AB253428, AF107771, AF286237, AY322190, closely clustering with the MAL isolate in PBS region, all MAL-like isolates studied in the present work (except for the CRF01_AE isolate JX446666) contain an additional nucleotide U between U210 and the first nucleotide A211 of the duplicate insertion. As compared to HXB2, this nucleotide is marked by red (G214U transversion). Following the nucleotide numbering of the MAL isolate (Baudin et al., 1993; Goldschmidt et al., 2002; Goldschmidt et al., 2004; Isel et al., 1995; Isel et al., 1998; Isel et al., 1999; Isel et al., 2010; Sleiman et al., 2012), this additional nucleotide may be considered as an insertion.
- *** Among all MAL-like templates studied in the present work are four isolates with a single mutation in the PBS sequence: EF036531 (tggtgccccgaacagggac), EF036534 (-ggcgccccgaacagggac), U54771 (tggcgccccgaacagggac) and L39106 (tggcgccccgaacagg-ac). The subtype G isolate (EU786670) contains a doubled duplicate insertion and the CRF02_AG isolate (AB231895) contains a repeat of a larger portion of the duplicate insertion and a downstream sequence up to the 5'-end of DIS hairpin.