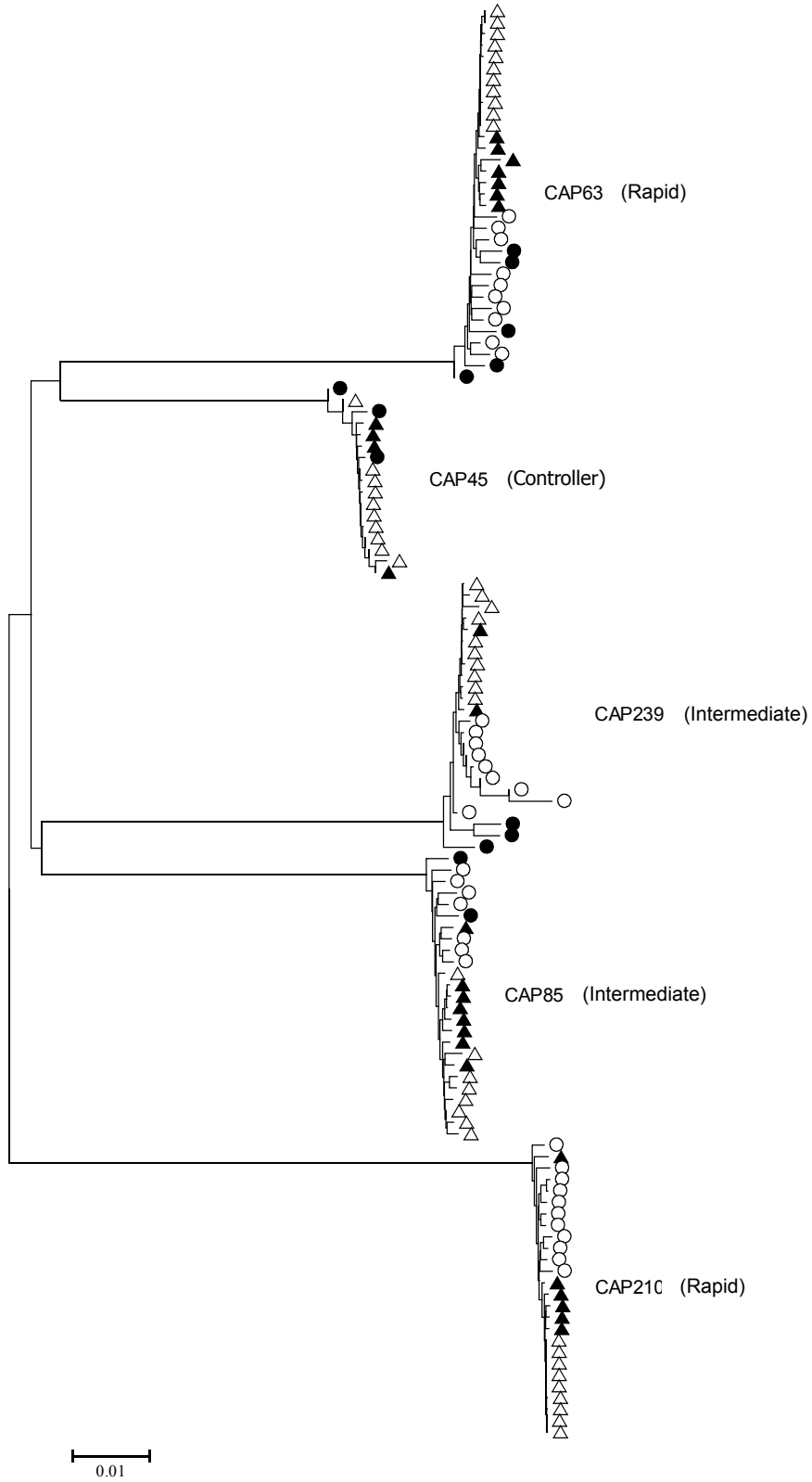

Figure S1. Neighbour-Joining tree of near full-length genome HIV-1 subtype C sequences from five women in the CAPRISA 002 Acute Infection cohort. Timepoints sequenced are indicated as two to five weeks (open triangles), three months (solid triangles), six months (open circles) and later than six months (solid circles). Disease progression classification for each participant is indicated. (Scale bar = 0.01)

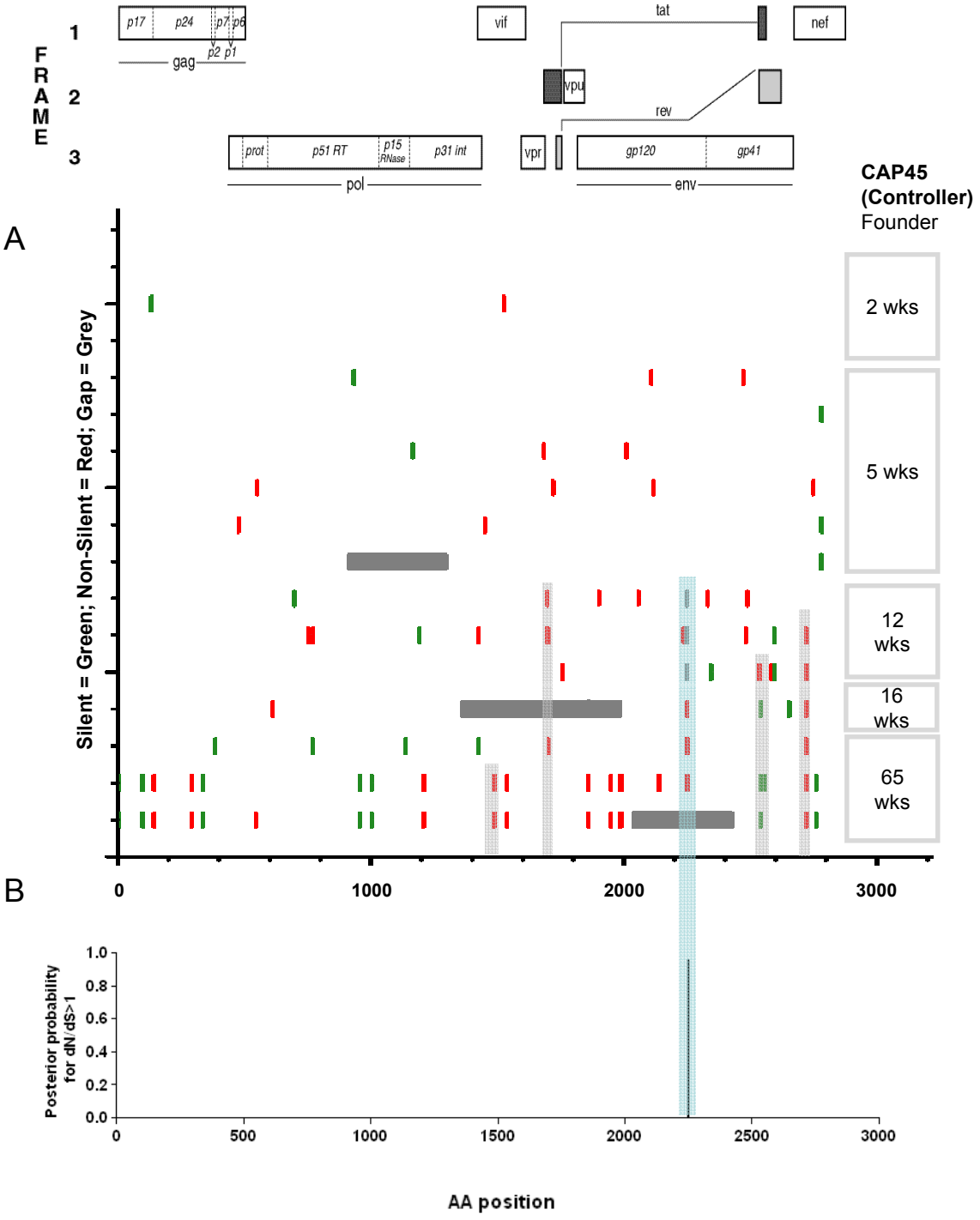
FRAME
1
2
3
p17
p24
p7
p6
p2
p1
gag
vif
tat
nef
vpu
rev
prot
p51 RT
p15 RNase
p31 int
pol
vpr
gp120
gp41
env
CAP45 (Controller) Founder
A
Silent = Green; Non-Silent = Red; Gap = Grey
2 wks
5 wks
12 wks
16 wks
65 wks
0
1000
2000
3000
B
Posterior probability for dN/dS>1
1.0
0.8
0.6
0.4
0.2
0.0
0
500
1000
1500
2000
2500
3000
AA position

Figure S2

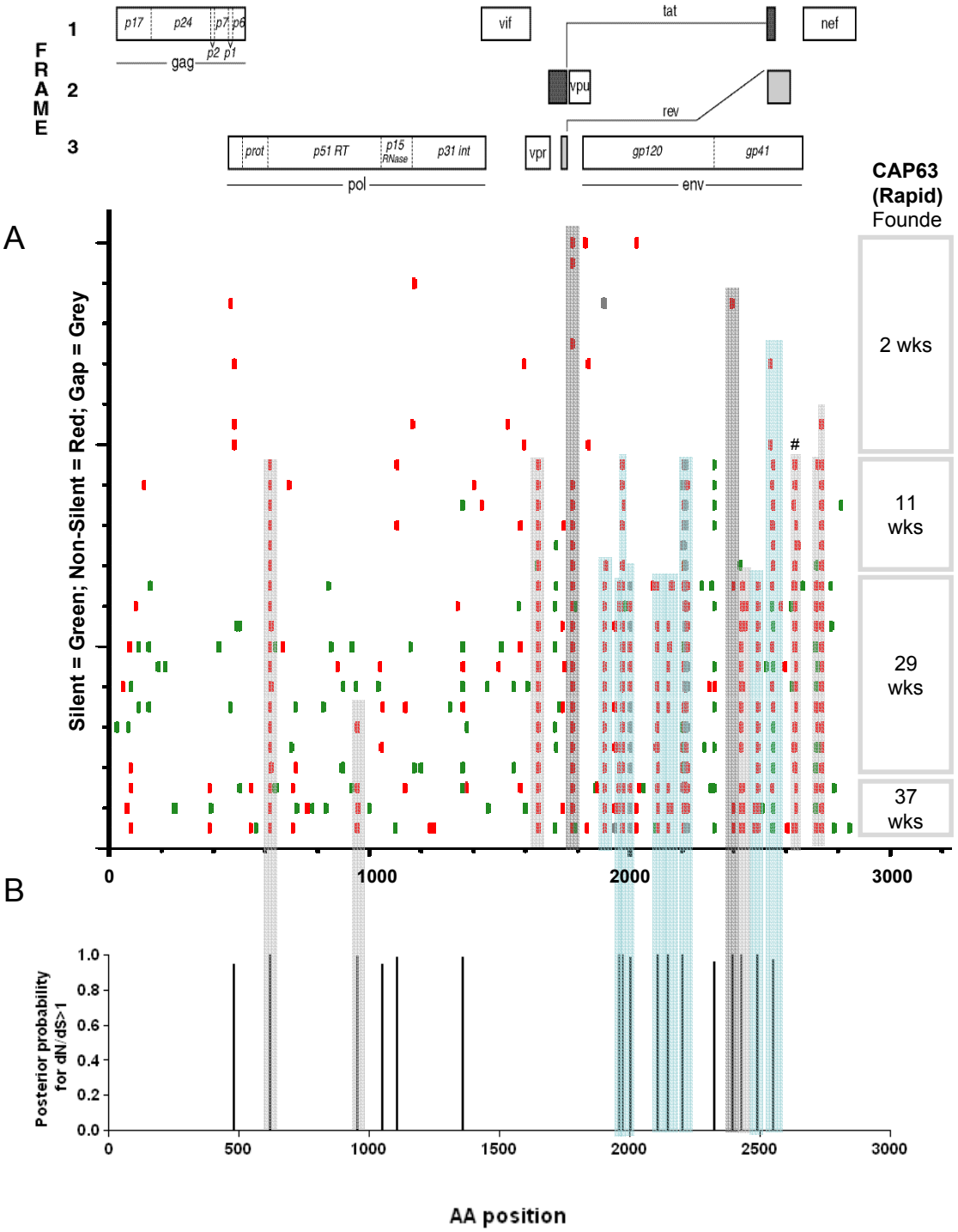
FRAME
1
2
3
p17
p24
p7
p6
p2
p1
gag
vif
tat
nef
vpu
rev
prot
p51 RT
p15
RNase
p31 int
pol
vpr
gp120
gp41
env
A
CAP63
(Rapid)
Founde
2 wks
11
wks
29
wks
37
wks
#
Silent = Green; Non-Silent = Red; Gap = Grey
0
1000
2000
3000
B
Posterior probability
for dN/dS>1
1.0
0.8
0.6
0.4
0.2
0.0
0
500
1000
1500
2000
2500
3000
AA position

Figure S3

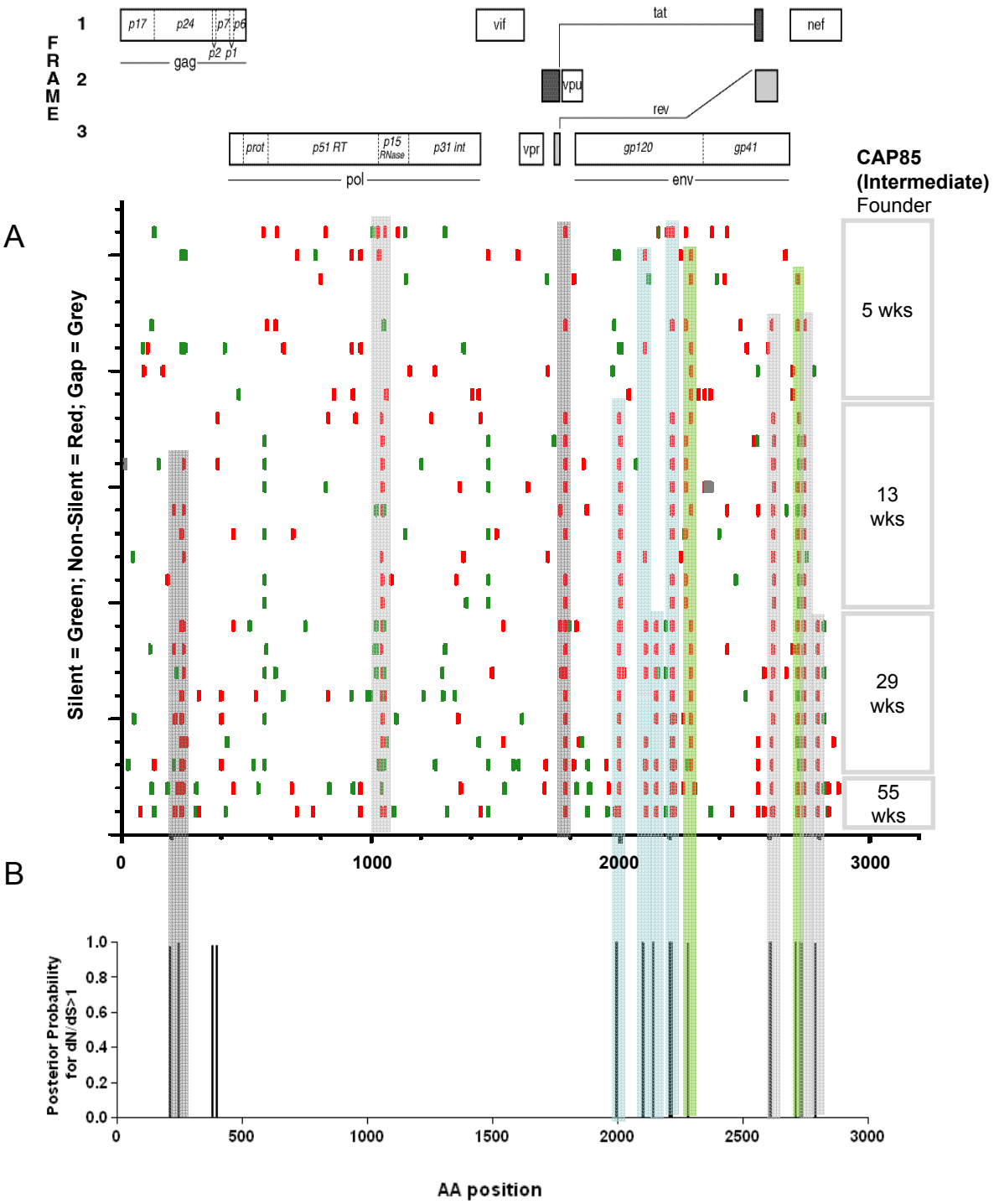
FRAME
1
2
3
p17
p24
p7
p6
p2
p1
gag
vif
tat
nef
vpu
rev
prot
p51 RT
p15
RNase
p31 int
vpr
gp120
gp41
pol
env
CAP85
(Intermediate)
Founder
A
Silent = Green; Non-Silent = Red; Gap = Grey
5 wks
13 wks
29 wks
55 wks
0
1000
2000
3000
B
Posterior Probability for dN/dS>1
1.0
0.8
0.6
0.4
0.2
0.0
0
500
1000
1500
2000
2500
3000
AA position

Figure S4

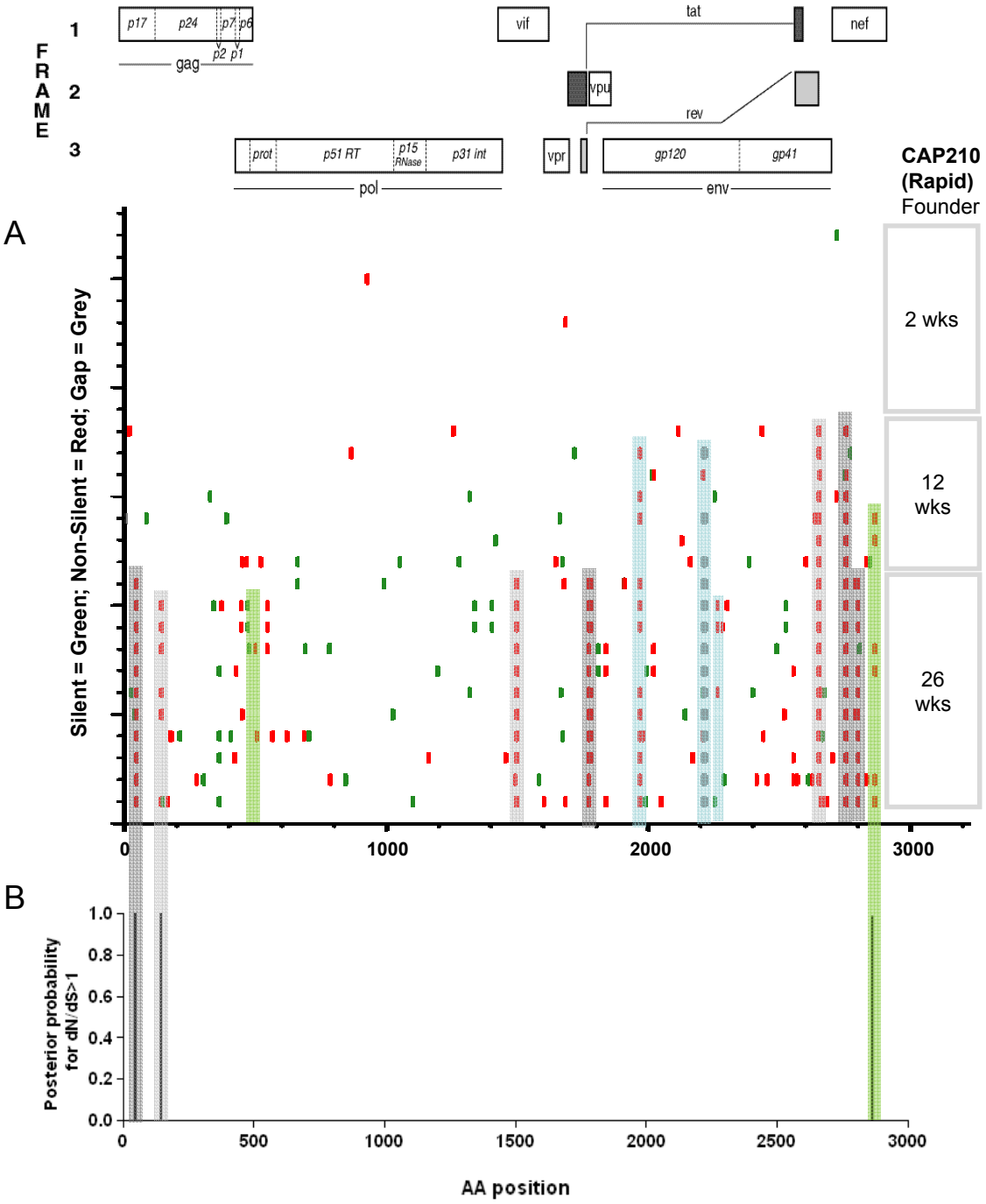
FRAME
1
2
3
p17
p24
p7
p6
p2
p1
gag
vif
tat
nef
vpu
rev
prot
p51 RT
p15
RNase
p31 int
pol
vpr
gp120
gp41
env
CAP210
(Rapid)
Founder
A
Silent = Green; Non-Silent = Red; Gap = Grey
2 wks
12
wks
26
wks
0
1000
2000
3000
B
Posterior probability
for dN/dS>1
1.0
0.8
0.6
0.4
0.2
0.0
0
500
1000
1500
2000
2500
3000
AA position

Figure S5

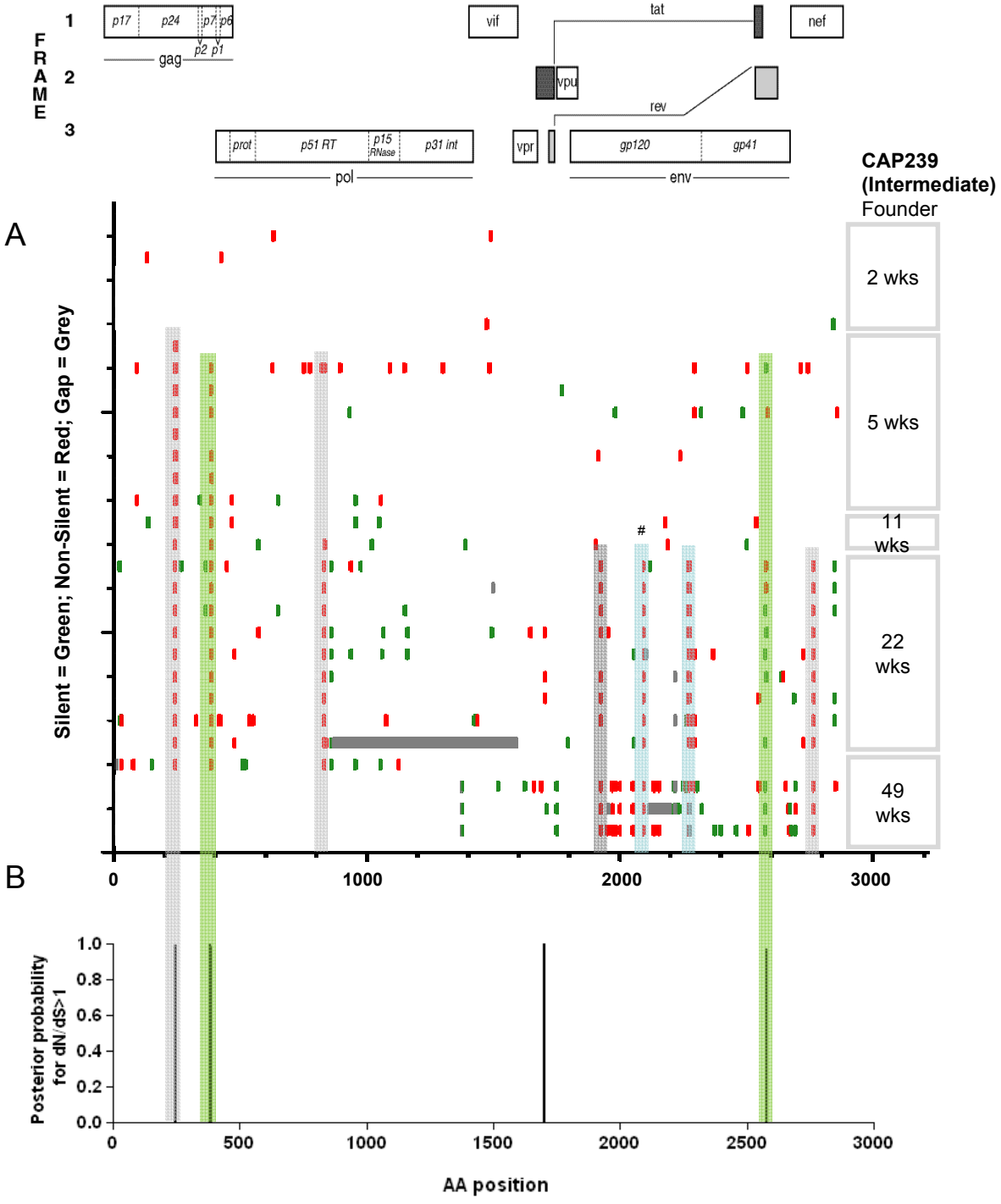
FRAME
1
2
3
p17
p24
p7
p6
p2
p1
gag
vif
tat
nef
vpu
rev
prot
p51 RT
p15 RNase
p31 int
vpr
gp120
gp41
pol
env
CAP239 (Intermediate) Founder
A
Silent = Green; Non-Silent = Red; Gap = Grey
2 wks
5 wks
11 wks
22 wks
49 wks
#
0
1000
2000
3000
B
Posterior probability for dN/dS>1
1.0
0.8
0.6
0.4
0.2
0.0
0
500
1000
1500
2000
2500
3000
AA position

Figure S6

Figure S2 to S6. Supplemental Digital Content 2. Immune selective pressure and positive selection in near full-length genomes from five subtype C-infected women. (**A**) *Highlighter* plots (http://www.hiv.lanl.gov/content/sequence/HIGHLIGHT/HIGHLIGHT_XYPLOT/highlighter.html) of synonymous and nonsynonymous changes from the transmitted/founder virus over time, and (**B**) amino acid sites across the genome under positive selection given as the posterior probability of a dN/dS ratio >1 for each site. Shaded bars indicate where mutating sites in the *Highlighter* plots correspond to sites under positive selection. Pale grey bars = putative cytotoxic T-lymphocyte escape sites, dark grey = sites reverting to consensus, blue bars = sites under putative antibody-mediated pressure and green bars = sites with changes which could not be classified. Time points followed by an * indicate that one or more sequence was excluded due to large deletions. A $^{\#}$ indicates that mutations were found to arise earlier than indicated on the *Highlighter* plot by subsequent focused epitope sequencing.

Table S1. Genome regions containing putative reversion mutations

Participant ID	ORF	Genome region sequence[1]	HXB2 position	Amino acid (% frequency[3]) change
CAP63	Vpu	VGALIIALILT**V**VVWIIA	9-26	V to I (0.9% to 97.85%)
	Gp41	LAVERYL**R**DQQLLGIWGC[2]	70-87	R to K (7.28% to 76.18%)
CAP85	Gag	TS**N**LQEQIAWMT**A**NPPVPVG**E**	240-260	N to T (10% to 85%) A to S (4.4% to 54%) E to D (31.72% to 68.04%)
	Vpu	IIAIIVWTI**T**YLEYRKVV	17-34	T to A (1.6% to 48%)
CAP210	Gag	LERFAL**D**PGLLETSGGCK	41-58	D to N (1.5% to 98%)
	Vpu	YRLGVGA**F**IVALIIAIVV	5-22	F to L (19% to 80%)
	Nef	EVGFPV**K**PQVPLRPMTYK	65-82	K to R (8.5% to 88%)
	Nef	LIYSKKRQDILDLW**I**YNT	100-117	I to V (1% to 98%)
CAP239	Gp120	NDMVDQMH**K**DIISLWDQS	98-115	K to E (0.19% to 93.2%)

[1]Bold amino acids indicate those undergoing reversion; underlined amino acids indicate those evolving under positive selection. Genome regions may also contain changes from high to low frequency amino acids not associated with known CTL escape, not shown here.

[2]Participant A*02:01/A*23:01 HLA association according to HIV molecular Immunology 2008 (http://www.hiv.lanl.gov)

[3]Amino acid frequencies according to Los Alamos subtype C database sequences for each gene (http://www.lanl.gov)

Table S2. Mutational shuffling/toggling in B*45:01 Nef EV11 epitope

Weeks post-infection	CAP63 Nef EV11	Frequency of variant	Weeks post-infection	CAP85 Nef EV11	Frequency of variant
2	EEVGFPVRPQV*	8/9	5	KEVGFPVRPQV*	9/10
	KEVGFPVRPQV	1/9		KEVGFP**I**RPQV	1/10
4	E**K**VGFPVRPQV	15/19	13	**ED**VGFPVRPQV	5/9
	EEVGFPVRPQV	1/19		KEVGFPVRPQV	2/9
	E**D**VGFPVRPQV	1/19		**EG**VGFPVRPQV	1/9
	E**G**VGFPVRPQV	1/19		K**G**VGFPVRPQV	1/9
	VEVGFPVRPQV	1/19			
5	E**G**VGFPVRPQV	6/13	29	**ED**VGFPVRPQV	4/7
	E**D**VGFPVRPQV	2/13		**EG**VGFPVRPQV	2/7
	KEVGFPVRPQV	2/13		**GD**VGFPVRPQV	1/7
	E**K**VGFPVRPQV	2/13			
	E**A**VGFPVRPQV	1/13			
7	E**G**VGFPVRPQV	9/10	39	**EG**VGFPVRPQV	1/1
	E**D**VGFPVRPQV	1/10			
11	E**G**VGFPVRPQV	4/7	55	**ED**VGFPVRPQV	2/4
	E**D**VGFPVRPQV	2/7		**GD**VGFPVRPQV	2/4
	DEVGFPVRPQV	1/7			
29	E**D**VGFPVRPQV	8/10			
	E**G**VGFPVRPQV	2/10			
37	E**D**VGFPVRPQV	5/5			

*Sequence of epitope in founder virus. Sites undergoing mutation are indicated in bold.