

THE ROLE OF RECOMBINATION IN THE EMERGENCE OF A COMPLEX AND DYNAMIC HIV EPIDEMIC

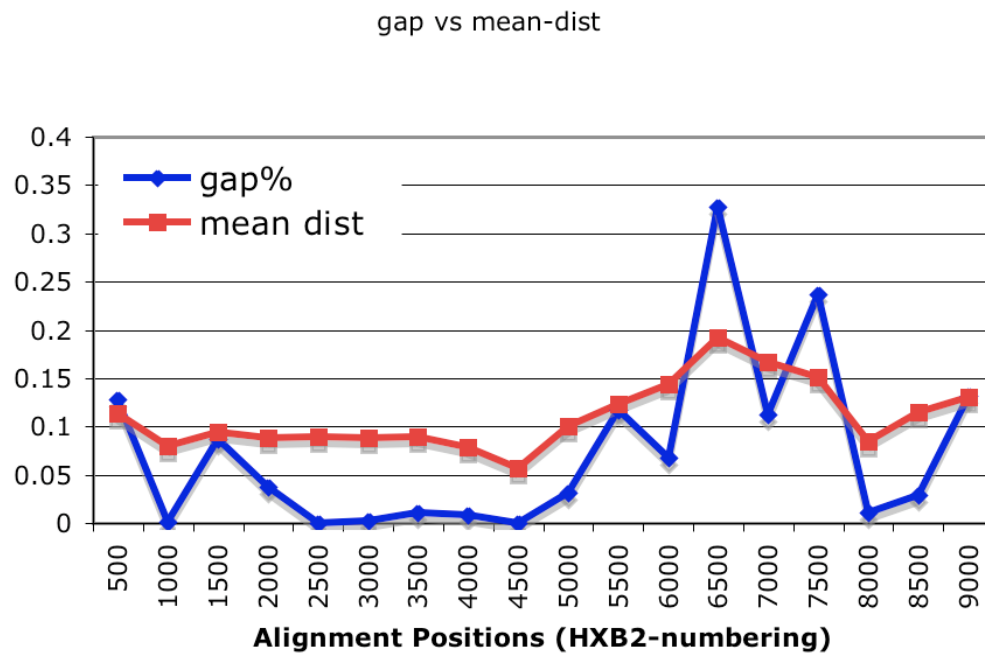
Ming Zhang, Brian Foley, Anne-Kathrin Schultz, Jennifer P. Macke, Ingo Bulla, Mario Stanke,
Burkhard Morgenstern, Bette Korber, Thomas Leitner

Supplementary figures:

Suppl. figure 1. Plot showing that our CRF02 conclusion, that is, CRF02 was derived from old and new recombination events, is not biased by the CRF02 sequence alignment quality.

Gap%: percentage of alignment gaps within every 500 nt window.

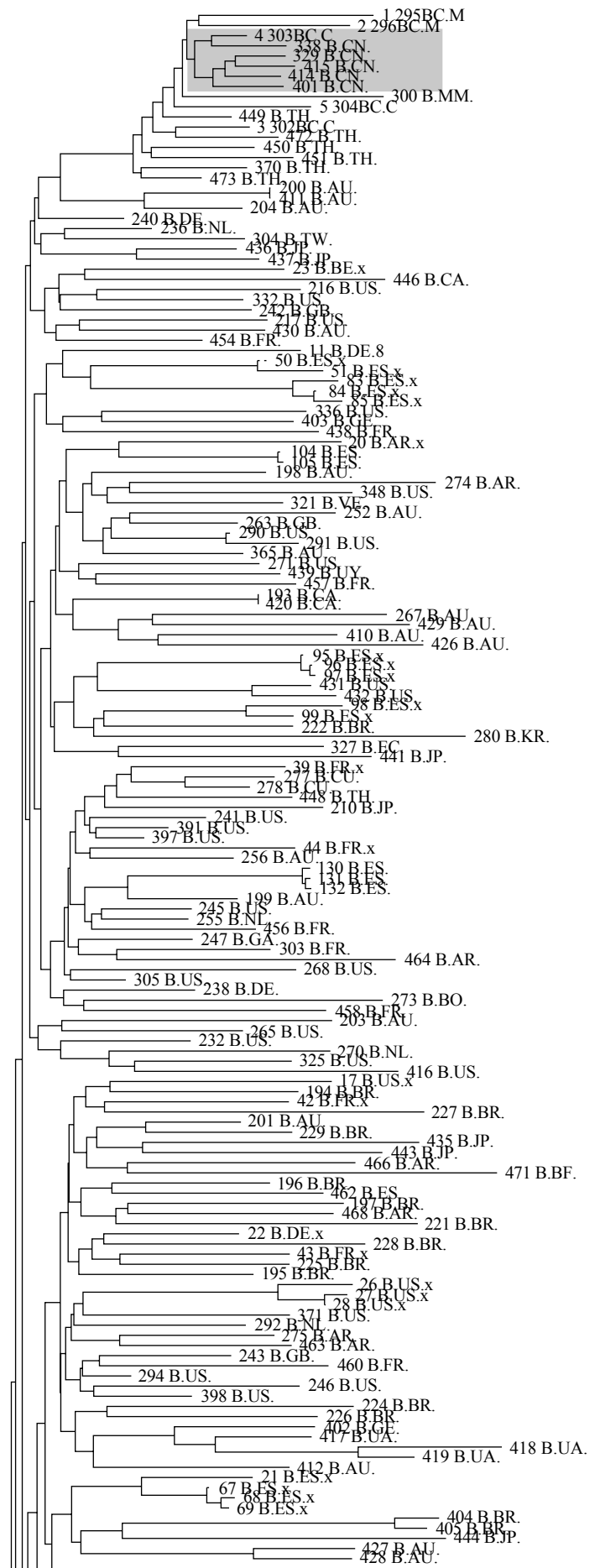
Mean dist: mean evolutionary distance (F84) within every 500 nt window.

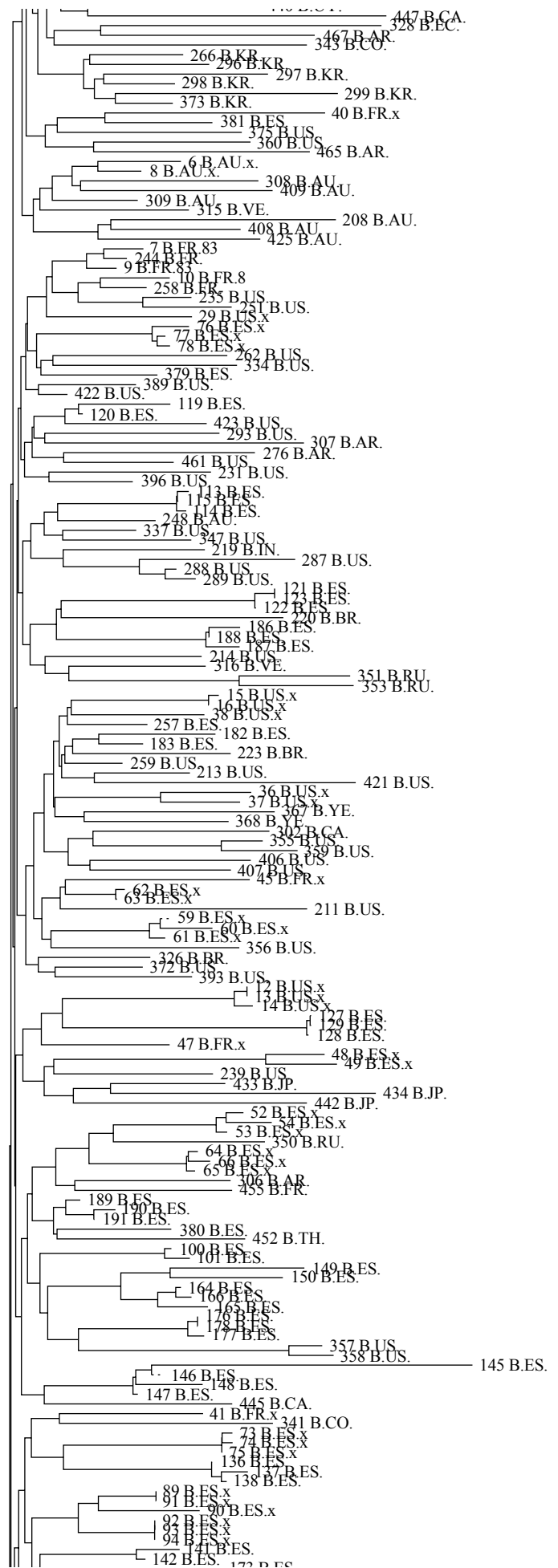


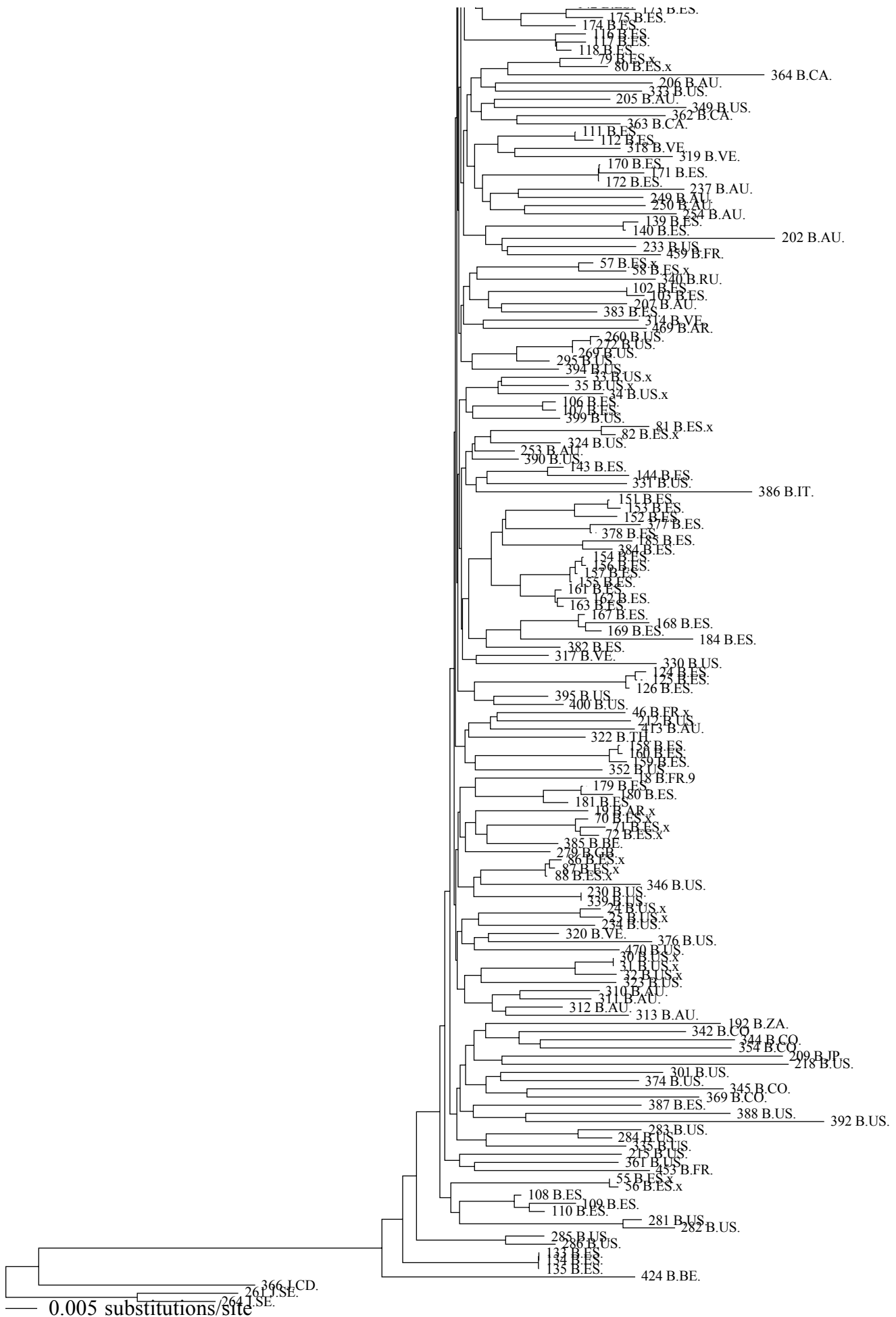
Suppl. figure 2. The BC epidemic in China is unique compared to China's neighboring countries.

Two neighbor-joining trees showing that China B and C are more restrictedly derived from China's neighboring countries, while the B and C sequences from China neighboring countries have more contacts with B and C from other regions of the world. Sequences used here were obtained from the worldwide sources (Set 3, as described in Methods). Each sequence name is in a standard format of number.subtype.country, where the number is a sequential identifier used in the sequence alignments.

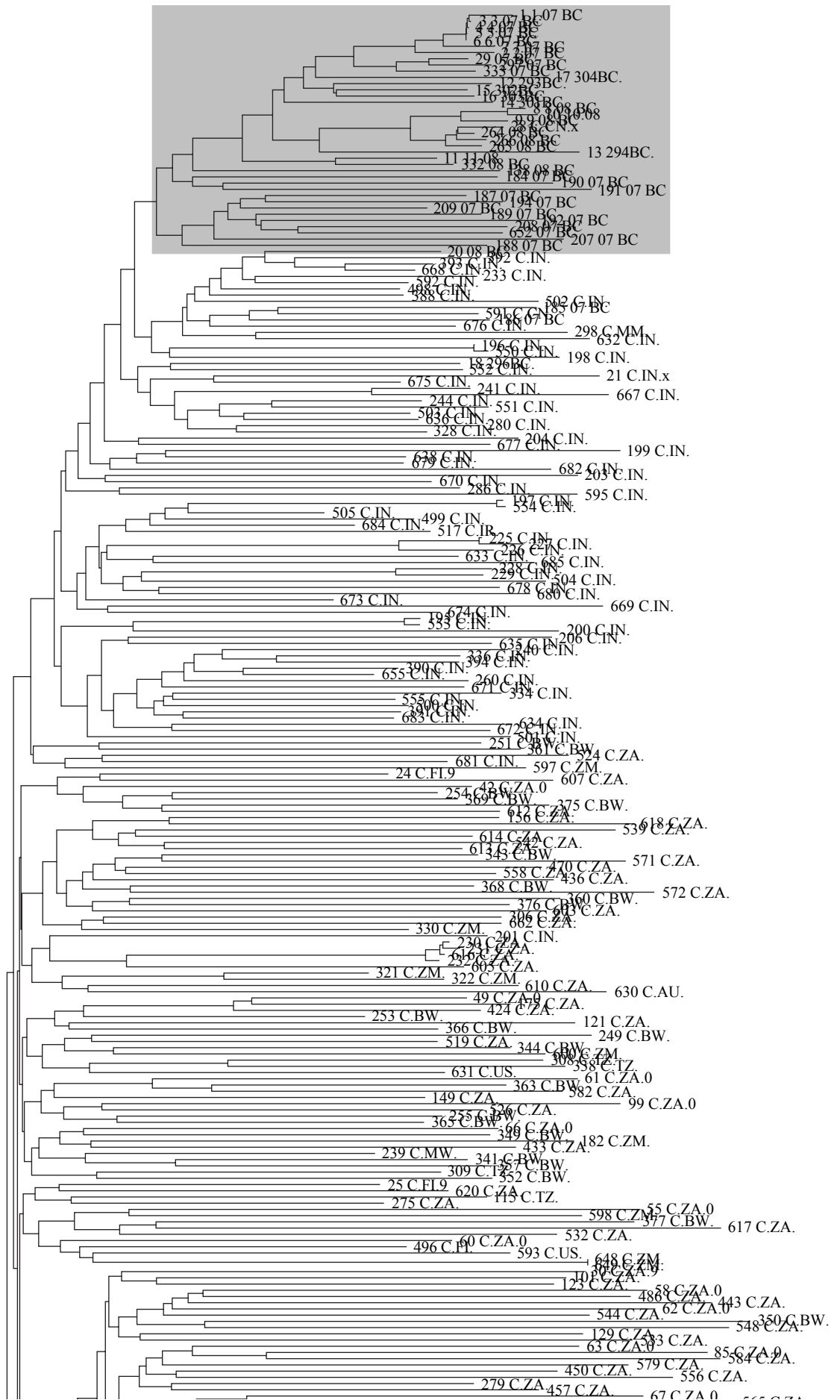
(A) Subtype B sequence fragments (HXB2 positions: 3497-4473)

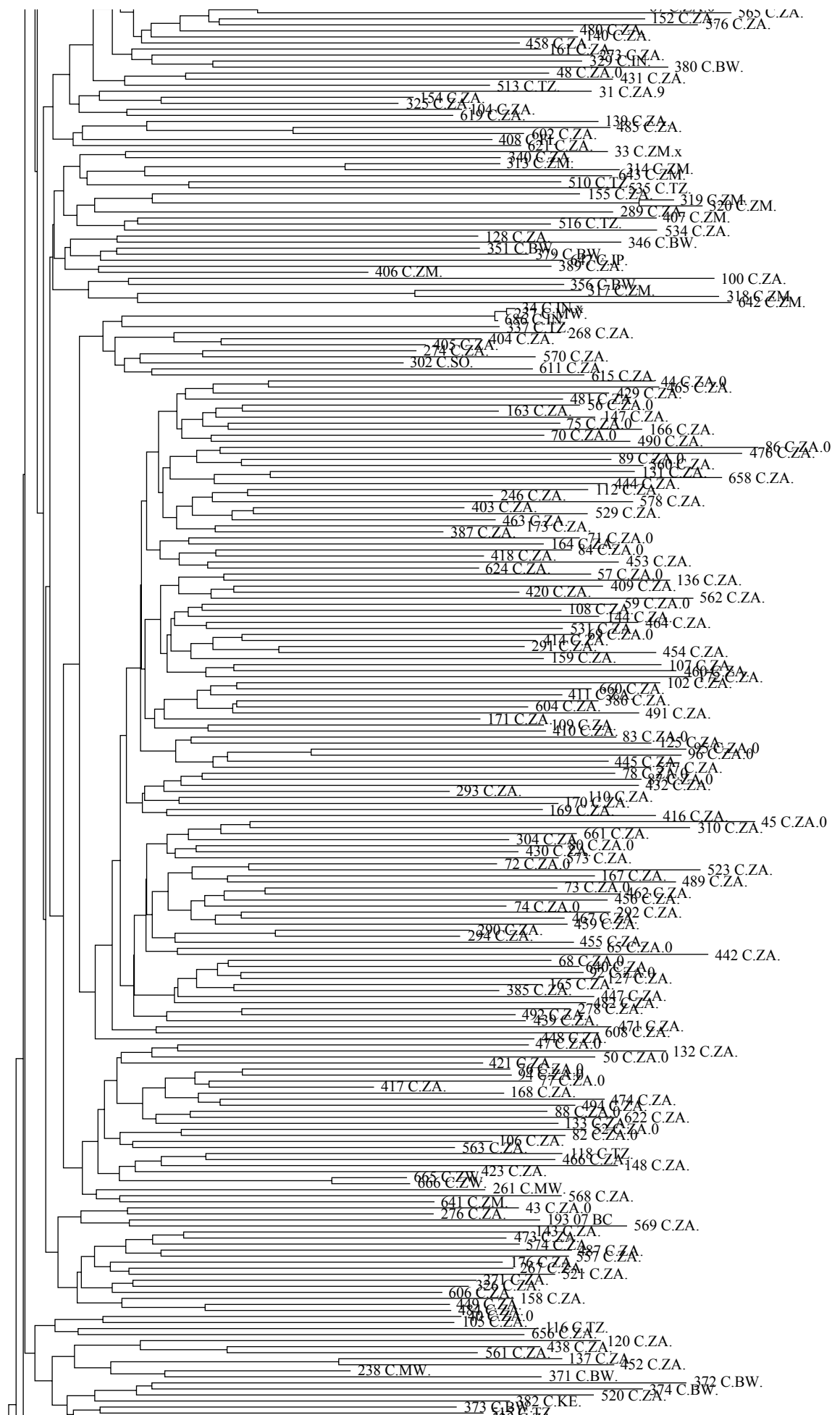


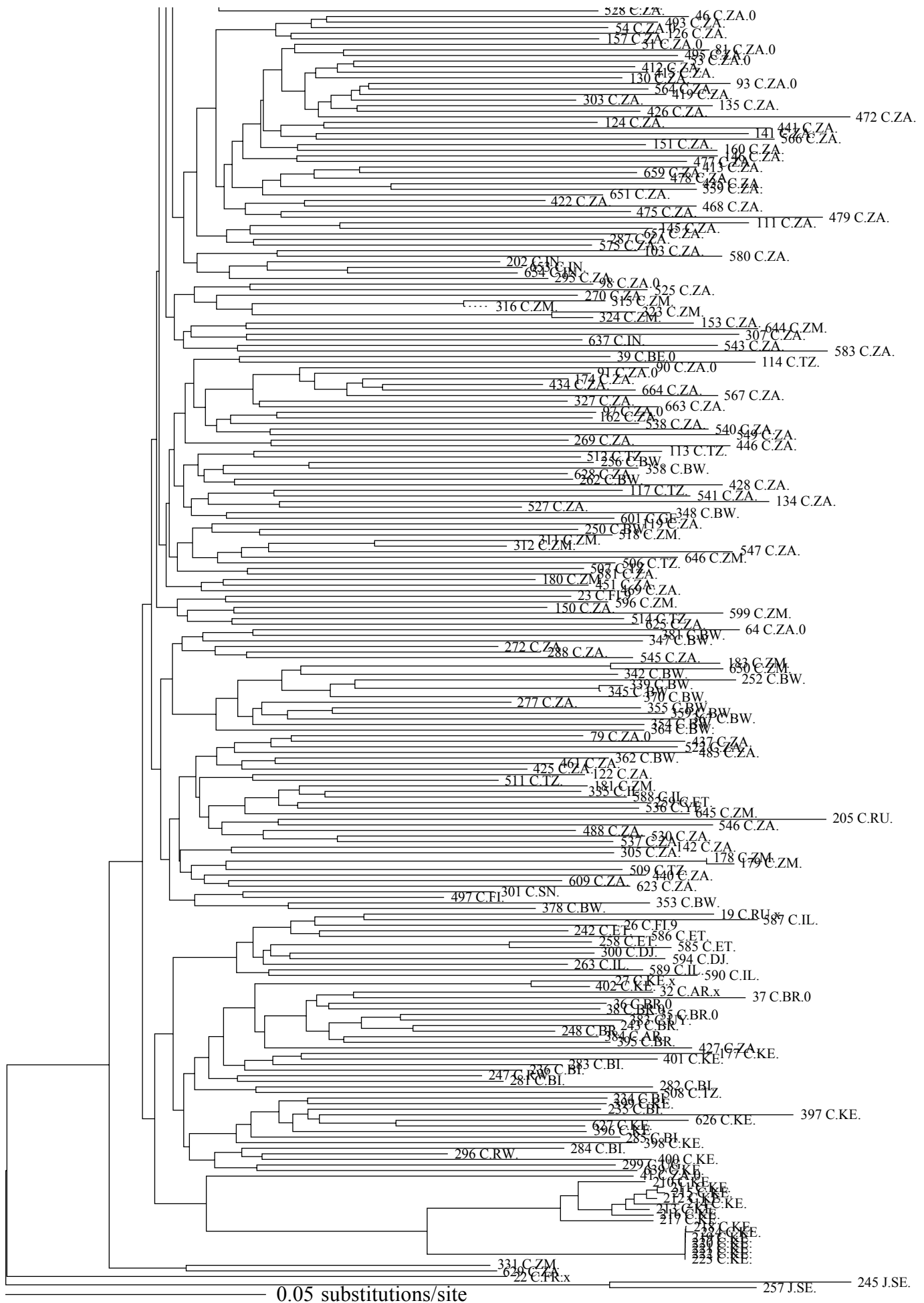




(B) Subtype C sequence fragments: HXB2 positions: 6582-7349.







Suppl. figure 3. The breakpoint frequency of CRF12, CRF28, and CRF29 sequences.

The breakpoint frequency = N/M , where N = the total number of BF recombinants that meet the following criteria: (1) breakpoints are within group breakpoint median value $\pm 98\text{nt}$, (2) the same subtypes flanking the breakpoint as in the CRF group; M = the total number of BF recombinants that span this genomic region. Solid arrow: breakpoints at fixed positions; open arrow: breakpoint region.

CRF12 group

Breakpoint number	1	2	3	4	5	6	7
Genome graph							
Breakpoint locations of this CRF							
Breakpoint range	953	2982	3679-3812	5946	6193-6229	8450-8485	8635-8669
Breakpoint median	953	2982	3713	5946	6229	8475	8635
Interquartile range (mean)	const. (953)	const. (2982)	3692-3713 (3722)	const. (5946)	6200-6229 (6216)	8475-8484 (8474)	8635-8669 (8649)
# of complete BF recombinants	Sequence total: 56 (Argentina: 22, Brazil: 23)						
Frequency of full-length BF recombinants that bear bk within median $\pm$ 98nt (world)	20/56	13/56	30/56	27/56	24/56	26/56	25/56
Frequency of full-length BF recombinants that bear bk within median $\pm$ 98nt (Brazil)	0/23	1/23	8/23	1/23	0/23	0/23	2/23
Frequency of full-length BF recombinants that bear bk within median $\pm$ 98nt (Argentina)	15/22	10/22	15/22	18/22	15/22	17/22	15/22
# of SouthAmerica BF fragments	Sequence total: 751 (Argentina: 639, Brazil: 109)						
Frequency of fragmental BF recombinants that bear bk within median $\pm$ 98nt (S. America)	0/2	333/685	11/11	0/0	25/28	4/5	1/5
Frequency of fragmental BF recombinants that bear bk within median $\pm$ 98nt (Brazil)	0/2	11/80	0/0	0/0	0/0	4/5	1/5
Frequency of fragmental BF recombinants that bear bk within median $\pm$ 98nt (Argentina)	0/0	322/603	11/11	0/0	25/28	0/0	0/0

CRF28 group:

Breakpoint number	1	2	
Genome graph			
Breakpoint locations Of this CRF			
Breakpoint range	1227-1398	2538-2565	
Breakpoint median	1329	2538	
Interquartile (mean)	1278-1364 (1318)	2538-2552 (2547)	
# of complete BF recombinants	Sequence total :56 (Argentina: 22, Brazil: 23)		
Frequency of full-length BF recombinants that bear bk within median $\pm$ 98nt (world)	15/56	29/56	
Frequency of full-length BF recombinants that bear bk within median $\pm$ 98nt (Brazil)	10/23	10/23	
Frequency of full-length BF recombinants that bear bk within median $\pm$ 98nt (Argentina)	2/22	10/22	
# of SouthAmerica BF fragments	Sequence total: 751 (Argentina: 639, Brazil:109)		
Frequency of fragmental BF recombinants that bear bk within median $\pm$ 98nt (S. America)	6/9	313/621	
Frequency of fragmental BF recombinants that bear bk within median $\pm$ 98nt (Brazil)	3/4	38/61	
Frequency of fragmental BF recombinants that bear bk within median $\pm$ 98nt (Argentina)	3/5	273/558	

CRF29 group

Breakpoint number	1	2	3	4	
Genome graph					
Breakpoint locations Of this CRF					
Breakpoint range	1236-1351	2518-2538	3734-3927	5233-5437	
Breakpoint median	1260	2538	3746	5368	
Interquartile (mean)	1248-1306 (1282)	2528-2538 (2531)	3740-3837 (3802)	5301-5403 (5346)	
# of complete BF recombinants	Sequence total: 56 (Argentina: 22, Brazil: 23)				
Frequency of full-length BF recombinants that bear bk within median $\pm$ 98nt (world)	16/56	29/56	30/56	7/56	
Frequency of full-length BF recombinants that bear bk within median $\pm$ 98nt (Brazil)	8/23	10/23	8/23	6/23	
Frequency of full-length BF recombinants that bear bk within median $\pm$ 98nt (Argentina)	4/22	10/22	15/22	1/22	
# of SouthAmerica BF fragments	Sequence total: 751 (Argentina: 639, Brazil: 109)				
Frequency of fragmental BF recombinants that bear bk within median $\pm$ 98nt (S. America)	4/9	313/621	11/11	0/0	
Frequency of fragmental BF recombinants that bear bk within median $\pm$ 98nt (Brazil)	1/4	38/61	0/0	0/0	
Frequency of fragmental BF recombinants that bear bk within median $\pm$ 98nt (Argentina)	3/5	273/558	11/11	0/0	