African Journal of Biotechnology Vol. 10(24), pp. 4784-4793, 6 June, 2011
Available online at http://www.academicjournals.org/AJB
DOI: 10.5897/AJB10.560
ISSN 1684–5315 © 2011 Academic Journals

Full Length Research Paper

Immune pressure analysis of protease and reverse transcriptase genes of primary HIV-1 subtype C isolates from South Africa

Pascal O. Bessong* and McLaughlin N. Rathogwa

AIDS Virus Research Laboratory, Department of Microbiology,University of Venda,PMB X5050 Thohoyandou 0950, South Africa.

Accepted 7 January, 2011

Synonymous (ds) and non-synonymous (dn) substitution rates and their ratios (ds/dn) were analyzed for 33 HIV-1 subtype C protease (PR) and reverse transcriptase (RT) nucleotide sequences each from antiretroviral naïve South African chronically infected individuals. The ds/dn ratios were calculated using the synonymous/non-synonymous analysis programme (SNAP). The ds/dn ratio for PR was found to be 5.3801, while for RT it was 8.6098. The RT had a higher ds/dn ratio when compared with PR, suggesting a greater conservation in the RT gene. Generally, these values point to proteins that have not been subjected to strong immune pressure. Analysis of the evolutionary distance of PR and RT based on synonymous and non-synonymous mutations was done by phylogenetic analysis. The clustering of viruses was different when synonymous and non-synonymous substitutions were considered in both gene regions. Overall, the data indicated that the viruses had not experienced high immune pressure at the time the sequences were obtained. It may also mean that, the PR and the RT of the sequences did not present enough epitopes to elicit antibody responses.

Key words: HIV-1 subtype C, chronic infection, immune pressure, protease gene, reverse transcriptase gene, South Africa.

INTRODUCTION

Human immunodeficiency virus (HIV) presents a high degree of genetic variability permitting the classification of isolates into types, groups, subtypes and recombinant forms. Genetic variability is attributed to the error prone reverse transcriptase (RT) that lack proof reading functions (3'-5' exonuclease activity), the diploid nature of the viral genome, the propensity for recombination and the high viral replication turn over. Understanding the genetic diversity of HIV is important for the following reasons: the genetic landscape impacts on the efficiency of diagnostics used for infection detection and patient monitoring such as viral RNA measurements; influences the efficacy of therapy and guides the selection of genes for vaccine development (Esparza and Bhamarapravati, 2000; Kantor et al., 2005; Bessong, 2008; Barouch and Korber, 2010).

Characterization of HIV sequences from the Limpopo Province has mainly focused on genetic variability and drug resistance (Bessong et al., 2005, 2006). However, it is important to have information on the mutation profiles due to immune pressure, since this can indicate the presence of epitopes elicitin viral immune response. The aim of this study was to evaluate the immune pressure on protease (PR) and RT gene regions of HIV-1 subtype C isolated from chronically infected individuals from the Limpopo Province of South Africa. The study specifically sought to determine the synonymous (ds) and non-synonymous (dn) mutation rates and the synonymous to non-synonymous (ds/dn) ratios of the reverse transcriptase and protease nucleotides sequences of HIV-1 subtype C primary isolates.

*Corresponding author. E-mail: bessong@univen.ac.za. Tel: +27 15 962 8301.

Table 1. HIV-1 subtype C isolates and accession numbers analyzed for immune pressure (Thirty three isolates were analyzed on both protease and reverse transcriptase regions).

S/N	Protease nucleotides sequence		Reverse transcriptase nucleotides sequence	
	Isolate number	GenBank accession number	Isolate number	GenBank accession number
1	01PB05ZA	DQ222243	01PB5ZA	DQ222283
2	01PB06ZA	DQ222244	01PB6ZA	DQ222284
3	01PB09ZA	DQ222245	01PB9ZA	DQ222285
4	01PB10ZA	DQ222246	01PB10ZA	DQ222286
5	01PB11ZA	DQ222247	01PB11ZA	DQ222287
6	01PB24ZA	DQ222249	01PB24ZA	DQ222289
7	01PB25ZA	DQ222250	01PB25ZA	DQ222290
8	01PB31ZA	DQ222251	01PB31ZA	DQ222291
9	01PB32ZA	DQ222252	01PB32ZA	DQ222292
10	01PB34ZA	DQ222253	01PB34ZA	DQ222293
11	01PB35ZA	DQ222254	01PB35ZA	DQ222294
12	01PB36ZA	DQ222255	01PB36ZA	DQ222295
13	01PB39ZA	DQ222256	01PB39ZA	DQ222296
14	01PB40ZA	DQ222258	01PB40ZA	DQ222297
15	01PB41ZA	DQ222257	01PB41ZA	DQ222298
16	01PB42ZA	DQ222259	01PB42ZA	DQ222299
17	02TS03ZA	DQ222262	02TS3ZA	DQ222301
18	02TS14ZA	DQ222263	02TS14ZA	DQ222302
19	02TS16ZA	DQ222266	02TS16ZA	DQ222303
20	02TS17ZA	DQ222267	02TS17ZA	DQ222304
21	02TS18ZA	DQ222268	02TS18ZA	DQ222305
22	02TS19ZA	DQ222269	02TS19ZA	DQ222306
23	02TS22ZA	DQ222270	02TS22ZA	DQ222307
24	02TS26ZA	DQ222272	02TS26ZA	DQ222308
25	02TS30ZA	DQ222265	02TS30ZA	DQ222309
26	04J8129ZA	DQ222273	04J8129ZA	DQ222310
27	04J8717ZA	DQ222274	04J8717ZA	DQ222311
28	04J8820ZA	DQ222276	04J8820ZA	DQ222312
29	04J8832ZA	DQ222277	04J8832ZA	DQ222313
30	04J8839ZA	DQ222278	04J8839ZA	DQ222314
31	04J8840ZA	DQ222279	04J8840ZA	DQ222315
32	04J8946ZA	DQ222280	04J8946ZA	DQ222316
33	04J8980ZA	DQ222281	04J8980ZA	DQ222317

MATERIALS AND METHODS

Ethical considerations

Ethical clearance for the study was obtained from the Safety, Health and Research Ethics Committee of the University of Venda, to use already described anonymous non-linked DNA sequences.

Nucleotides sequences

The protease (PR) and reverse transcriptase (RT) nucleotide sequences of HIV -1 subtype C used in the analyses are shown in Table 1. Thirty three isolates were used; a PR and a RT gene for each isolate. The isolates were obtained from antiretroviral naïve South African patients who have been chronically infected. The mean genetic distances and drug resistance associated mutations for the data set had been previously described (Bessong et al., 2006) and showed that, there was very little intra-genetic diversity among the viruses and further suggests that the viruses would be susceptible to the current treatment regimen in South Africa.

Phylogenetic analysis of sequences based on synonymous and non-synonymous mutations

Nucleotides sequences were aligned using SeqPublish, a program available on the HIV sequence database (www.hiv.lanl.gov/content/sequence). All aligned PR and RT nucleotide sequences were used to create a phylogenetic N-J tree based on synonymous distances and a phylogenetic N-J tree based on non-synonymous distances. Codon by codon XY-plot graph were drawn for PR and RT nucleotides sequence based on the synonymous (ds) and non-synonymous (dn) substitution using the synonymous non-

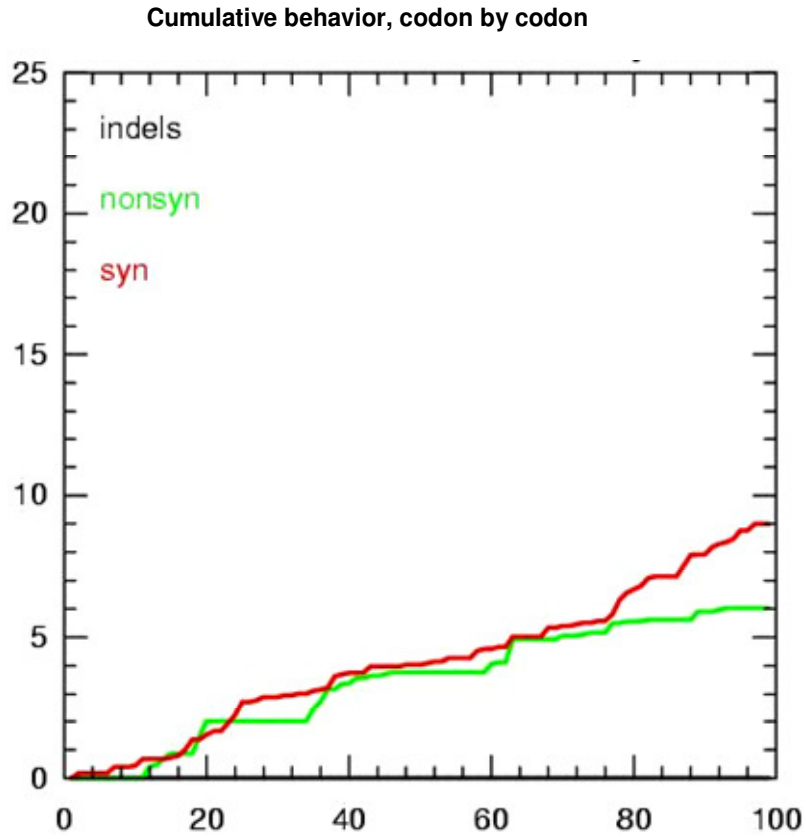

Figure 1. Codon by codon XY-plot of synonymous (ds) and non-synonymous (dn) substitution of the protease sequences. Protease sequences described in genbank from South African HIV subtype C isolates were aligned using seqpublish. XY-plot of ds and dn substitution was constructed using a program available in the SNAP (synonymous and non-synonymous analysis program). The graph shows the sequence relationship among sequences based on synonymous and non-synonymous mutations rates, and the behavior of each functional codon along the coding region of the protease gene.

synonymous analysis program (SNAP) available from HIV database (www.hiv.lanl.gov/content/sequence/HIV/HIVTools.html).

Synonymous (ds) and non-synonymous (dn) substitution analysis

Aligned PR and RT were analyzed for the rates of synonymous (ds) to non-synonymous (dn) substitution using the SNAP program.

RESULTS

Synonymous and non-synonymous substitution mutation of protease nucleotides sequence

Codon by codon XY-plot of the protease nucleotides sequence based on the Synonymous (ds) and non-synonymous (dn) substitution

Both synonymous and non-synonymous substitutions started with a low constant rate at the start codon of the PR gene. However, the non-synonymous rate was lower when compared with the synonymous substitution. There was a high rate of synonymous substitution and a low rate of non-synonymous substitution in PR, moving across the coding region of the gene. In the coding region between positions 22 and 37 and position 42 and 59, there was a constant rate of non-synonymous substitution mutations. Generally, the rate of synonymous substitutions was constant between positions 41 and 44. It was observed that, both synonymous and non-synonymous substitutions were constant between positions 61 and 63 (Figure 1).

Phylogenetic analyses of protease sequences based on synonymous and non-synonymous rates of mutations

Separate analyses on the phylogeny of the isolates with regards to synonymous and non-synonymous mutations were made. It was observed that for some sequences,

the clustering differed when synonymous mutations were considered independently of non-synonymous mutations (Figures 2 and 3).

Synonymous and non-synonymous substitution mutation of reverse transcriptase nucleotides sequence

Codon by codon XY-plot of the reverse transcriptase nucleotides sequence based on the synonymous (ds) and non-synonymous (dn) substitution

The behavior with respect to changes across the RT nucleotide sequences is shown in Figure 4. Generally, a high rate of synonymous mutation was observed across the gene for all the sequences when compared with the non-synonymous mutations.

Phylogenetic analyses of reverse transcriptase sequence on synonymous and non-synonymous rates of mutation

Separate phylogenetic analyses of RT sequences based on synonymous and non-synonymous substitutions were constructed (Figure 5 and 6). The generated trees showed that the clustering pattern of sequences was not the same for both set of analysis. This indicates that the types of mutations influence the appearance of the evolutionary distance of a sequence. Sequences with more synonymous rates of mutations tended to cluster together, likewise sequences with non-synonymous rates of mutations.

Synonymous (ds) to non-synonymous (dn) substitution ratio of PR and RT

The 33 nucleic acid sequences of PR and RT of HIV-1 subtype C (Table 2) were also analyzed using SNAP for the synonymous to non-synonymous amino acid substitution ratio (ds/dn) as a measure of natural selection at protein level. The analysis showed that, the values of ds were higher when compared with dn in PR (0.1458 ds versus 0.0271 dn). The values of ds were also higher when compared with RT (0.1920 ds versus 0.0223). The values of ds were substantially lower in PR (0.1458) when compared with the value of ds in RT (0.1920), but the value of dn in PR (0.0271) was higher when compared with dn in RT (0.223). The RT had a ds/dn ratio higher than PR (8.6098 ds/dn RT and 5.3801 ds/dn PR) thus, suggesting a greater conservation in the RT gene.

DISCUSSION AND CONCLUSION

The synonymous or silent (ds) and non-synonymous or non-silent (dn) substitution rates that occurred in PR and RT genes of HIV-1 subtype C viruses from chronically infected, antiretroviral naïve patients were analyzed. In the absence of positive selection, substitution that are silent (synonymous mutations) would be expected to exceed substitution that lead to amino acid changes (non-synonymous mutations), since most of the substitutions that lead to amino acid changes or structural changes in a protein are many times deleted by the DNA repair mechanism of the cell in a bid to maintain functionality.

A ds/dn ratio of less than 1 indicates a strong immune pressure since non-synonymous mutations are more than synonymous mutations. In the study, the ds/dn ratio of both the PR and RT regions were more than 1. In fact ds/dn for PR was 5.3801 and a ds/dn for RT was 8.6098. Generally, these values point to proteins that have not been subjected to strong immune pressure. The lower ds/dn ratio for the PR sequences is a reflection of the higher variability in the PR gene when compared with the RT gene that had been generally observed (Kantor et al., 2005; Bessong et al., 2006). In an analysis of PR and RT sequences from Mozambique of antiretroviral naïve patients, a ds/dn ratio for the PR and RT genes of HIV-1 subtype C was found to be 0.116 and 0.093, respectively (Ballocchi et al., 2005). These values are much lower than those obtained in this study. A high ds/dn ratio indicates that, the sequences did not undergo adequate immune pressure to lead to changes in amino acids. This could be explained by the fact that, the RT and PR genes are internal components of the virus unlike the envelope proteins which are external and exposed to high immune pressure from the host.

An analysis of the evolutionary distance of PR and RT based on synonymous and non-synonymous mutations was also done by phylogenetic analysis. The clustering of viruses was different when non-synonymous and synonymous substitutions were considered in the PR phylogenetic analysis. For example, isolates 01PB39ZA and 04J8946ZA appeared to have evolved the most in the synonymous analysis than in the non-synonymous analysis. On the non-synonymous estimate isolates, 04J-8820ZA and 02TS17ZA appeared to have evolved the most. On the other hand, isolates 01PB11ZA and 04J8980ZA; 01PB31ZA and 01PB05ZA appeared to have evolved the least with the pairs having similar evolutionary distances.

The clustering of viruses was also different for both the non-synonymous and synonymous in the RT phylogenetic analysis. For example, isolates 02TS17ZA, 04J8980ZA and 04J8832ZA appeared to have evolved the most in the synonymous analysis than in the non-synonymous analysis. Isolates 01PB05ZA, 02TS26ZA and 02TS22ZA appeared to have evolved the least having the same evolutionary distance in the synonymous analysis. On the non-synonymous estimate, isolates 01PB39ZA and 0J8832ZA appeared to have evolved

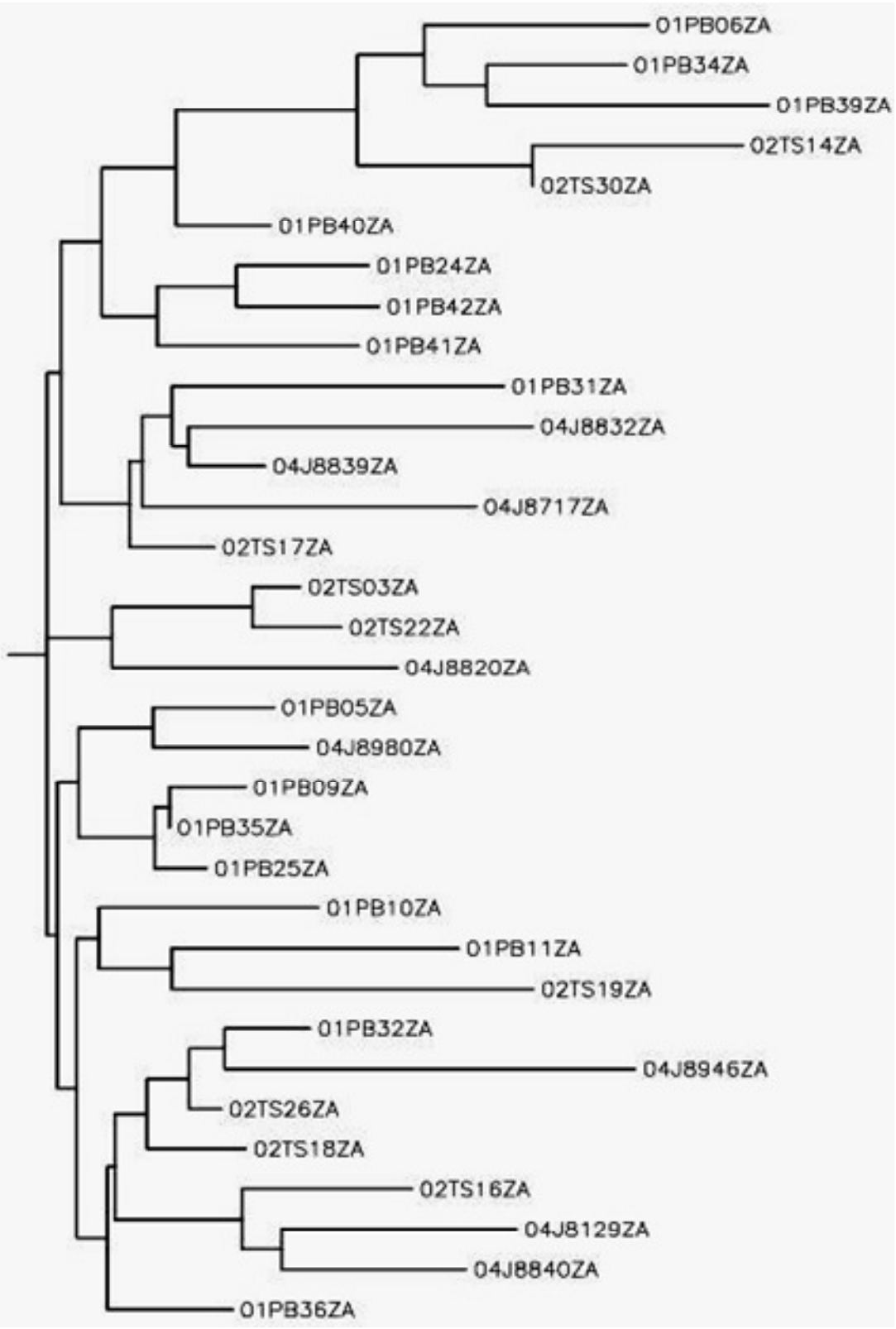

Figure 2. Phylogenetic analysis of protease sequences based on synonymous rates of mutation. Protease sequences described in GenBank from South African HIV subtype C isolates were aligned using Seqpublish. A synonymous (ds) tree of protease sequences was constructed using a program available in the SNAP (synonymous and non-synonymous analysis program). The tree shows the relationship among the sequences based on synonymous rates.

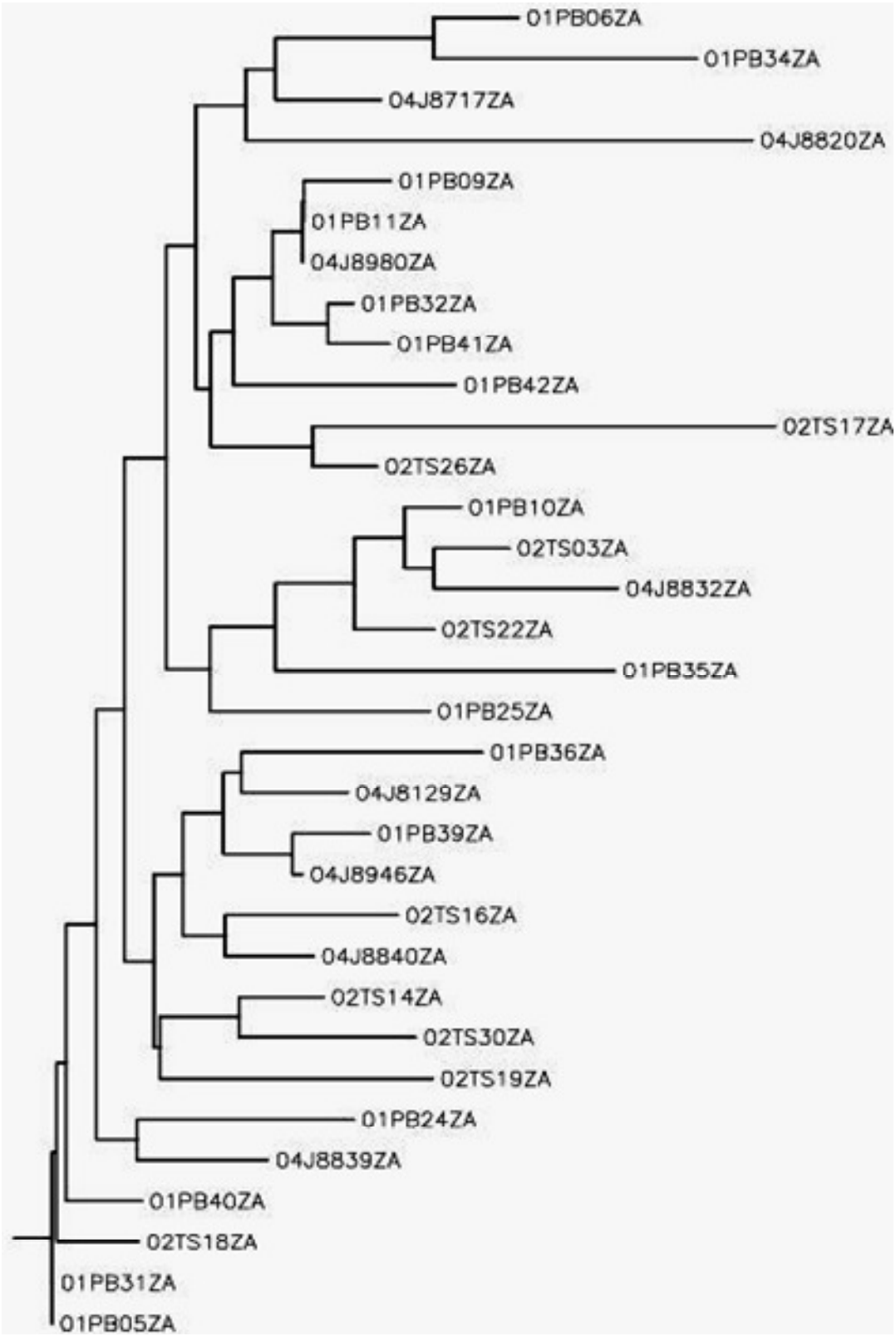

Figure 3. Phylogenetic analysis of protease sequences based on non-synonymous rates of mutation. Protease sequences described in GenBank from South African HIV subtype C isolates were aligned using Seqpublish. A non-synonymous (ds) tree of protease sequences was constructed using a program available in the SNAP (synonymous and non-Synonymous analysis program). The tree shows the relationship among the sequences based on non-synonymous rates.

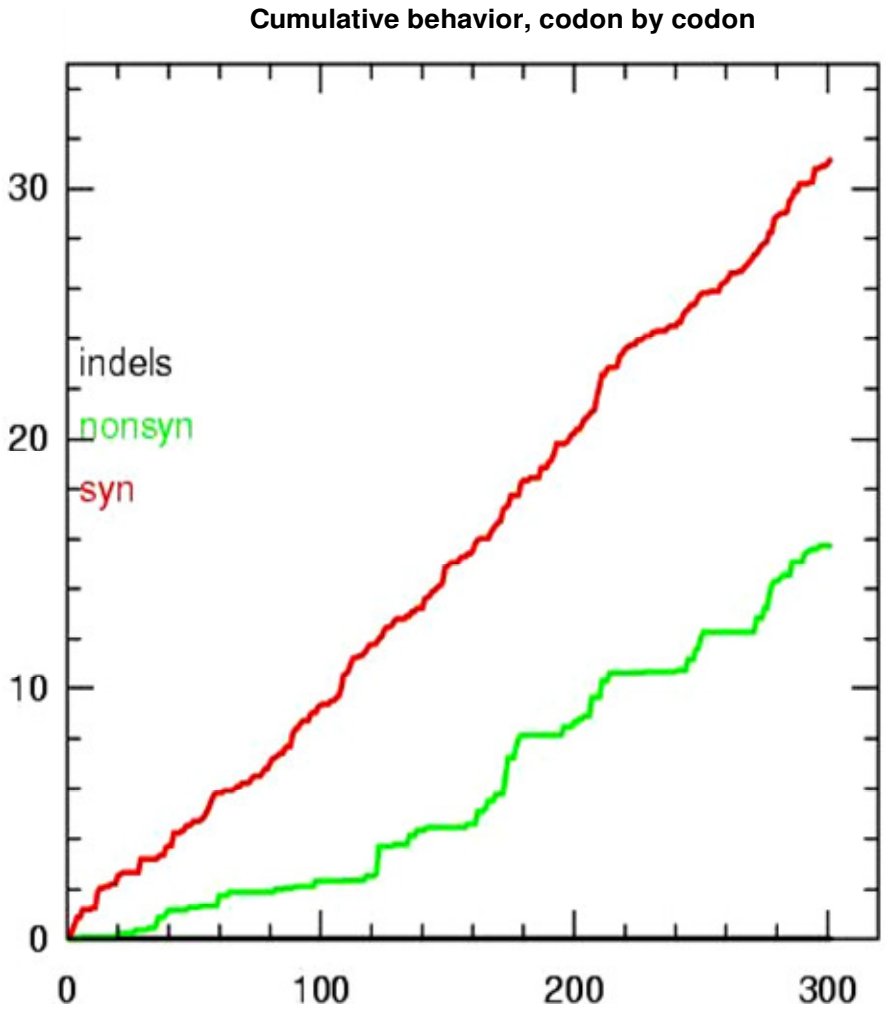

Figure 4. Codon by codon XY-plot of synonymous (ds) and non-synonymous (dn) substitution of the reverse transcriptase sequences. Reverse transcriptase sequences described in GenBank form South African HIV subtype C isolates were aligned using seqpublish. XY-plot of ds and dn substitution was constructed using a program available in the SNAP (synonymous and non-synonymous analysis program). The graph shows the sequence relationship among sequences based on synonymous and non-synonymous mutations rates, and the behavior of each functional codon along the coding region of the gene.

Table 2, The ratio of synonymous to non-synonymous substitution (ds/dn) of both PR and RT sequences.

Gene region	Synonymous substitution (ds)	Non - synonymous substitution (dn)	Ratio of ds/dn
PR	0.1458	0.0271	5.3801
RT	0.1920	0.0223	8.6098

Nucleic acid sequences of each of 33 reverse transcriptase and protease GenBank were analyzed for ds/dn ratio using SNAP (synonymous to non-synonymous analysis program). The software program used does not allow the insertion of genetic distances. However, according to the publish work of Bessong et al. (2006), the genetic distance in the phylogenetic analysis was 0.1.

the most, while isolates 02TS17ZA appeared to have evolved the least in the non-synonymous analysis than in the synonymous analysis. Mutations do affect the behavior of sequences in a phylogenetic analysis particularly when the analysis assumes a constant rate of change for all the sequences in the data set. Consequently, the change in the clustering of sequences in the phylogenetic trees based on mutations was not surprising. In conclusion, these data indicate that, the viruses had not experienced high immune pressure at the time the

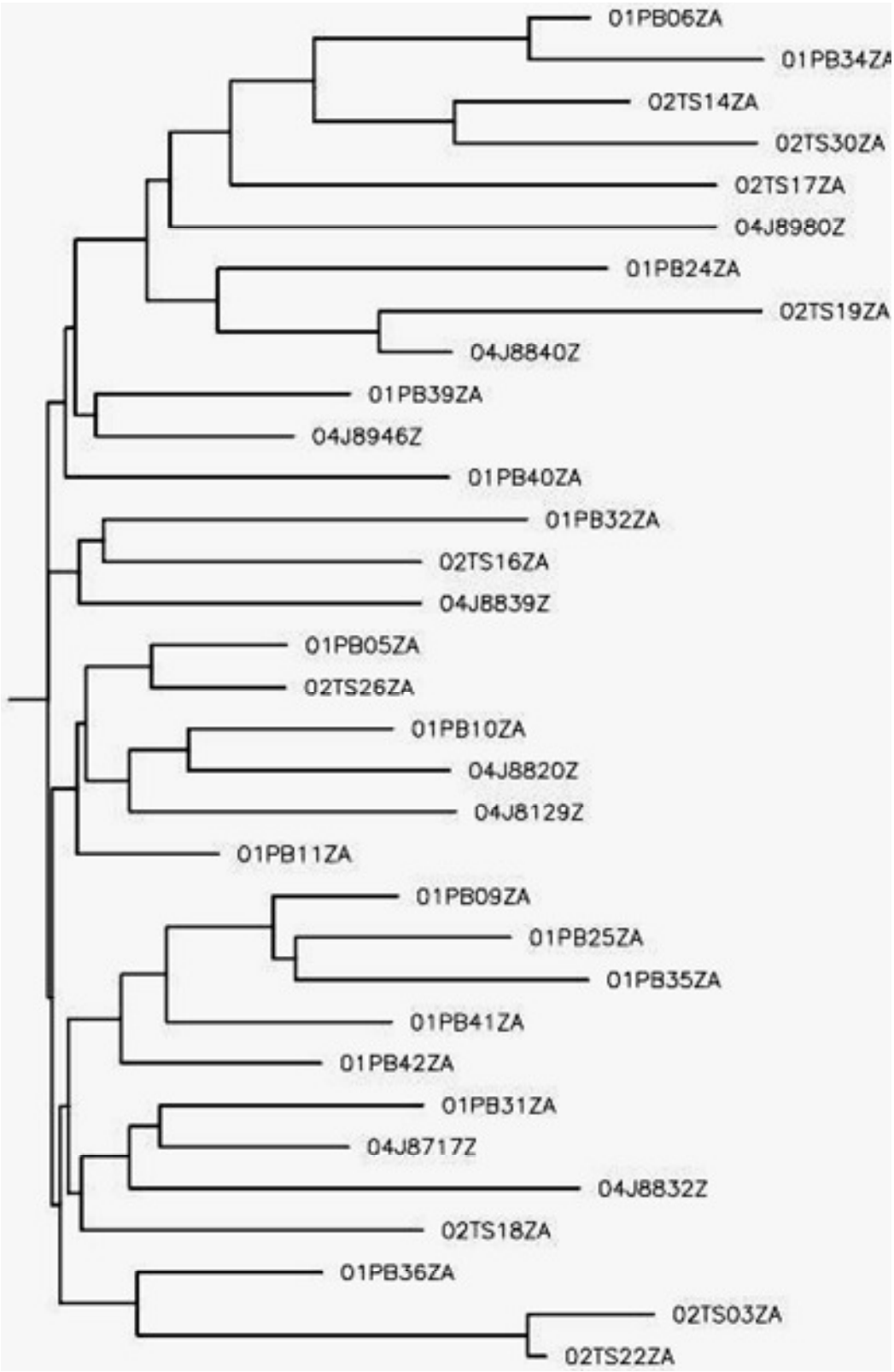

Figure 5. Phylogenetic analysis of reverse transcriptase sequences based on synonymous rates of mutation. Reverse transcriptase sequences described in Genbank from South African HIV subtype C isolates were aligned using Seqpublish. A synonymous (ds) tree of reverse transcriptase sequences was constructed using a program available in the SNAP (synonymous and non-synonymous analysis program). The tree shows the relationship among the sequences based on synonymous rates.

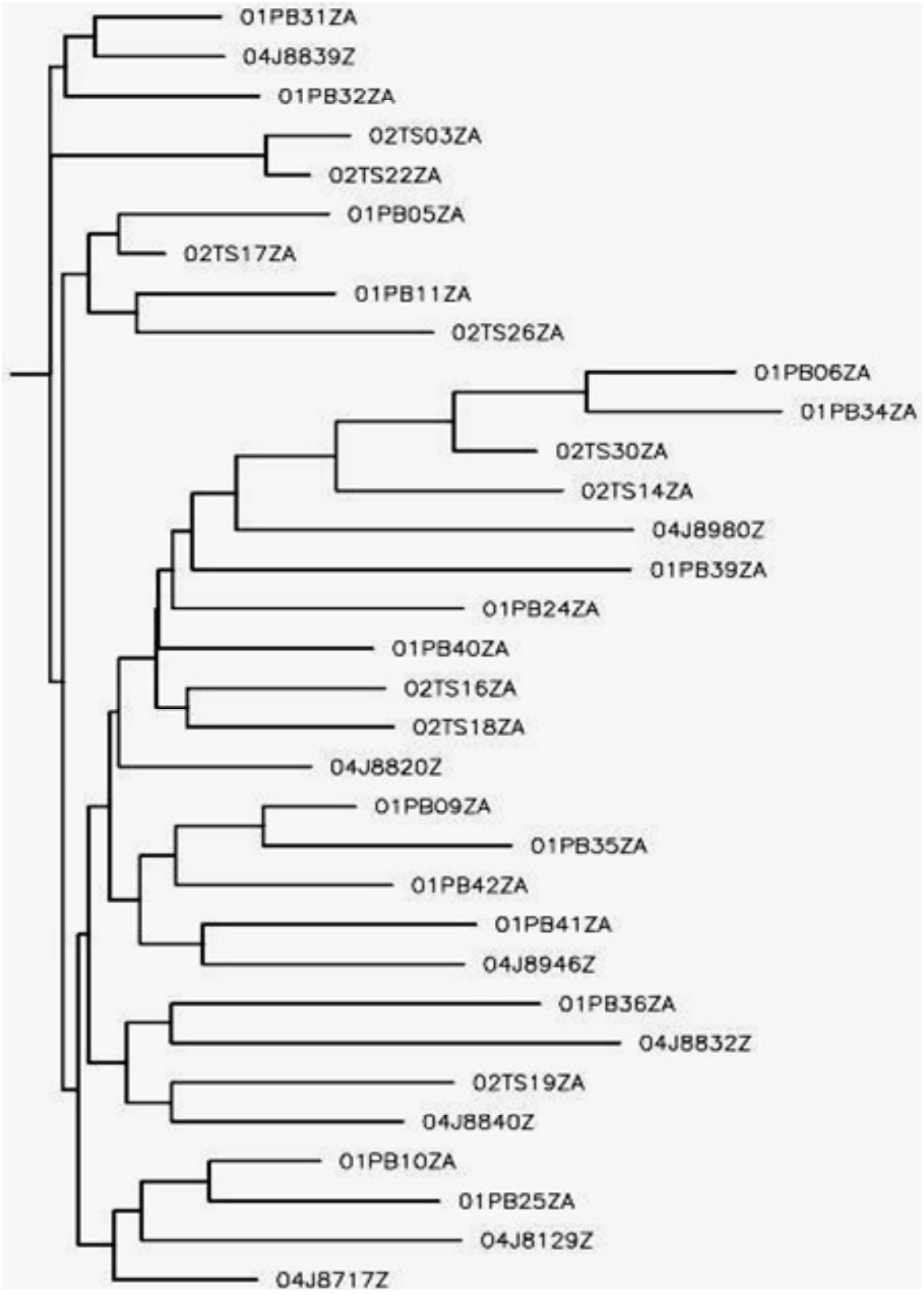

Figure 6. Phylogenetic analysis of reverse transcriptase sequences based on non-synonymous rates of mutation. Reverse transcriptase sequences described in Genbank from South African HIV subtype C isolates were aligned using Seqpublish. A non-synonymous (ds) tree of reverse transcriptase sequences was constructed using a program available in the SNAP (Synonymous and non-synonymous analysis program). The tree shows the relationship among the sequences based on non-synonymous rates.

sequences were obtained. It may also mean that the PR and RT genes did not present enough epitopes to elicit strong immune responses.

REFERENCES

Ballocchi MC, Forbic F, Palombi L, Gori C, Coelho E, Svicher V, D'Arrigo R, Emberti-Gialloreti L, Ceffa S, Erba F, Marazzi MC, Silberstein FC, Perno CF (2005). Subtype analysis and mutations to antiviral drugs in HIV-1 infected patients from Mozambique before initiation of antiretroviral therapy: results from the DREAM programme. J. Med. Virol. 76: 452-458.

Barouch DH, Korber B (2010). HIV-1 vaccine development after STEP. Ann. Rev. Med. 61:153-167.

Bessong PO (2008). Polymorphisms in HIV-1 subtype C proteases and their potential impact on protease inhibitors. Trop. Med. Int. Health, 13: 141-152.

Bessong PO, Mphahlele J, Choge I, Obi CL, Morris L, Hammarskjold ML, Rekosh D (2006). Resistance mutational analysis of HIV-1 subtype C among rural South African drug naïve patients prior to large scale availability of antiretrovirals. AIDS Res. Hum. Retroviruses, 22: 1306-1312

Bessong PO, Obi CL, Cilliers T, Phoswa M, Papathanasopoulos M, Moris L (2005). Characterization of human immunodeficiency virus type-1 from a previously unexplored region of South Africa with a high HIV prevalence. AIDS Res. Hum. Retroviruses, 21: 103-109.

Esparza J, Bhamarapravati N (2000) Accelerating the development and future availability of HIV-1 vaccines: why, when, where, and how? Lancet, 355: 2061-2066.

Kantor R, Katzenstein DA, Efron B, Carvalho AP, Wynhoven B, Cane P, Clarke J, Sirivichayakul S, Soares MA, Snoeck J, Pillay C, Rudich H, Rodrigues R, Holguin A, Ariyoshi K, Bouzas MB, Cahn P, Sugiura W, Soriano V, Brigido LF, Grossman Z, Morris L, Vandamme AM, Tanuri A, Phanuphak P, Weber JN, Pillay D, Harrigan PR, Camacho R, Schapiro JM, Shafer RW (2005). Impact of HIV-1 subtype and antiretroviral therapy on protease and reverse transcriptase: results of a global collaboration. PLOS Med. 2:e112.