The COLATE trial: Comparison of the Evolutionary Distance for Protease and Reverse Transcriptase Sequences

J Kjær[1], Z Fox[1], UB Dragsted[1], J Gerstoft[2], JD Lundgren[1] and the COLATE study group

1 Copenhagen HIV Programme, Copenhagen, Denmark; 2 Rigshospitalet, Copenhagen, Denmark

BACKGROUND & OBJECTIVES

Virological failure on a regimen containing lamivudine (3TC) is associated with emergence of the M184I/V mutation. Previous studies suggest a higher fidelity of the mutant reverse transcriptase. Hypothetically, the presence of the M184I/V mutation would lead to a reduced rate over time in the evolution of other mutations in patients continuing 3TC compared to those discontinuing 3TC.

METHOD

In the COLATE trial, patients failing a 3TC containing regimen (HIV-RNA >1000 c/ml) were randomised to:

- continue 3TC (3TC arm)

  or
- discontinue 3TC (No-3TC arm)

whilst continuing or initiating a new three-drug regimen. All available cryopreserved plasma samples were used for protease and reverse transcriptase sequencing at baseline and during follow-up if HIV-RNA >500 c/ml. Patients were only included if baseline and ≥ 1 follow-up sequence was available.

Phylogenetic trees were created using the Neighbour-Joining distance method and the PAUP program was used to calculate the number of changed nucleotides since baseline (Jukes-Cantor distance).

Two additional evolutionary nucleotide methods were applied:

- The Kimura 2-parameter distance that measures distances after accounting for the transversion/transition ratio*.
- The Syn-Scan application that counts both synonymous (dS)/non-synonymous (dN) nucleotide changes# and takes into account the information held by the ambiguous nucleotide letters@ .

Kruskal-Wallis non-parametric tests were used to test the differences between the 3TC arm and the No-3TC arm at all scheduled visits. Using the Bonferroni correction for multiple testing a p-value of ≤0.01 was considered to be significant.

RESULTS

We obtained a total of:

- 67 sequences in 19 patients (3TC arm) and
- 67 sequences in 23 patients (No-3TC arm)

All sequences were 1302 nucleotides long.

Four sequences in 4 patients in the No-3TC arm and one sequence for 1 patient in the 3TC arm were outliers in the phylogenetic trees. These patients were excluded from the analyses so that the evolutionary drift from baseline could be measured for the remaining patients without the interference of the outliers. The sequences for these patients have been marked in the phylogenetic trees shown in Figure 2 and Figure 3. After these exclusions we had 63 sequences for 18 patients (3TC) and 54 sequences for 19 patients (No-3TC) to include in our analysis.

The Jukes-Cantor distances show an increase in number of nucleotide changes from baseline to week 48, comparable between the two arms at all time points (Figure 1);

At Week 48 the median (IQR) nucleotide distances from baseline were:

2.01 (0.00, 17.41) (n=9) vs. 4.05 (1.01, 11.22) (n=7)

in the 3TC arm vs. the No-3TC arm, respectively (p=0.49).

Similar results were found using the:

Kimura 2-parameter distance (in 10^{-3}):

1.54 (0.00, 13.39) vs. 3.11 (0.78, 8.63); p=0.49 (Figure 4)

and the

Syn-Scan synonymous/non-synonymous method:

9.0 (5.0, 28.4) vs. 12.0 (8.5, 20.0); p=0.71. (Figure 5)

The observed increase in mutation rate for the No-3TC arm, at week 48 specifically, in the Jukes-Cantor and the Kimura 2-parameter methods cannot be confirmed when adjusting for the ambiguous nucleotide letters as we did in the Syn-Scan method (Figure 5). The Syn-Scan method shows a constant difference between the two arms of approximately 3 nucleotides at week 24, 36 and 48. It is important to note that the loss of the M184V mutation in No-3TC arm is expected to show a distance of one nucleotide change from baseline.

LIMITATIONS

Due to the small numbers of patients who could be included in these analyses there may not be sufficient power to identify any differences between evolutionary distances if they did exist.

This is a subgroup analysis of COLATE patients who failed to suppress viral replication therefore, even though these data are derived from a RCT, the randomisation procedure may not have ensured a comparable prior risk of viral evolution between treatment arms. However, the rate of suppression was comparable in the two arms in the parent trial implying a reduction in the possible bias.

CONCLUSION

In this substudy to the COLATE trial comparable mutation rates were found over 48 weeks among patients failing a 3TC-containing regimen who either continued or discontinued 3TC as part of HAART. Data at week 48 in two of the three analyses suggest there may be a trend in a higher mutation rate in the No-3TC arm. Larger datasets and a longer follow-up time is needed to see if this is the case.

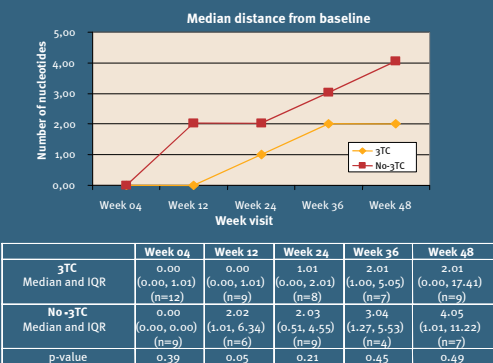### Figure 1 — Jukes-Cantor distance from baseline

Median distance from baseline

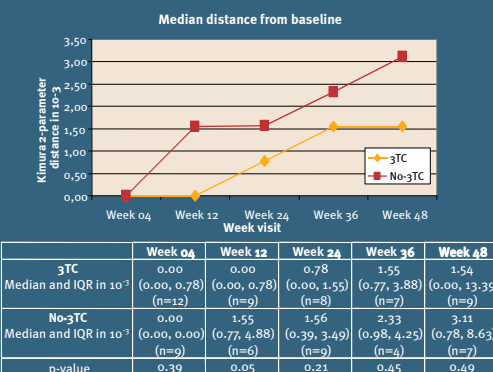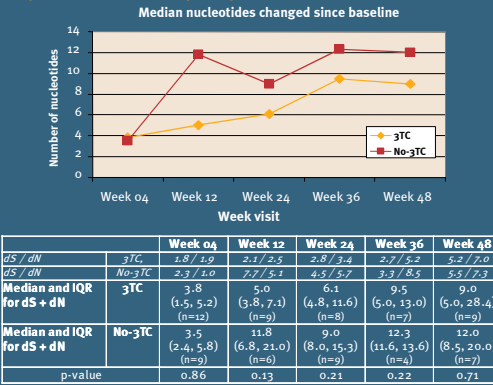|	Week 04	Week 12	Week 24	Week 36	Week 48
3TC Median and IQR	0.00 (0.00, 1.01) (n=12)	0.00 (0.00, 1.01) (n=9)	1.01 (0.00, 2.01) (n=8)	2.01 (1.00, 5.05) (n=7)	2.01 (0.00, 17.41) (n=9)
No-3TC Median and IQR	0.00 (0.00, 0.00) (n=9)	2.02 (n=6)	2.03 (n=9)	3.04 (1.27, 5.53) (n=4)	4.05 (1.01, 11.22) (n=7)
p-value	0.39	0.05	0.21	0.45	0.49

Values listed are the products of:
The Jukes Cantor distance * sequence length (1302 nucleotides) = number of nucleotides.

Figure 4 — Kimura 2-parameter distance from baseline

Median distance from baseline

	Week 04	Week 12	Week 24	Week 36	Week 48
3TC Median and IQR in 10^{-3}	0.00 (0.00, 0.78) (n=12)	0.00 (0.00, 0.78) (n=9)	0.78 (0.00, 1.55) (n=8)	1.55 (0.77, 3.88) (n=7)	1.54 (0.00, 13.39) (n=9)
No-3TC Median and IQR in 10^{-3}	0.00 (0.00, 0.00) (n=9)	1.55 (0.77, 4.88) (n=6)	1.56 (0.39, 3.49) (n=9)	2.33 (0.98, 4.25) (n=4)	3.11 (0.78, 8.63) (n=7)
p-value	0.39	0.05	0.21	0.45	0.49

Figure 5 — Synonymous (dS)/non-synonymous (dN) changes from baseline

Median nucleotides changed since baseline

		Week 04	Week 12	Week 24	Week 36	Week 48
dS / dN	3TC	1.8 / 1.9	2.1 / 2.5	2.8 / 3.4	2.7 / 5.2	5.2 / 7.0
dS / dN	No-3TC	2.3 / 2.0	2.7 / 5.1	4.5 / 5.7	3.3 / 8.5	5.5 / 7.3
Median and IQR for dS + dN	3TC	3.8 (1.5, 5.2) (n=12)	5.0 (3.8, 7.1) (n=9)	6.1 (4.8, 11.6) (n=8)	9.5 (5.0, 13.0) (n=7)	9.0 (5.0, 28.4) (n=9)
Median and IQR for dS + dN	No-3TC	3.5 (2.4, 5.8) (n=9)	11.8 (6.8, 21.0) (n=6)	9.0 (8.0, 15.3) (n=9)	12.3 (11.6, 13.6) (n=4)	12.0 (8.5, 20.0) (n=7)
p-value		0.86	0.13	0.21	0.22	0.71

Neighbour-Joining distance tree for all sequence in the 3TC arm (Figure 2) and the no-3TC arm (Figure 3):

- Leaves marked with a shape have been excluded from the analysis.
- First number in the leave name indicates the presence of the M184V mutation (1 = present, 0 = not present)
- The number between the two hyphens is the patient identifier
- The last two-digit number is the week visit to which the sequence belongs:
  01 = Baseline; 04 = Week 04; 12 = Week 12; 24 = Week 24; 36 = Week 36; 48 = Week 48
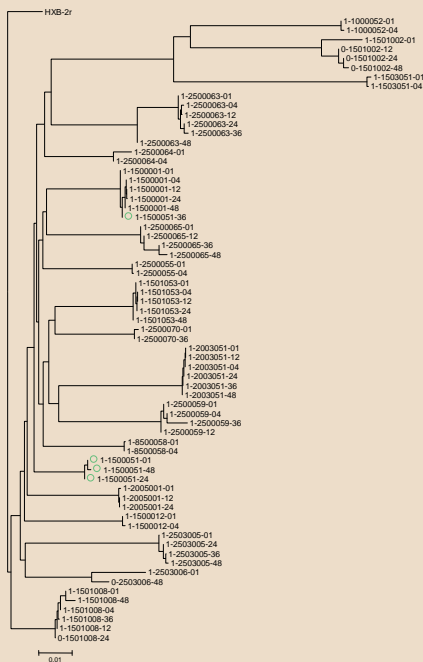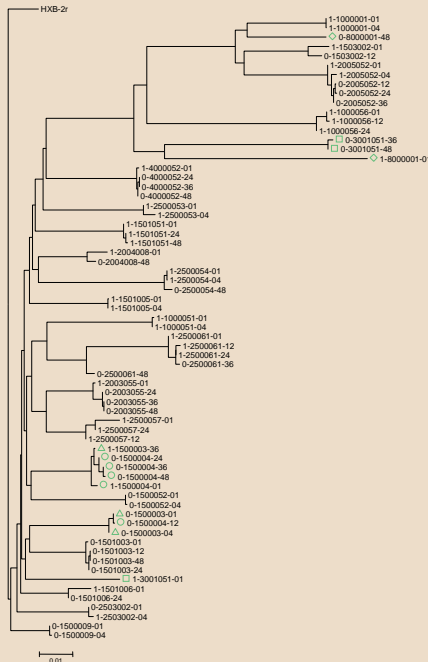

Figure 2: 3TC arm

Figure 3: No-3TC arm

*Transition/Transversion

- A transition is a mutation either between:
  two purines (A ↔ G) or two pyrimedines (C ↔ T)
- A transversion is a mutation between:
  a purine and a pyrimedine (A or G ↔ C or T)

Transitions happen more often than transversions. The Kimura 2-parameter model takes the transition/transversion ratio into account.

#Synonymous (dS)/Non-synonymous (dN)

- A synonymous mutation is a nucleotide change that does not result in change of an amino acid, a synonymous mutation is always in the third position of a codon.
- A non-synonymous mutation is a nucleotide change that does result in change of an amino acid, a change at the first or second position of a codon will always be a non-synonymous mutation.

The synonymous/non-synonymous ratio gives an indication of positive or purifying selection. A ratio less than 1 indicates positive selection.

@Ambiguous nucleotide letters (IUPAC codes)

- Letters that code for a mixture of A, C, T, G
  - E.g. R codes for purines (A and G) and Y for pyrimidines (C and T).
- Common in HIV-1 sequence data
  - HIV-1 is a quasispecies and the letters reflect the diversity rather than the ambiguity
- Letters are treated very differently in sequence programmes
  - PAUP applies the shortest possible distance:
    - Distance between R and A is zero even though R codes for a mixture of A and G
  - Syn -Scan (http://hivdb.stanford.edu/pages/synscan.html): calculates the average distance:
    - Distance between R and A is 0.5 for a mixture of A and G

PRINCIPAL INVESTIGATORS

Belgium:	N. Clumeck
Denmark:	J. Gerstoft, L. Mathiesen, C. Pedersen
France:	P. Dellamonica , C. Katlama
Germany:	F.-D. Goebel, S. Staszewski
Greece:	G. Panos
Ireland:	F. Mulcahy
Italy:	A. d'Arminio Monforte
Norway:	J. N. Bruun
Poland:	A. Horban
Spain:	J. M. Gatell, B. Clotet
Sweden:	K. Gyllensten
United Kingdom:	
	D. Churchill, M. Fischer, B. Gazzard , M. Johnson, C. Leen

CHIP COPENHAGEN HIV PROGRAMME