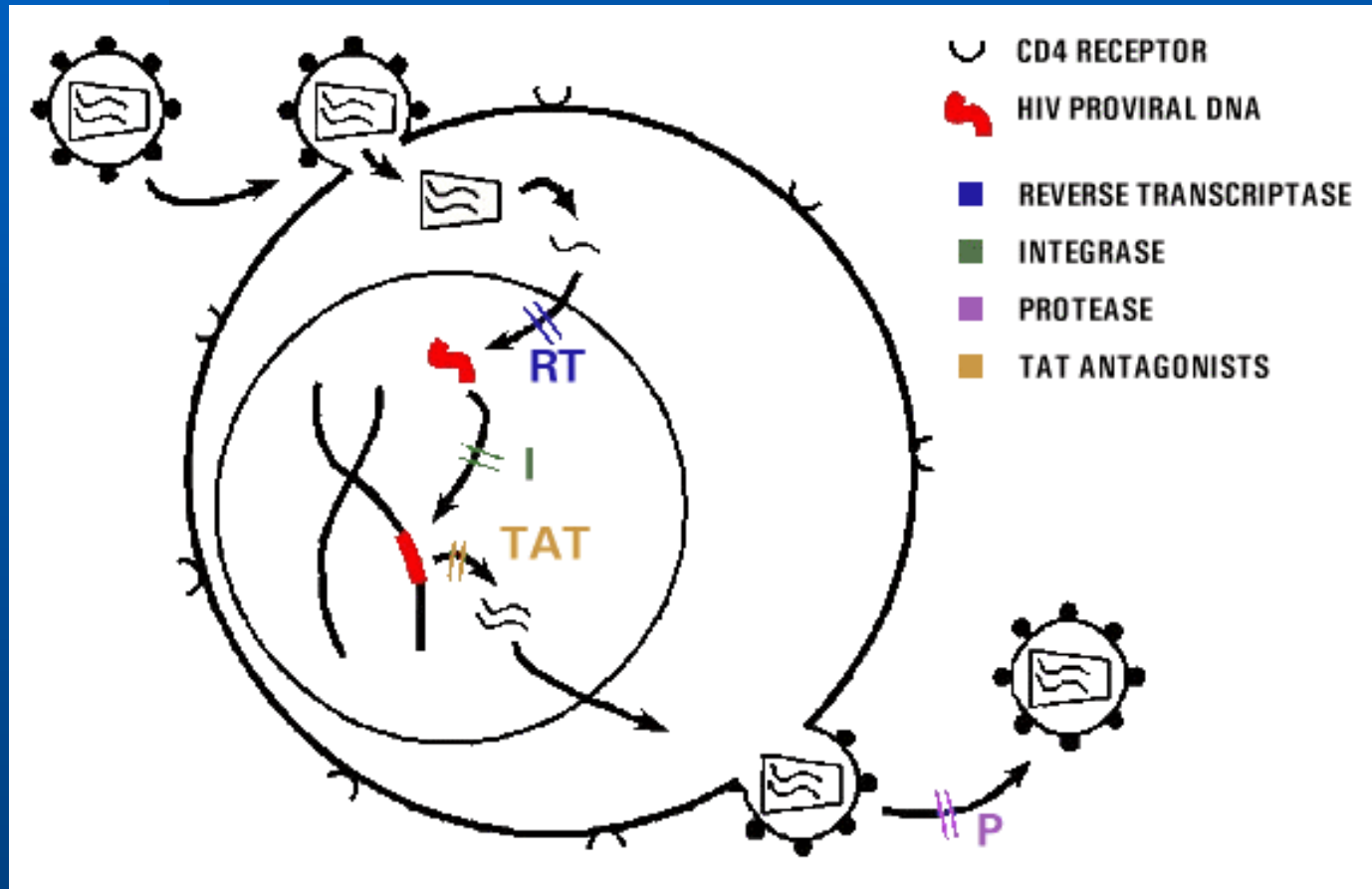# Analysis of HIV Sequences From Patients Data

By Patricia Buendia

Why study HIV data:

- **Viruses have a high evolution rate**
- **Large amount of data in public databases**
- **Evolutionary pattern closely related to the immunological status of host**
- **Analysis of serially sampled data significant to medical assessment of disease**

Drugs and the HIV Life Cycle
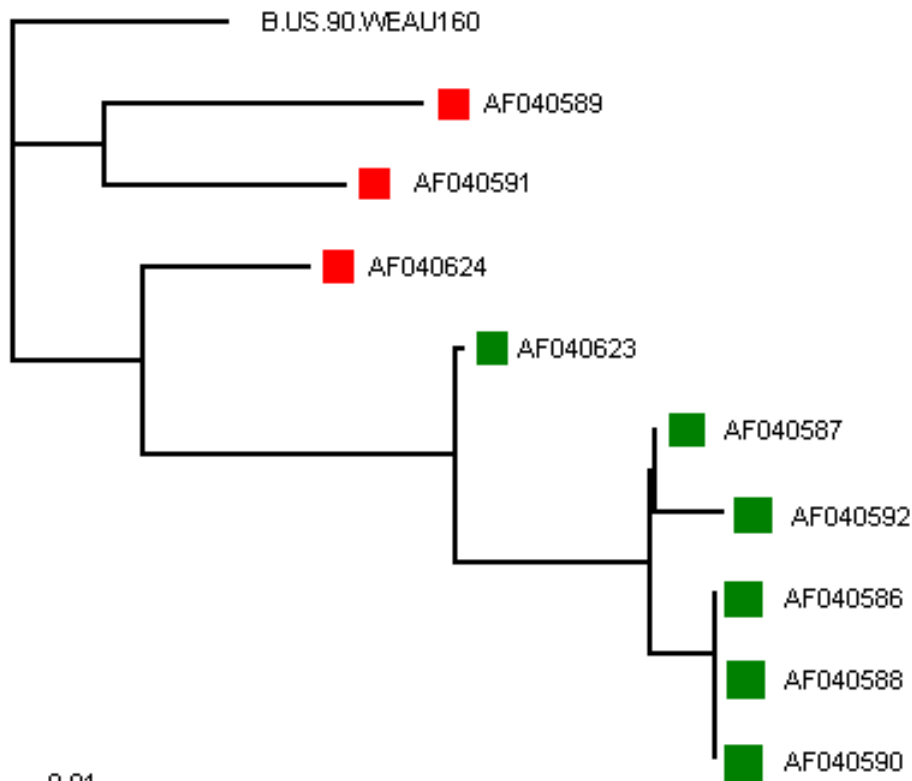
Analysis of First Set of Data

- **1 Year study of 10 patients before and during therapy**

- **Drug therapy: AZT(ZDV)-3TC-IDV**

- **5 to 9 sequences per patient in Genbank**

Reference

Günthard et al. – Journal of Virology 1998

Human immunodeficiency virus replication and genotypic resistance in blood and lymph nodes after a year of potent antiretroviral therapy

ML Tree for Patient F

- Generated by Phylip's DNAML program
- Implements maximum likelihood method

geno2pheno

- **Input: HIV-1 pol-gene DNA sequence**
- **Output: predictions of phenotypic resistance to 17 antiretroviral drugs**
- **How: 2-class phenotype predictions using decision trees and SVM**

FOR MORE INFO...

http://217.89.67.10/cgi-bin/geno2pheno.pl/

Patient F: Drug Resistance Prediction Results

Sample AF040589 before receiving ZDV-3TC- IDV drug therapy

Sample AF040592 after a year of ZDV-3TC-IDV drug therapy (March 1998)

Drug	Cutoff	Decision tree classification[1] [confidence factor]	SVM classification[2]	Predicted fold-resistance (SVM regression[2])
ZDV	8.5	resistant [0.90]	resistant	13.6
ddC	2.5	susceptible [0.80]	susceptible	1.5
ddI	2.5	susceptible [0.86]	susceptible	1.6
d4T	2.5	susceptible [0.71]	susceptible	1.8
3TC	8.5	susceptible [0.80]	susceptible	5.8
ABC	2.5	resistant [0.89]	susceptible	2.2
TDF	2.5	susceptible [0.88]	susceptible	2.2
NVP	8.5	susceptible [0.89]	susceptible	1.5
DLV	8.5	susceptible [0.89]	susceptible	1.5
EFV	8.5	susceptible [0.91]	susceptible	1.1
SQV	3.5	susceptible [0.89]	susceptible	1.1
IDV	3.5	susceptible [0.90]	susceptible	1.2
RTV	3.5	susceptible [0.91]	susceptible	1.1
NFV	3.5	susceptible [0.89]	susceptible	1.0
APV	3.5	susceptible [0.92]	susceptible	1.3
LPV	3.5	susceptible [0.86]	susceptible	1.1
ATV	3.5	susceptible [0.84]	susceptible	1.4

Drug	Cutoff	Decision tree classification[1] [confidence factor]	SVM classification[2]	Predicted fold-resistance (SVM regression[2])
ZDV	8.5	resistant [0.90]	resistant	26.7
ddC	2.5	resistant [0.73]	resistant	2.6
ddI	2.5	resistant [0.58]	resistant	2.4
d4T	2.5	susceptible [0.71]	susceptible	1.9
3TC	8.5	resistant [0.98]	resistant	211.0
ABC	2.5	resistant [0.89]	resistant	4.0
TDF	2.5	resistant [0.76]	susceptible	1.8
NVP	8.5	susceptible [0.89]	susceptible	5.3
DLV	8.5	susceptible [0.89]	susceptible	1.5
EFV	8.5	susceptible [0.91]	susceptible	2.2
SQV	3.5	resistant [0.88]	susceptible	3.0
IDV	3.5	resistant [0.87]	susceptible	3.0
RTV	3.5	resistant [0.89]	susceptible	2.7
NFV	3.5	resistant [0.93]	susceptible	3.3
APV	3.5	susceptible [0.92]	susceptible	1.7
LPV	3.5	susceptible [0.86]	susceptible	1.5
ATV	3.5	resistant [0.83]	susceptible	2.6

Analysis of Second Set of Data

- **5 Year study of HAART therapy**
  - 2 patients under 3TC + d4T+ Idv drug therapy
  - 1 patient under 3TC + d4T + Nfv + Sqv
- **approx. 250 sequences per patient**
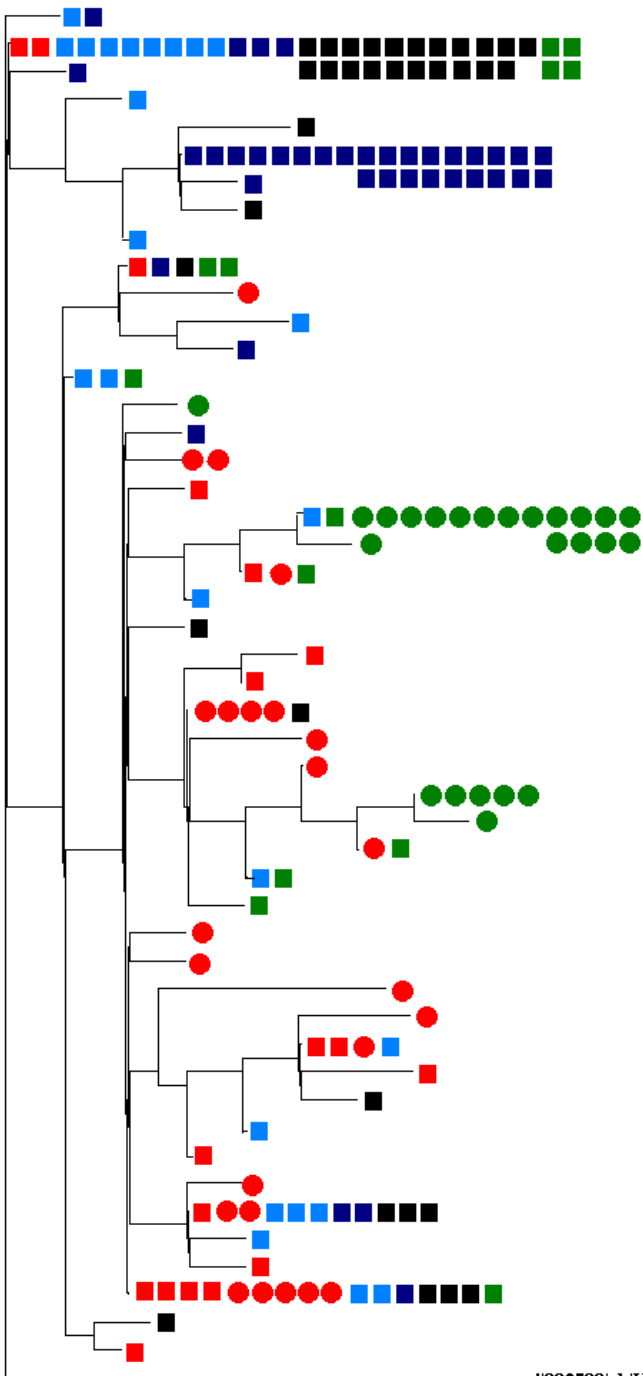- **Observation: Viral load increased after discontinuation of therapy**

Reference

Imamichi et al. – Journal of Infectious Diseases 2001

Human immunodeficiency virus Type 1 quasi species rebound after discontinuation of higly active antiretroviral therapy…

ML Protease Tree

Tree for Patient 2 was constructed using DNAML

	HIV-1 RNA in plasma	HIV-1 DNA in plasma
Baseline (-5 months and prior to start of HAART)	● (red circle)	■ (red square)
0 mo - 1st documented HIV RNA <50 copies/mL		■ (light blue square)
12 mo after 1st <50 copies/m_		■ (navy square)
18 mo after 1st <50copies/mL		■ (black square)
29 mo - Time point of relapse after discontinuation of HAART	● (green circle)	■ (green square)

gi|902798|gb|U26942.1|HIV1U269

Time 1: -5 months before first count of <50 copies/mL &before therapy start

2 3 4 6 23 26 27 30 31 34 35 37 38 39

1 5 22 24 25 28 29 36 40 44

Time 2: 0 months and first documented count of <50 copies/mL

19 9 41 24 8 13 48 7 28 20 36 40 10 11 44

Time 3: 12 months after first count of <50 copies/mL

12 22 13 28 36 43 42 49 14 44

Time 4: 18 months after first count of <50 copies/mL

15 22 28 16 17 29 36 18 45 44

Time 5: 29 months after first count of < 50copies/mL 3 weeks after discontinuation of therapy & relapse

19 22 25 20 21 33 48 31 32 36 46 47 44

Find Closest Ancestor

- **Use a distance matrix**
- **Find closest distance between sequences from consecutive time periods**

	1	2	3	4	5	6	7	8	9	10	11
1											
2	0.0067										
3	0.0101	0.0101									
4	0.0135	0.0135	0.0101								
5	0.0101	0.0101	0.0067	0.0101							
6	0.0101	0.0034	0.0135	0.0168	0.0135						
7	0.0135	0.0135	0.0101	0.0067	0.0101	0.0168					
8	0.0101	0.0101	0.0067	0.0101	0.0067	0.0135	0.0101				
9	0.0135	0.0135	0.0101	0.0135	0.0034	0.0168	0.0135	0.0101			
10	0.0135	0.0202	0.0168	0.0202	0.0168	0.0236	0.0202	0.0168	0.0135		
11	0.0135	0.0202	0.0168	0.0202	0.0168	0.0236	0.0202	0.0168	0.0135	0.0067	
12	0.0135	0.0202	0.0168	0.0202	0.0101	0.0236	0.0202	0.0168	0.0135	0.0202	0.0202
.	.	.	.	.	.	.	.	.	.	.	.

Neutral and Darwinian Selective Evolution

Sequence 1:	ATC	GTA	CCT
Sequence 2:	ATC	GTT	ACT
	-	SYN	NONSYN
	Ile	Val	Pro->Thr

- **Detect positive selection by comparing rates of nonsynonymous to synonymous substitutions $\omega = d_N/d_S$**

- **assume the nonsynonymous substitutions are proof of positive selection $\omega > 1$**

- **pairwise calculation of positive selection rate ω is done using Yang's codon-based model implemented in PAML**

Algorithm for Analysis of serially sampled data

1) **Calculate distance matrix for unique sequences (DNADIST)**

2) **Calculate pairwise positive selection rate ω (PAML)**

3) **Separate sequences into time groups $t_i, \ldots, t_n$**

4) **For each $seq_d \in t_i$ find $seq_a \in t_{i-1}$ so that $dist(seq_d, seq_a) = \min(dist(seq_d, seq_j)) \; \forall \; seq_j \in t_{i-1}$**

5) **For each such pair found in 4), get ω from matrix in 2), to determine positive selection ($\omega > 1$)**

Time 1: -5 months before first count of <50 copies/mL &before therapy start

2 3 4 6 23 26 27 30 31 34 35 37 38 39

1 5 22 24 25 28 29 36 40 44

20%

Time 2: 0 months and first documented count of <50 copies/mL

19 9 41 24 8 13 48 7 28 20 36 40 10 11 44

30%

Time 3: 12 months after first count of <50 copies/mL
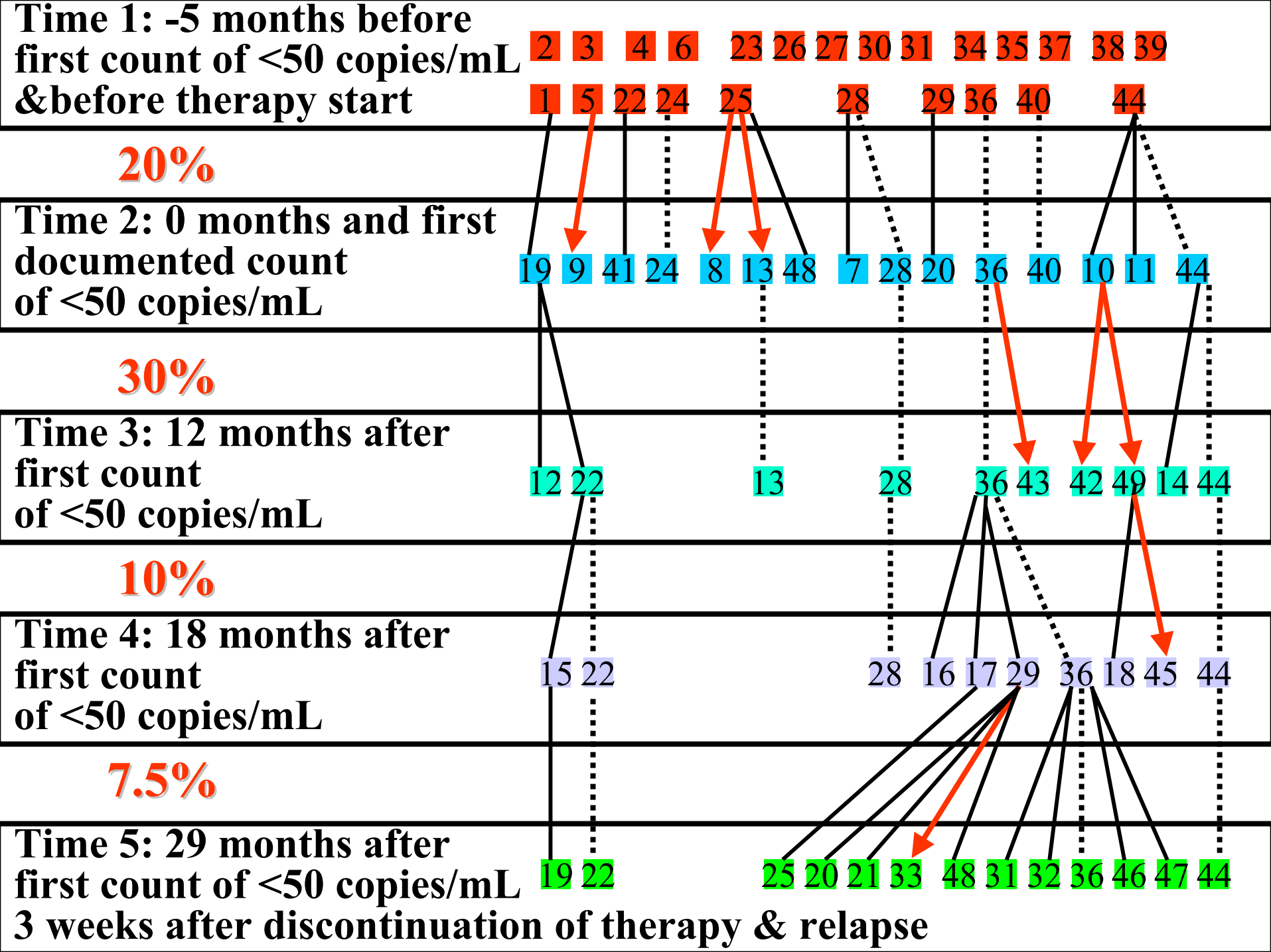
12 22 13 28 36 43 42 49 14 44

10%

Time 4: 18 months after first count of <50 copies/mL

15 22 28 16 17 29 36 18 45 44

7.5%

Time 5: 29 months after first count of <50 copies/mL 3 weeks after discontinuation of therapy & relapse

19 22 25 20 21 33 48 31 32 36 46 47 44

Conclusion

- **Reasons for drug therapy failure:**
  - **appearance of multidrug-resistant virus:**
  - **non-drug resistant virus continues to replicate in virus reservoirs of special cells**
- **New methods needed for analysis of time consecutive HIV-1 data**