Evidence for more major HIV-1M lineages from the early stages of the HIV-1 epidemic

Marcel Tongo[1,2,3], Wendy A Burgers[4], Eitel Mpoudi-Ngole[2] Jeffrey R Dorfman[1,3]and Darren P Martin[5]

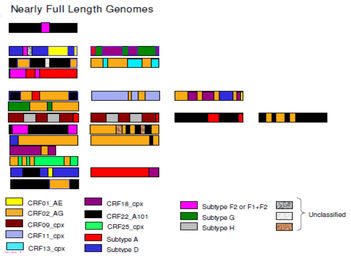
[1]International Centre for Genetic Engineering and Biotechnology, Cape Town, South Africa

[2]Institute of Medical Research and Study of Medicinal Plants (IMPM), Yaoundé, Cameroon

[3]Division of Immunology, [4]Division of Medical Virology and [5]Computational Biology Group, University of Cape Town, South Africa

Introduction

The current HIV-1M subtype and CRF classification system is largely based on the order in which different HIV-1M lineages were discovered with viruses discovered early tending to be classified as belonging to "pure" subtypes. Viruses discovered later that had inconsistent degrees of similarity across their genomes to these pure subtypes have subsequently been classified as "recombinant forms". Although this approach has worked very well for classifying HIV variants in areas where few subtypes predominate, it might not be optimal for classifying the multitudes of highly divergent viral lineages that still circulate in the Congo basin region. These lineages are commonly too divergent to be placed within any existing subtype or CRF grouping. These lineages and the presence of sequences of indeterminate origin within established CRFs, provide strong evidence that HIV diversity is not fully represented under the current classification system. Without invoking the existence of numerous undiscovered divergent HIV-1 group M lineages, it is also difficult to explain how viruses belonging to apparently low-prevalence parental subtypes (e.g. F, J, H) have contributed to so many of the currently described HIV-1 group M recombinant forms.



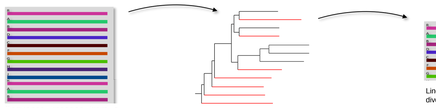
Carr et al, *Retrovirology* 2010

In an effort to better characterise HIV-1M diversity, we re-analysed the phylogenetic and recombinational relationships of a set of viruses fully representing the known diversity of HIV-1M .

Methods

Sequences:

We selected a set of 481 sequences representative HIV-1M near full length sequences including (i) 423 sequences from 76 characterised subtypes/CRFs, (ii) 12 sequences classified as "U", (iii) 37 divergent sequences from Genbank from the Congo basin and (iv) an additional nine near full-length divergent sequences from Cameroon (Tongo et al, 2012) .



The representative sequences were specifically selected to include the broadest diversity of sequences previously identified as belonging to known HIV-1 group M subtypes and CRFs

Analyses

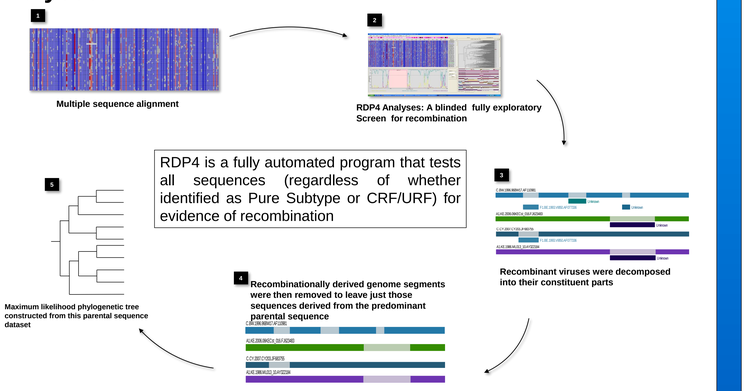
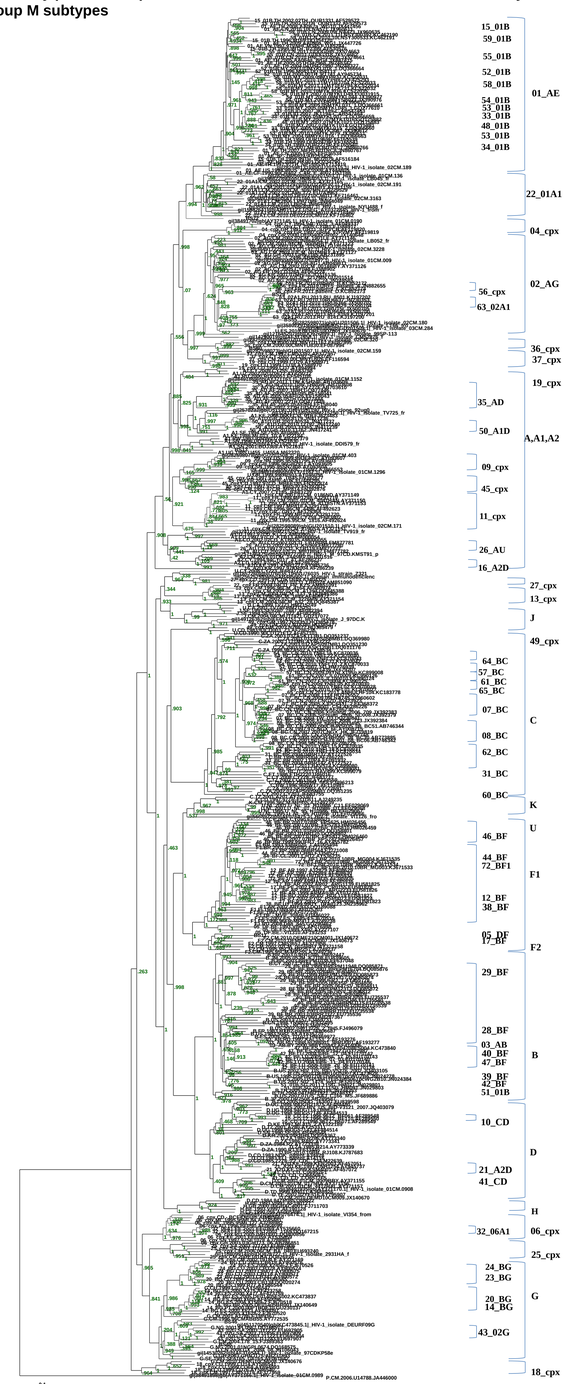


Figure 1: Maximum likehood tree indicating the phylogenetic relationships between 471 full length sequences of HIV-1. The tree was constructed from these sequences with 100 boostrap replicates. The tree was rooted using a sequence from HIV-1 group P. Green values indicate the midpoint support.

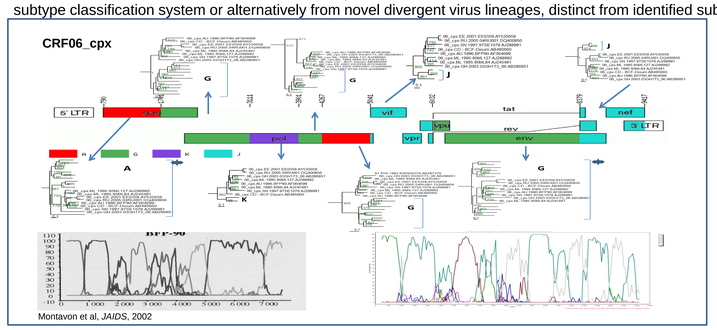
Results

Phylogenetic analyses of mostly recombination-free HIV-1 group M sequences indicated that many parental sequences of CRFs are not classifiable into the currently defined HIV-1 group M subtypes

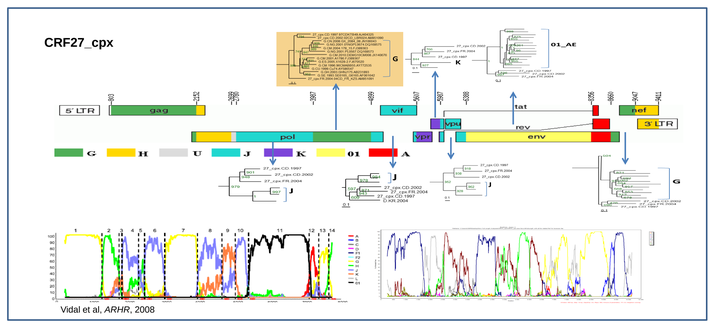


Novel divergent virus lineages

A reconstruction of some CRFs that are presently described in the literature was made in order to reconcile and/or understand the analytical issues causing differences between the established genome structures of known recombinants and those described during our exploratory recombination screen. ML methods were used to test whether segments of recombinationally derived sequence were from parental viruses falling within the existing subtype classification system or alternatively from novel divergent virus lineages, distinct from identified subtypes.



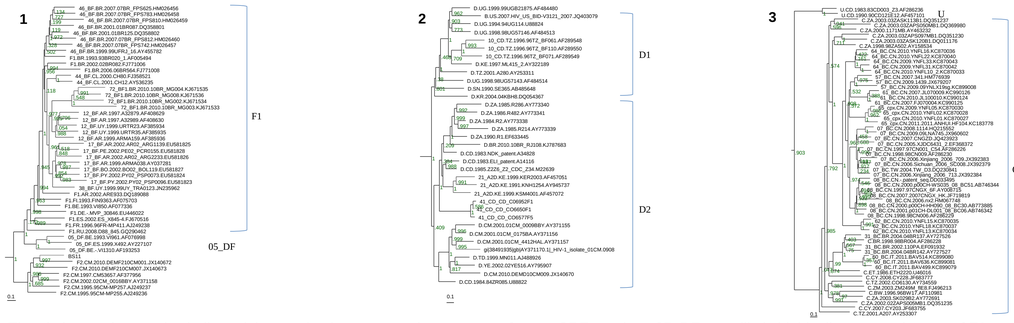
Montavon et al, *JAIDS*, 2002



Vidal et al, *ARHR*, 2008

- **Most of the recombinationally derived genome segments are apparently from parental viruses that fall within novel divergent lineages. (See displayed examples, CRF06_cpx and CRF27_cpx.)**
- **Interestingly, one of the subtype G-assigned sequences within 27_cpx appears, in fact, to have been derived from an ancestral subtype G from a virus in the 27_cpx lineage (see orange-shaded subtree).**

Highly divergent subgroup of HIV-1



- **The phylogenetic tree also indicated that predominant parental sequences of some CRFs such as 05_DF were likely highly divergent subgroup of F1 viruses and may constitute a sub subgroup of F1 (subtree 1).**
- **Some major subtypes like subtype D exhibited underappreciated diversity and can be divided in two sub-subtypes (subtree 2).**
- **Our analysis also demonstrated that some sequences previously identified as U represent viruses that are genuinely highly divergent and are possibly extant descendants of previously unknown early diverging HIV-1M lineages (subtree 3).**

Conclusions

Our analyses implies that there is potentially a far more diverse pool of HIV-1M sequences circulating amongst humans than the current classification system of subtypes and CRFs might suggest.

Our fully exploratory recombination screen suggested that many CRFs contain sequence fragments from previously uncharacterised diverging HIV-1group M lineages.

Such sequences are critical to efforts to piece together the early evolutionary history of HIV-1 group M.

Acknowledgements: **Funding for this study was provided by the Atomic Energy International Agency (IAEA); WHO-Africa Vaccine Programme (AAVP); International Centre for Genetic Engineering and Biotechnology (ICGEB), Cape Town**