The Diversification of HIV-1: a comparison of groups O and M

John Archer and David L. Robertson
Michael Smith Building, Faculty of Life Science, University Of Manchester
Contact: john.archer@postgrad.manchester.ac.uk

Abstract

Objective: To quantify the similarity (or lack of) between the phylogenetic substructure of HIV-1 in groups O and M.

Methods: Two phylogenetic tree statistics the subtype diversity ratio*, SDR, and the subtype diversity variance** - this is a measure of the variation within the ratio of the mean intra-cluster pairwise distance to the mean inter-cluster pairwise distance calculated for each cluster on the tree. The lower the value the more symmetrical the tree.

Results: We show that as expected the established global-group M subtypes have a high degree of phylogenetic symmetry in relation to each other in terms of inter- and intra-subtype diversification. They are significantly different from the substructure present amongst the random trees. To the contrary the group O diversification does not display this highly symmetrical substructure and is not significantly different from the substructure present on randomly generated trees. Phylogenies comprised of group M strains from the epicentre of the HIV/AIDS pandemic, the Democratic Republic of Congo (DRC), exhibit a substructure more similar to group O than to global-group M.

Conclusions: The substructure present within groups O and M is quantifiably different. The well-defined clades that characterize the group M diversification, are not present in group O or amongst group M strains from the DRC. The group M subtypes are thus unique and a signature of pandemic HIV-1.

Introduction

Three different cross species transmission events involving Simian Immunodeficiency Virus (SIVm) from apes has resulted in three distinct phylogenetic lineages of HIV-1 in humans. These lineages termed “groups” are labeled M (main), O (Outlier) and N (non-M/O) [1] (Fig. 1). Group M is almost entirely responsible for the global HIV pandemic. Group O has remained endemic to Cameroon and is also probably the result of a single cross species transmission event involving P. t. troglodytes. The HIV-1 strains that comprise group M cluster into distinct and strongly supported clades in phylogenetic trees (Fig. 1A). These clades are referred to as “subtypes” and have been labeled A to D, F to H, J and K [1]. The significance of the HIV-1 subtypes, and whether or not group O can be subdivided into clusters that are biologically equivalent to group M subtypes, is unclear [2, 3]. Resolution of this issue will have implications for any future group O vaccine design and possibly for drug based intervention strategies.

Materials and Methods

To compare group O and global group M phylogenies neighboring joining trees were constructed from p24 (gag), p32 (pol) and gp160 (env) genomic regions.

Figure 1. Phylogenetic tree showing the relationship between the three HIV-1 groups.

Figure 2. The relationship between SDR and tree topology.

Materials and Methods

To compare group O and global group M phylogenies neighboring joining trees were constructed from p24 (gag), p32 (pol) and gp160 (env) genomic regions. SDR [4] values were calculated for the trees and compared between both groups and to the SDR distribution produced by randomly generated trees. SDV values were also calculated. To calculate the SDR and SDV values each tree was first divided into clusters. For group M the designated subtypes within the LANL HIV sequence database were used, for group O previously proposed clades were used, while for the random trees a heuristic clustering algorithm was used.

To investigate the emergence of the “global” group M subtypes and their relevance within the center of the pandemic a neighbor joining tree, containing 197 partial env sequences sampled in 1997 [8] and 46 sequences sampled in the 1980s [8], was constructed. Both these datasets consisted of sequences from the DRC region. SDR and SDV values for this tree where calculated.

Following this in order to get a fuller picture of the diversity within the group M subtypes, 258 more sequences sampled during 2002 [8] were added to the DRC tree along with two other strains that were previously proposed to be a new subtype [8]. 100 global group M strains from the first part of the analysis were truncated and also added to the tree.

Figure 3. HIV-1 groups M and O phylogenetic trees. Panel (a) displays a tree that was constructed using global-group M envelope gp160 sequences. In panel (b) the tree was constructed using group O gp160 sequences. The “*” indicates bootstrap support greater than 90%. The scale bar corresponds to nucleotide substitutions per site. The mean SDV value for the group M gp160 trees was 0.017 (±0.002) while for the group O gp160 tree it was 0.03.

Discussion

Distinct group M “subtypes” have emerged during the course of a complicated epidemiological history arising from the HIV-1 pandemic (Fig. 6). The non-random nature of their emergence is quantified by the significant difference obtained between the group M SDR distribution and the SDR distribution produced by random tree topologies (Fig. 4).

The subtype classification system has been devised through the global sampling of strains. Strains from the centre of the pandemic (DRC) are classified according to the subtype that they fall closest to. However the distinctness of these globally defined subtypes is due to individual strains being exported from the DRC region followed by localised diversification. Thus the current subtypes do not accurately represent the diversity that is present within the centre of the pandemic.

In relation to group O the absence of strong founder effects followed by relatively isolated diversification has reduced the extent of strain clustering (Fig. 3b). As a result the groups SDR values fall significantly inside the values produced by the clusters present on the randomly generated trees (Fig. 4). Group O clusters are not equivalent in nature to the global group M subtypes although they are comparable to the loosely defined lineages present at the centre of the group M pandemic.

Acknowledgments

We are grateful to Andy Rambaut for his insightful comments and his advice concerning the generation of random trees. We would also like to thank Michael Worobey and John Pinney for helpful comments and discussion. This research is supported by a grant from the BBSRC.

References

[3] We are grateful to Andy Rambaut for his insightful comments and his advice concerning the generation of random trees. We would also like to thank Michael Worobey and John Pinney for helpful comments and discussion. This research is supported by a grant from the BBSRC.

Contact: john.archer@postgrad.manchester.ac.uk

The University of Manchester