Hybrid Origin of SIV in Chimpanzees

Elizabeth Bailes,[1] Feng Gao,[2,3] Frederic Bibollet-Ruche,[2] Valerie Courgnaud,[4] Martine Peeters,[4] Preston A. Marx,[5] Beatrice H. Hahn,[2] Paul M. Sharp[1]*

The ancestry of HIV-1 (human immunodeficiency virus–1) has been traced to SIVcpz (simian immunodeficiency virus) infecting chimpanzees (*Pan troglodytes*)] in west central Africa (*1*), but the origin of SIVcpz itself remains unknown. Species- specific strains of SIV have been identified in more than 20 species of African primates (*2*), but all, except SIVcpz, infect monkeys. Here, we present evidence that SIVcpz arose through successive cross-species transmission and recombination events of SIVs infecting monkeys on which chimpanzees prey.

The known strains of SIV cluster into six distinct major lineages (*2*), whose order of divergence has been difficult to disentangle. SIVcpz forms a deep branch within the primate lentivirus tree and could reflect an ancient infection. However, SIVcpz has only been found in central (*P. t. troglodytes*) and eastern (*P. t. schweinfurthii*) subspecies (*1–3*), whereas members of a third, western subspecies (*P. t. verus*) appear not to be infected. Thus, chimpanzees more likely acquired SIV recently, subsequent to the split between subspecies.

SIV phylogeny is complicated by a number of viruses that occupy different positions in the evolutionary tree dependent on the region of genome analyzed, indicative of recombination events among viruses from different major lineages. Two recently characterized viruses, SIVrcm (*4*), from red-capped mangabeys (*Cercocebus torquatus*), and SIVgsn (*5*), from greater spot-nosed monkeys (*Cercopithecus nictitans*), are most closely related to SIVcpz, but only in certain genomic regions, and so both have been interpreted as recombinant viruses (*4*, *5*). However, these regions do not seem to overlap, and it is possible that SIVcpz, rather than SIVrcm or SIVgsn, is the recombinant.

To investigate this, we have conducted extensive phylogenetic analyses on different subsets of SIV strains, comparing the topologies among four regions of the proteome: Gag, Pol(PR-RT), Pol(IN), and Env. Strong discordance among the topologies from different regions would be evidence of recombination. Representatives of eight clades of SIV for which full-length sequences are available were analyzed: the six previously defined, approximately equidistant lineages (*2*, *6*) and the two putatively recombinant lineages (SIVrcm and SIVgsn). We examined all 70 possible combinations of four lineages ("tetrads") selected from these eight, and we compared the likelihoods of the three possible (unrooted) topologies, asking which was the best topology and whether the other topologies could be rejected statistically. By restricting the individual analyses to four lineages at one time, we aimed to identify at least a subset of nonrecombinant lineages that did not produce conflicting topologies for different regions.

Only 11 of the 70 tetrads yielded the same maximum likelihood topology for all four proteome regions (table S1), indicative of the difficulty in resolving the basal divergences among SIV lineages. None of the 35 tetrads including SIVcpz were among these 11. However, in many cases, the likelihood for the best topology was not significantly greater than that of an alternative. SIVcpz was the only lineage that gave rise to significant discordance among tree topologies for different regions in more than half of the tetrads in which it was included. Exclusion of SIVcpz yielded evidence of significant discordance in only 3 of the 35 different tetrads. In contrast, with exclusion of either SIVgsn or SIVrcm, 15 or 16 discordant tetrads remained. Thus, among the eight lineages analyzed, SIVcpz was the most clearly recombinant.

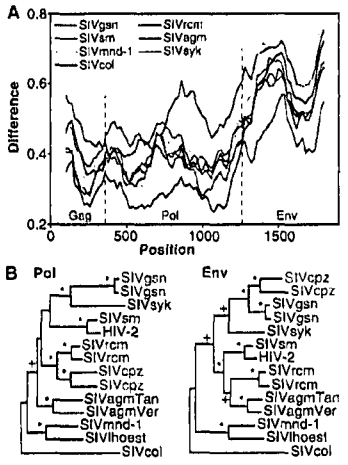

Fig. 1. (A) Diversity plot comparing SIVcpz (strain US) to seven other SIV lineages, across concatenated Gag, Pol, and Env protein sequence alignments. (B) Maximum likelihood phylogenies of primate lentivirus Pol and Env sequences. Internal branches found in at least 70% and 95% of bootstrap replicates are indicated by + and *, respectively (*6*).

Diversity plots across concatenated Gag, Pol, and Env sequences identified one clear breakpoint in SIVcpz between Pol and Env (Fig. 1A). In phylogenetic analyses, SIVcpz clustered closely with SIVrcm in Pol, but closely with SIVgsn in Env, whereas other features of the topologies did not differ significantly (Fig. 1B). This is consistent with a more recent origin of SIVcpz by recombination between ancestors of SIVs infecting red-capped mangabeys and greater spot-nosed monkeys, the ranges of which overlap with *P. t. troglodytes* in west central Africa. Because chimpanzees are known to hunt smaller monkey species, the simplest explanation appears to be that both SIVrcm and SIVgsn have been acquired by chimpanzees and recombined in that host. The other SIV-infected subspecies, *P. t. schweinfurthii*, is thought to have arisen recently due to eastward expansion from *P. t. troglodytes* (*7*); whether they were already infected by SIVcpz, or whether the virus followed later, is unclear.

The finding of a hybrid origin of SIV in chimpanzees has important implications. First, it provides evidence that, in addition to humans, another ape species acquired SIV by cross-species transmission under natural conditions. Second, the endemic infection of two chimpanzee subspecies (*3*) indicates substantial secondary spread of the initial hybrid. Third, the recombinant chimpanzee virus was capable of spreading to humans (*2*). It will be important to examine whether chimpanzee predation on monkeys has led to other SIV acquisitions and whether the resulting chimpanzee-adapted SIVs are more likely to infect humans.

References and Notes

1. F. Gao *et al.*, *Nature* **397**, 436 (1999).
2. B. H. Hahn, G. M. Shaw, K. M. De Cock, P. M. Sharp, *Science* **287**, 607 (2000).
3. M. Santiago *et al.*, *Science* **295**, 465 (2002).
4. B. E. Beer *et al.*, *J. Virol.* **75**, 12014 (2001).
5. V. Courgnaud *et al.*, *J. Virol.* **76**, 8298 (2002).
6. Materials and Methods are available as supporting online material on *Science* Online.
7. T. L. Goldberg, M. Ruvolo, *Mol. Biol. Evol.* **14**, 976 (1997).
8. This work was supported by NIH grants RO1 AI44596, R01 AI50529, and N01 AI85338.

Supporting Online Material
www.sciencemag.org/cgi/content/full/300/5626/1713/DC1
Materials and Methods
Table S1
References

[1]University of Nottingham, Nottingham NG7 2UH, UK. [2]University of Alabama at Birmingham, Birmingham, AL 35294, USA. [3]Duke University, Durham, NC 27710, USA. [4]University of Montpellier 1, 34032 Montpellier cedex 1, France. [5]Tulane National Primate Research Center, Covington, LA 70433, USA.

*To whom correspondence should be addressed. E-mail: paul@evol.nott.ac.uk