Discovery of a New Endogenous Type C Retrovirus (FcEV) in Cats: Evidence for RD-114 Being an FcEV<sup>Gag-Pol</sup>/Baboon Endogenous Virus BaEV<sup>Env</sup> Recombinant

ANTOINETTE C. VAN DER KUYL, 1* JOHN T. DEKKER, 1 AND JAAP GOUDSMIT 2

Department of Human Retrovirology, Academic Medical Centre, 1105 AZ Amsterdam, 1 and Amsterdam Institute of Viral Genomics, 1105 BA Amsterdam, The Netherlands 2

Received 29 March 1999/ Accepted 17 June 1999

Analysis of a cat genomic DNA library showed that cats harbor a previously unrecognized endogenous type C retrovirus, whose env gene has homology to the murine Fe-4 resistance gene. This unique retrovirus, designated FcEV (Felis catus endogenous retrovirus), has a type C pol gene, closely related to the primate Papio cynocephalus endogenous virus (PeV) pol, not overlapping the env gene, unlike in other type C retroviruses, and is presumably present in a higher copy number than RD-114. Phylogenetic analysis of FcEV and RD-114 fragments amplified from cat species and comparison with baboon endogenous virus (BaEV) fragments from monkeys suggested that RD-114 does not represent the cat strain of BaEV but is actually a new recombinant between FcEV type C genes and the env gene of BaEV. Although BaEV did appear to have infected an ancestor of the domestic cat lineage, it was a de novo recombinant that made its way into the cat germ line.

All 38 known cat species, including the big cats, arose during the last 10 to 15 million years (13), with the lineage leading to the domestic cat being formed 7 to 9 million years ago (11). Extant cats belonging to the species Felis probably have a common ancestor much more recently, since they are genetically closely related. An interesting feature of the genus Felis (composed of Felis catus, F. chaus, F. nigripes, F. margarita, F. silvestris, and F. libyca) is the presence of two endogenous retroviruses not found in any other cat species, e.g., endogenous feline leukemia virus (FeLV) and RD-114 (2, 20). RD-114 showed a high level of homology to the monkey virus baboon endogenous virus (BaEV), which was attributed to an ancient cross-species transmission. RD-114 can still be expressed, but there is probably only a single active copy on cat chromosome B3 (19). Analysis of RD-114 proviruses in the domestic cat genome showed that there are approximately 20 related integrations per cell but that most have either lost their env genes or replaced them by a completely unrelated env (20, 23). No complete sequence is available for RD-114; only the env gene and the 3′ end of the pol gene have been sequenced (7), and some long terminal repeat (LTR) sequences are available (22).

BaEV is a complete inducible endogenous retrovirus (1), which is found in the genomic DNA of a subset of African monkeys, e.g., baboons, geladas, mandrills, mangabeys, and African green monkeys (26). Phylogenetic analysis of BaEV sequences from monkeys indicated that two BaEV strains have entered independently into the germ line of the monkey species (26). The last common ancestor of the Cercopithecinae supposedly lived at least 9 million years ago, indicating that exogenous BaEV was spreading in Africa after that time. BaEV is a recombinant retrovirus, containing type C gag and pol genes and a type D env gene, which probably arose by recombination of two primate viruses (simian endogenous retrovirus [SERV] and Papio cynocephalus endogenous virus [PeEV]), whose genomes are also present in monkey genomic DNA (17, 27). RD-114 has the same mosaic genomic structure as BaEV. The env gene of RD-114 is most probably derived from BaEV and not from the primate type D virus SERV, one of the ancestors of BaEV. Although the three env genes are generally very homologous, BaEV and RD-114 have extensive homology at the extreme C-terminal part of this protein, and the 30 nucleotides nt downstream of the env stop codon are highly homologous, in contrast to SERV. Recombination probably took place just downstream of the env gene. The env-LTR intergenic region in type D viruses containing a constitutive transport element, which is absent from type C retroviruses.

To expand our knowledge of the evolution of RD-114 and BaEV, we have analyzed two env genes of presumed RD-114 proviruses isolated from a cat genomic library and have examined RD-114 sequence variation in different species of cats and compared it with BaEV variation in monkey species.

MATERIALS AND METHODS

Cat genomic library. A cat genomic library in the lambda FIX<sup>®</sup> II vector was obtained from Stratagene (La Jolla, Calif.). This library was constructed from whole blood of an adult female domestic cat of mixed breed.

Isolation of type C proviral clones. The cat genomic library was screened by using a <sup>32</sup>P-labelled probe homologous to the 3′ end of the RD-114 pol sequence. Lambda DNA was isolated from purified positive plaques by using the Wizard Lambda Preps DNA purification system from Promega (Madison, Wis.).

Cat and primate samples. DNA samples were obtained from F. catus (domestic cat), F. chaus (jungle cat), and F. silvestris (European wild cat). The F. chaus and F. silvestris DNA samples were kindly donated by Stephen J. O’Brien (National Cancer Institute, Frederick, Md.). Primate samples (peripheral blood mononuclear cells [PBMCs] and serum for Cercocebus aterrimus) were obtained from the following African monkeys: Papio ursinus (chacma baboon), Cercocebus aterrimus (black mangabey), Cercocebus aestuas aestuas (grivet), Cercocebus aestuas peyerdythrus (vervet), and Cercocebus aestuas sabaeus (green monkey). The origin of the samples was as published previously (26).

DNA extraction and amplification. Total DNA was extracted from the samples by a procedure involving silica and guanidine thiocyanate (5). A 590-bp nested fragment of BaEV or RD-114 was amplified by using an outer primer set designated RD. The RD primer set consisted of an upstream primer, 5′ GACC TTAGAGACTGGC 3′ (nt 5655 to 5670 of the BaEV sequence [14]), located in the pol gene and a downstream primer, 5′ GCTGCAATCGCATGG 3′ (nt 6343 to 6357 of BaEV), located in the env gene. A nested set, consisting of primers upstream 5′ GAGAGGCCCTCTTATCTC 3′ (nt 5681 to 5696 of BaEV) and downstream 5′ GCGAGGGTCTGTAACCC 3′ (nt 6289 to 6308 of BaEV),

* Corresponding author. Mailing address: Department of Human Retrovirology, Academic Medical Centre, University of Amsterdam, Meibergdreef 15, 1105 AZ Amsterdam, The Netherlands. Phone: 31 20 566 4522. Fax: 31 20 691 6531. E-mail: a.c.vanderkuyl@amc.uva.nl.

Received 29 March 1999/Accepted 17 June 1999

Copyright © 1999, American Society for Microbiology. All Rights Reserved.

Downloaded from http://jvi.asm.org/ on August 14, 2017 by guest
was used in a second amplification reaction, if necessary. A PCR fragment could be obtained from F. chaus only by using the nested primer set directly on the genomic DNA, probably because of mismatches with the outer primer set. For FcEV amplification, the upstream primers were combined with downstream primers based on the env sequence of this virus: 5′ GGGAAGTTGTCAGGGAGGT 3′ and the nested primer 5′ GAGGGGAACCCACATAGGT 3′. This primer combination was designated FC. First-round PCR amplifications were performed under the following conditions: denaturation for 5 min at 95°C; amplification consisting of 35 cycles of 1 min at 95°C, 1 min at 55°C, and 2 min at 72°C; followed by an extension of 10 min at 72°C. The protocol for the nested reactions was identical, except that only 25 cycles were performed. The obtained fragments were cloned into the TA vector (Invitrogen, San Diego, Calif.) and sequenced. At least two clones from a single individual were analyzed.

DNA sequencing. Sequencing of the clones with an ABI 373A or ABI 377 automated sequencer was done in both directions directly from lambda DNA with purified specific primers, as specified by the manufacturer. PCR fragments obtained from genomic DNA were sequenced directly from plasmid DNA with T7 and M13 dye primers.

Analysis of the Fc21 and Fc41 env genes. Provirial clones Fc21 and Fc41 were obtained by hybridizing a cat genomic library to a RD-114/FeLV recombinant virus (4). The intergenic region of Fc21 and Fc41 is almost 430 nt encoding part of the endonuclease (integrase) was additionally sequenced (see Fig. 4). The pol sequences from Fc21 and Fc41 were then analyzed together with the amplified fragments from cat and monkey genomic DNA (see Fig. 5). The Fc21 and Fc41 pol genes are C type and are closely related to RD-114 pol. The intergenic region of Fc21 and Fc41 is almost identical to that of RD-114 (see Fig. 4). However, based upon the divergent env sequences of the Fc21 and Fc41 genomic clones, we postulated that the proviral genomes present in these clones represent two closely related virus strains distinct from RD-114. This novel endogenous virus we named Felis catus endogenous virus (FcEV).
FIG. 1. (A) Alignment of the cat genomic proviral clones Fc21 and Fc41. The stop codons of the pol and env genes and the start codon of the env gene are indicated by arrows and underlined. (B) Alignment of amino acid sequences of Fc21 and Fc41 derived from the sequences in Fig. 1A. The Fc41 reading frame has been modified to allow translation by the removal of 1 nt and the insertion of 2 nt, according to the Fc21 ORF. The putative signal peptide and the gp70 and P15E proteins are indicated. Processing sites were obtained by comparison of Fc21 and Fc41 with the GA-FeLV-B env protein (8). The putative immunosuppressive peptide is shaded.
the env gene. Translation of the pol ORF showed that many substitutions are silent and do not lead to amino acid replacement compared to BaEV. The RD-114 fragments amplified from the three species of cats are almost identical, suggesting that the time of divergent evolution is short and that probably the same single locus has been amplified. Cat FcEV fragments showed a little more variation than did the RD-114 sequences. BaEV fragments amplified with primer set RD from monkey DNA were even more variable, suggesting that different loci have been amplified or that the total time of divergent evolution is longer. Defective clones, containing premature stop codons or small deletions, were amplified from both monkey and cat DNA.

With the pol-env fragment used, no distinction can be made between the BaEV for strain infecting mangabeys and mandrills and the BaEV sav strain infecting baboons, geladas, and African green monkeys, since the strains were separated earlier by using LTR sequences and, to a lesser extent, env sequences.
(26). The few nucleotide differences between the pol-env fragment of the cultured RD-114 isolate (7) and the sequences obtained from cat genomic DNA (Fig. 4) could be due either to mutations acquired during culturing or to sequence artifacts.

**Phylogenetic analysis of BaEV, RD-114, and FcEV fragments.** NJ analysis of amplified fragments showed that both endogenous cat viruses clustered together and away from monkey BaEV (Fig. 5A). Surprisingly, the Fc41 fragment was basal to all cat sequences and was most closely related to monkey viruses, which suggested that Fc41 represented the most ancient integration. Analysis of only pol sequences (Fig. 5B) showed more clearly that all cat virus pol fragments had evolved from Fc41 pol. This observation was confirmed in a linearized NJ tree (not shown), where the rate constancy is calculated with respect to an outgroup and subsequently forced upon the sequences. Therefore, RD-114 pol sequences are almost certainly derived from FcEV pol genes. The near absence of divergence in RD-114 from different cat species again suggested that a single locus has been amplified. An NJ analysis of pol-derived amino acid sequences generated a tree identical to the nucleotide pol tree (not shown).

**DISCUSSION**

To investigate the evolution of the monkey retrovirus BaEV after its presumed transmission to cats, as suggested by sequence comparison, we have sequenced the env genes and adjacent genome parts of two RD-114 proviruses integrated in the genome of the domestic cat (*F. catus*). Analysis of the two clones (Fc21 and Fc41) revealed that their env genes are of type C, in contrast to the type D env found in RD-114. Earlier hybridization studies had already shown that all putative RD-114 proviruses analyzed possessed env genes unrelated to inducible RD-114 (20, 23). Most of these integrations are full-length proviruses with gag, pol, and env genes flanked by twoLTRs. All or most proviruses formerly attributed to RD-114 are more likely to be FcEV integrations, and we assume that the number of 15 to 20 copies per haploid genome is probably the copy number of FcEV. It would be worthwhile to determine the actual FcEV copy number, its gag sequence, and its distribution in feline species. Earlier experiments involving Southern hybridization showed that the RD-114-unrelated env sequences were also unique for the genus *Felis* and do not occur in primates (20). It is possible that the single inducible RD-114 locus is the only representative of the true RD-114 virus in the cat genome. Domestic cats are polymorphic with respect to the presence or absence of the inducible RD-114 provirus on chromosome B3 (19), suggesting that this integration is recent in evolutionary terms.

The origin of the FcEV env gene is most probably a murine type C virus. Database homology searches and phylogenetic analysis of type C Env proteins showed that FcEV env, and the env-LTR intergenic region, are related to the Fv-4 resistance gene and its downstream sequence of Asian *Mus musculus* (10). The Fv-4 locus contains the env gene of an ecotropic MuLV, whose expression protects carrier mice from new ecotropic MuLV infections. It would be interesting to see if expression of FcEV env genes occurs and, if so, if it protects cats from related infections. It has been shown previously that FeLV is also derived from rodent viruses. Elder and Mullins
noted homologies between the env genes of FeLV-B and a murine mink cell focus-forming virus, while Benveniste et al. observed a larger degree of homology to rat sequences. The pol gene of Fc41 was found to be closely related to monkey type C pol genes, especially to the recently characterized PcEV pol gene, and less closely related to the BaEV pol gene. PcEV is a full-length type C virus identified in baboons but is present in most Old World monkey genomes (17). PcEV is one of the putative ancestors of BaEV. Phylogenetic analysis showed that all RD-114-like pol genes amplified or isolated from cats, including Fc21 and inducible RD-114, have evolved

FIG. 4. Alignment of a pol-env fragment amplified from monkey and cat species. The pol and env parts of the sequence are shown translated into amino acids, while the intergenic region is shown as nucleotides. The sequence corresponds to nt 5738 to 6288 of the BaEV reference sequence (14). Gaps introduced for optimal alignment are indicated by dots, and identical nucleotides are indicated by dashes. Stop codons are indicated by an asterisk, and incomplete codons (due to single nucleotide deletions) are indicated by a question mark. The primer set (RD or FC) used in generation of the fragments is identified. (8)
from the Fc41 pol gene. Thus, the viral integration contained in the Fc41 clone is more ancient than inducible RD-114 but contains a type C env gene. Fc41 thus represents an endogenous type C retrovirus distinct from the prototype RD-114 and should be classified separately as FcEV, together with the closely related Fc21 provirus. Probably, the FcEV provirus present in clone Fc21 represents a less ancient integration of FcEV, integrated much later in the cat genome. Exogenous retroviruses have an increased substitution rate compared to endogenous retroviruses.

RD-114 contains a pol gene distinct from BaEV pol, while cat genomes harbor multiple copies of a type C retrovirus closely related to RD-114 pol. Also, the pol-env intergenic region and the LTRs of RD-114 and FcEV are closely related. The primer binding site of RD-114 and all RD-114-related viruses was found to be complementary to tRNAGly (22), in contrast to the PBS of BaEV, which utilizes tRNAPro to initiate replication (14).

From the phylogenetic analyses, the copy number (15 to 20 per haploid genome, as suggested in reference 20) and the non-inducibility of the FcEV proviruses, it can be assumed that they are older than the RD-114 integration. Although newly integrated proviruses can also be defective, proviruses generally become increasingly suppressed by their host as a function of time. Phylogenetic analysis showed that it is unlikely that the pol gene of RD-114 and the untranslated parts of its genome
are derived from BaEV. However, the RD-114 env gene is highly homologous to BaEV env, suggesting that RD-114 is a recombinant virus between the cat virus FcEV and primate BaEV. Recombination could have been facilitated by the two homologous stretches, one located in the signal peptide sequence of the env gene and the other just downstream of the env stop codon. The presence of a separate env gene in FcEV, as in type D viruses, facilitated the exchange of this gene. BaEV thus did infect a cat ancestor, but it was a newly recombinant virus that made its way into the cat germline. Recombination is a common strategy in retroviruses, as is presently seen in the human immunodeficiency virus type 1 pandemic, where approximately 10% of viruses isolated are recombinants (21), either between different subtypes or between different strains. In simian immunodeficiency virus, the virus harbored by sabaeus monkeys is a recombinant between African green monkey and sooty mangabey simian immunodeficiency viruses (12). Also, BaEV itself is a recombinant between C-type (PcEV) and D-type (SERV) primate retroviruses (17, 27).

Cats of the genus Felis thus harbor at least three unique retroviruses in their genome: (i) FcEV, a type C virus, of which presumably 15 to 20 copies exist; (ii) the FcEV-BaEV recombinant virus RD-114, probably present as a single inducible copy; and (iii) multiple endogenous FeLV integrations, which...
can recombine with exogenous FeLVs (18, 24). Since all three viruses must have infected the common ancestor of the domestic cat lineage, this suggests that in the ancestral cat population, two rodent retroviruses and a primate retrovirus were simultaneously spreading. It should be noted that cats also harbor endogenous virus sequences with homology to the macaque type C virus MAC-1 (6).

REFERENCES