
Access from the University of Nottingham repository:
http://eprints.nottingham.ac.uk/2476/1/Talintoncanine.pdf

Copyright and reuse:

The Nottingham ePrints service makes this work by researchers of the University of Nottingham available open access under the following conditions.

This article is made available under the Creative Commons Attribution licence and may be reused according to the conditions of the licence. For more details see:
http://creativecommons.org/licenses/by/2.5/

A note on versions:

The version presented here may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the repository url above for details on accessing the published version and note that access may require a subscription.

For more information, please contact eprints@nottingham.ac.uk
Characterisation of a group of endogenous gammaretroviruses in the canine genome


School of Veterinary Medicine and Science, University of Nottingham, Sutton Bonington Campus, Leicestershire LE12 5RD, UK
Aaron Diamond AIDS Research Centre, 455 First Avenue, 7th Floor, New York, NY 10016, USA

Abstract

Bioinformatics were used to identify and characterise 39 pol, 34 gag and five env gammaretroviruses within the canine (Canis lupus familiaris) reference genome. These endogenous retroviruses are monophyletic to the Canidae, predate the divergence of dogs and foxes and are fixed in 20 canine breeds examined. They are transcribed in normal canine tissue but are unlikely to be replication competent in dogs.

Introduction

Retroviruses exist as exogenous forms, which are infectious, and endogenous forms, which are integrated within the genome of the host. In the retroviral lifecycle, the retroviral reverse transcriptase (RT) creates a double stranded DNA copy of the virus that integrates into the DNA of the host cell and may become inherited if integrated into germ line cells (Boeke and Stoye, 1997).

Endogenous retroviruses (ERVs) are ubiquitous in vertebrates and comprise a large percentage of the host genome, e.g. approximately 10% in mice (Stocking and Kozak, 2008). These ERVs become increasingly degenerate with time due to the accumulation of mutations and internal deletions (indels) (Boeke and Stoye, 1997). More recent integrants tend to resemble their exogenous counterparts and may have pathogenic potential, e.g. in feline leukaemia virus (FeLV), recombination between endogenous and exogenous viruses can generate strains with altered pathogenic properties (Roy-Burman, 1995).

Leukaemias and lymphomas are relatively common in dogs (Canis lupus familiaris) and the frequency differs among breeds (Onions, 1984; Modiano et al., 2005; Lurie et al., 2008). Retroviral particles and RT (polymerase) activity have been reported in canine lymphomas and cell lines (Perk et al., 1992; Modiano et al., 1995; Ghernati et al., 1999, 2000), but endogenous retroviral sequences have not been fully characterised in dogs and no exogenous retroviruses have been identified in this species.

At around 2.4 Gb, the canine genome (Karlsson and Lindblad-Toh, 2008) is compact in comparison with other mammalian genomes. This smaller size is largely due to a lower number of repetitive elements in the genome; 34% compared to 40% in mice and 46% in humans. As a result, the number of ERVs and DNA transposons within the dog genome is low; 26,000 in the dog compared with 180,000 in humans (Lindblad-Toh et al., 2005).

Using the Retrotector programme, Martinez Barrio et al. (2011) identified 407 ERV sequences (0.15%) in the canine genome, which is lower than in other vertebrate genomes. Jo et al. (2012) identified a smaller number of ERVs (184 sequences), but also identified three ERVs that had not been detected by Martinez Barrio et al. (2011). These studies differed in their phylogenetic classification of the ERV elements.

The low number of ERVs and apparent lack of exogenous retroviruses in dogs is surprising, since the canine genome lacks functional copies of the retroviral restriction factor TRIM5α and canine cells support replication of several non-canid exogenous retroviruses (Fadel and Poeschla, 2011). In this study, we analysed the canine genome for endogenous gammaretroviral sequences.

Materials and methods

Ethics approval

The project was approved by the University of Nottingham School of Veterinary Medicine and Science Non-Animals (Scientific Procedures) Act Committee. Tissue samples were collected post-mortem from animals euthanased for clinical reasons with full informed consent of the owners.

RNA extraction

Samples of lung, liver, kidney, spleen, mesenteric lymph node (MLN) and jejunum were collected from four aged German shepherd dogs euthanased for medical reasons (Table 1). The tissues were stored in RNA later (Qiagen), held at 4 °C overnight and then stored at −20 °C until RNA was extracted using the NucleoSpin
RNA II kit (Macherey–Nagel). RNA from canine liver, kidney, brain and stomach (two ‘normal’ Beagles) was purchased from Ams Biotechnology (Promega) and Moloney murine leukaemia virus (Mo-MuLV) RT (Promega).

CDNA from canine cell lines, normal dog mammary epithelial cells (DMEMCs) and normal skin fibroblasts cells (DSFCs) was kindly provided by Cinzia Allegrucci, School of Veterinary Medicine and Science, University of Nottingham, Nottingham, UK. RNA was also extracted from MDCK (Madin-Darby canine kidney) and D17 (canine osteosarcoma) cell lines.

Data mining and phylogenetic analysis

Data mining was performed on the dog genome (build CanFam2, GenBank number AAX000000000.21) (Lindblad-Toh et al., 2005) Potential gammaretroviral ERVs were predicted using a pipeline of Perl scripts. The complete group antigen (Gag) polymerase (Pol) and envelope (Env) protein sequences of Mo-MuLV (Shinnick et al., 1981) were used with the Exonerate algorithm (Sater and Birney, 2005) to screen each chromosome of the dog genome sequentially (build CanFam2 for potential homologous ERV genes). A minimum of 200 predicted amino acids, without exons, was used as an initial cut-off for calling a putative ERV gene. For overlapping predictions, the highest score from Exonerate was used to determine the best predicted ERV gene, five env 40 gag and 123 pol genes were identified. These initial predictions were further parsed such that each prediction was ≥ 25% of the full length of the query sequence. This produced a filtered core set of predicted genes, which were analysed further.2 Long terminal repeat (LTR)-like sequences were located using BLAST.2

For the most intact locus, part of the predicted Pol amino acid sequence was identified using TBLASTN5 against a retroviral database (NCBI Viral reference genomes). This was aligned with amino acid sequences of known retroviral Pol sequences using MUSCLE6 (Edgar, 2004). A maximum likelihood phylogeny was generated with PhyML7 using the approximate likelihood ratio test model.

Results

Identification of canine endogenous retrovirus sequences

A core set of 39 pol, 34 gag and five env gammaretroviral genes were retrieved. The full data set of the chromosomal locations of these genes is provided in Appendix A (Supplementary Table 1). The search strategy utilised was conservative and only identified one full length proviral locus consisting of gag–pol–env on chromosome 20 (position 36654581–366577; Fig. 1). A further six gag–pol pairs and one gag–env pair were identified; none of these loci had recognisable LTRs. The remaining loci consisted of single genes (five with paired LTRs, 10 with single LTRs, the remainder with none). These loci were named according to Jern et al. (2005). None of these sequences is potentially replication competent, since all contain multiple stop codons and/or frame shifts in all coding domains. The full length locus on chromosome 20 contains no open reading frames >181 codons, with stop codons in all coding domains; only portions of the locus were robustly identified using the search parameters (Fig. 1).

Table 1

<table>
<thead>
<tr>
<th>Dog</th>
<th>Sex</th>
<th>Age</th>
<th>Breed</th>
<th>Clinical information</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Male</td>
<td>Unknown</td>
<td>German shepherd</td>
<td>No significant findings on post-mortem examination</td>
</tr>
<tr>
<td>2</td>
<td>Male</td>
<td>9 years</td>
<td>German shepherd</td>
<td>No significant findings on post-mortem examination</td>
</tr>
<tr>
<td>3</td>
<td>Female</td>
<td>9 years 10 months</td>
<td>German shepherd</td>
<td>Osteoarthritis of the left hip</td>
</tr>
<tr>
<td>4</td>
<td>Male</td>
<td>10 years</td>
<td>German shepherd</td>
<td>Treated for epilepsy. Splenic nodular hyperplasia</td>
</tr>
<tr>
<td>Ams Biotechnology (kidney, stomach, brain)</td>
<td>Male</td>
<td>4 years</td>
<td>Beagle</td>
<td>Healthy</td>
</tr>
<tr>
<td>Ams Biotechnology (liver)</td>
<td>Female</td>
<td>2 years 6 months</td>
<td>Beagle</td>
<td>Healthy</td>
</tr>
</tbody>
</table>

Quantitative real-time PCR (qPCR) was performed on cDNA derived from canine tissues with consensus primers for the pol gene of canine ERVs (CanERVs) identified in this study (forward: 3’-CAGGACAGGTAGAGGATGAACAG-2’; reverse: 3’-TGG GGATAACAGGTGGAGGAAG-2’). Relative expression of CanERVs was normalised against canine glyceraldehyde–3-phosphate dehydrogenase (GAPDH) expression (Table 2). The Absolute QPCR SYBR Green Low ROX master mix (Thermo Scientific) was used with a LightCycler 480 (Roche) qPCR machine. Thermal cycling conditions were initial denaturation of 95 °C for 15 min, then 45 cycles of 95 °C for 15 s, 55 °C for 30 s and 72 °C for 30 s. Samples were run in duplicate.

PCR for individual CanERV loci

Several matches to the consensus sequence of canine ERV genes identified in this study were matched to the Poodle whole genome shotgun sequence dataset (GenBank AACN000000001–AACN0108963) that did not have apparent equivalents in the reference Boxer genome (CanFam2). However, this may be due to gaps in the genome build or Y chromosome inserts (the poodle genome is for a male dog; the Boxer genome is for a female dog). Therefore, to check for genuine breed variation, PCR primers were designed for eight poodle CanERV loci (Table 2). Primer sets had one primer within the retroviral sequence and one within distinct flanking or indel genomic DNA sequences. GAPDH primers were used as positive controls for DNA quality.

Reaction mixtures contained 1× buffer (NEB), 2 mM MgCl₂, 5 U Taq polymerase, 20 μM F primer, 20 μM R primer, 20 μM deoxynucleotide triphosphates (dNTPs, NEB) and MgCl₂ concentrations as shown in Table 2. Thermocycling conditions using a Techne TC-512 PCR machine were 95 °C for 2 min, then 30 cycles of 95 °C for 30 s, 50–65 °C for 30 s (Table 2), 72 °C for 30 s and a final cycle of 72 °C for 2 min. Samples that tested negative had DNA re-extracted and were tested again before being accepted as negative. To confirm sequence identity of the loci, PCR bands of the expected size were cloned into the PGM-T easy vector system (Promega), transformed into Escherichia coli C29885 (NEB) and the plasmids were sequenced using Big Dye Terminator 2.1 and an ABI 3130 analyser (Applied Biosystems).

Statistical analysis

The relative standard curve method (Applied Biosystems) was used to calculate pol gene expression relative to a calibrator sample. Statistical analysis was performed using the Kruskal–Wallis test with Dunn’s multiple comparisons test used as a post hoc test. Statistical analyses were performed using GraphPad Prism version 5.04.
Transcription of canine endogenous retrovirus sequences

Expression of pol gene RNA by reverse transcriptase (RT)-qPCR was detected in all tissues and cell lines tested. Expression relative to GAPDH did not vary substantially (range 0.82–1.47). There were no statistically significant differences in expression between any tissues (P = 0.862) or animals (P = 0.753) (Figs. 3 and 4). The PCR primers amplified CanERV sequences (confirmed by Sanger sequencing) from canine and fox DNA and RNA, but did not produce PCR products from unrelated species, such as the horse (data not shown).

Endogenous retrovirus polymorphisms in canine breeds

PCR amplicons were produced from all blood samples with the GAPDH and CanERV primer sets (CanERV1–7).

Discussion

While the CanERVs identified here are defective and therefore not capable of forming replication competent retroviruses, they are monophyletic to the Canidae, as demonstrated by the pol phylogenetic analysis. The most closely related sequences are present in the red fox and LTR divergence analysis indicates that these ERVs entered this lineage of Canidae prior to the split that led to the red fox and modern dog.

Based on the pol phylogenetic tree, these CanERVs are firmly positioned between the group containing the known exogenous Mo-MuLV-like viruses and the reticuloendotheliosis virus (REV)-like viruses, with the nearest relative being an endogenous retrovirus from the opossum genome. The clustering with the opossum sequence probably reflects the lack of other Canidae sequences available rather than a real species jump.

Of the sequences identified in our study, 4/5 env, 32/34 gag and 25/39 pol sequences were reported previously by Jo et al. (2012). Our sequences match the CIERV γ1 group most closely, although Jo et al. (2012) report a larger number of loci (55) in this group and many of our sequences also have matches with their γ9, 10 and 17 groups. These latter groups were classified separately by Jo et al. (2012) because they are missing portions of pol. In our analysis, these sequences cluster with other CanERV sequences, implying that this is an artificial grouping. In our analysis, as well as that of Jo et al. (2012), these viruses group closely with the previously identified fox sequence, but their relationships with other gammaretroviruses vary; these are also different again from those reported in Martinez Barrio et al. (2011), who did not place any sequences within this clade.

These differences are likely to be due to the different search strategies employed in these different studies and the different gene regions utilised in phylogenetic analysis. The study by Martinez Barrio et al. (2011) utilised Retrotector, which detects consensus motifs in potential ERV sequences and their phylogenetic analysis was based on pol genes with reverse transcriptase motifs, whereas Jo et al. (2012) applied BLAST to pro/pol sequences and did not utilise their phylogenetic analysis into different structural motifs of the pro/pol gene. The BLAST-based approach is less sensitive for detecting ERVs, but reliably detects structurally intact ERVs (Garcia-Extebarria and Jugo, 2010).
genomes become available, this is an increasing problem in these types of studies. While our strategy robustly identifies Mo-MuLV-like sequences, it will have excluded many less conserved loci.

The loci examined for polymorphisms across different dog breeds are fixed within all breeds examined. This is to be expected given that the integration of these gammaretroviruses predates that of the species divergence of dogs and foxes. Mismatches between the Poodle and Boxer genomes are likely due to sequence artefacts. It is extremely unlikely that breed differences in lymphoma susceptibility in dogs are due to variation in these endogenous gammaretroviral integrations.

While these ERVs do not apparently encode for functional retroviral genes, they are transcribed in canine tissues. Transcription of (or variations) in repetitive elements, such as ERVs and long interspersed nuclear elements (LINEs), have been linked with diseases in humans (Cruickshanks and Tufarelli, 2009). There has also been considerable interest in retroviral contamination in biological products and clinical samples (Sakaguchi et al., 2008; Gray et al., 2011).

The pol gene of at least some of these CanERV sequences is transcribed in normal canine tissues and may be detectable in some screening tests for retroviral contamination of biological products. We did not detect any tissue-specific patterns of expression, but the possibility exists that some tissues in the dog may display differential expression, as has been reported for reproductive tissue in a variety of species (Heidmann et al., 2009). There has been one report of a syncytin-like gene in carnivores (Cornelis et al., 2012). The degenerate env gene in the CanERV sequences identified in our study makes them poor candidates for this type of function.
While this study has provided insights into the retroviral complement of the canine genome, it does not resolve the issue of whether an exogenous ‘canine lymphoma virus’ exists. The sequences reported here are unlikely to be responsible for the previous reports of retroviral particles in canine tissues (Modiano et al., 1995; Ghernati et al., 2000), which may even represent contamination by exogenous viruses from other species. The lack of recent retroviral integrants in the dog genome is surprising, although in this respect the dog is similar to humans, where the most recent ‘endogenisation’ event (that of the HERV-K family) is thought to have occurred 30 MYA; these sequences remained active to within the last 200,000 years and are polymorphic between individuals (Denner et al., 1995).

Conflict of interest statement

None of the authors of this paper has a financial or personal relationship with other people or organisations that could inappropriately influence or bias the content of the paper.

Acknowledgements

The authors wish to thank Dr. Cinzia Allegrucci (School of Veterinary Medicine and Science, University of Nottingham, Nottingham, UK), Dr. Robin Flynn (School of Veterinary Medicine and Science, University of Nottingham, Nottingham, UK) and Dr. Brian Catchpole (Royal Veterinary College, London, UK), for their kind donation of samples used in this project. This project was funded by a Blue Skies research grant from the Royal College of Veterinary Surgeons Charitable Trust and the School of the Veterinary Medicine and Science at the University of Nottingham.

References


Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version, at http://dx.doi.org/10.1016/j.tvjl.2012.08.011.

Fig. 4. Relative quantitative PCR (qPCR) expression of canine endogenous retrovirus (CanERV) pol RNA normalised against glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and calibrated against kidney expression via the relative standard curve method in different animals. Dogs 1–4 were German shepherds. Dogs 5 and 6 were Beagles. Cell lines were MDCK, D17, DMECs and DSFCs. Standard deviations are shown as error bars.