Infection, Genetics and Evolution 11 (2011) 686–693

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Short communication

# Distinctive differences in long terminal repeat sequences between γ1 endogenous retroviruses of African and Eurasian suid species☆

Fabrícia Ferreira do Nascimento [a,1], Jaime Gongora [a], Michael Tristem [b], Stewart Lowden [c], Chris Moran [a,*]

[a] Faculty of Veterinary Science, The University of Sydney, NSW 2006, Australia
[b] Division of Biology, Imperial College London, Silwood Park, Ascot, Berkshire SL5 7PY, United Kingdom
[c] Veterinary Health Research Pty Ltd., Armidale, NSW 2350, Australia

ARTICLE INFO

*Article history:*
Received 26 August 2010
Received in revised form 10 January 2011
Accepted 17 January 2011
Available online 21 January 2011

*Keywords:*
Endogenous retrovirus
Long terminal repeat
Suid
Pig
Hog
Evolutionary analyses

ABSTRACT

Diversity of long terminal repeats (LTRs) from γ1 endogenous retroviruses (ERVs) was analysed by DNA sequencing in 10 species of the family Suidae (suids, pigs and hogs). Phylogenetic analysis separated LTR sequences into two groups which correlated approximately with either the previously described cluster I and III, or the clusters II, IV and V. Interestingly, a specific LTR exhibiting a novel molecular rearrangement was identified exclusively within African host species when compared to LTRs previously reported from known ERVs in the domestic pig (*Sus scrofa*). Furthermore, other sections of LTRs appear to be unique to African suids as suggested by phylogenetic analysis. These differences between African and Eurasian ERV lineages show that these ERVs belong to different viral sub-populations, implying coevolution of endogenous viral sequences with their host species and providing no evidence of transfer of viral sequences between African and Eurasian suids.



## 1. Introduction

Endogenous retroviruses (ERVs) represent ancient germline infections in which the integrated provirus becomes part of the host genome and is subsequently transmitted vertically from one host generation to the next (Boeke and Stoye, 1997; Gifford and Tristem, 2003). Usually, the proviral genome consists of three main coding domains, the *gag*, *pro*/*pol* and *env* genes, flanked by long terminal repeats (LTRs) which are identical at the moment of integration (Gifford and Tristem, 2003).

In domestic pigs and wild boar (*Sus scrofa*), ERVs are known as porcine endogenous retroviruses (PERVs) and are classified into β and γ families (Patience et al., 2001; Klymiuk et al., 2002). Interest in PERVs has intensified due to their potential risk to xenotransplant recipients. The known human-tropic PERVs are classified as γ1 and are divided into three main envelope classes A, B, and C (Takeuchi et al., 1998). Approximately 50 copies are present in the host genome (Le Tissier et al., 1997; Patience et al., 1997; Akiyoshi et al., 1998; Patience et al., 2001), although most are defective. The LTR of PERV-NIH shares 78% nucleotide identity with PERV-C, having identical enhancer-like repeat sequences (Wilson et al., 2003). The two sequences from PERV-A and B generated by Wilson et al. (2003) showed slight differences in their LTR regions, being almost identical except for an extra 39 bp repeat element in the former. These A and B LTRs showed 64% nucleotide sequence identity with the PERV-C LTR (Wilson et al., 2003). Modifications to PERV LTRs might be responsible for their adaptation to human cells *in vitro*, specifically an increase in the length of the U3 region of the LTR. Insertion of direct repeats in this region may function as enhancers (Denner et al., 2001, 2003; Scheef et al., 2001; Wilson et al., 2003).

Furthermore, Klymiuk et al. (2008) analysed a higher number of published γ1 LTR sequences from genomic PERV sequences from pigs of various breeds, porcine cell lines, infected human primary cells and transcripts from virus particles released from porcine cell lines. They identified five different phylogenetic clusters (I–V); for convenience, the classical LTRs from PERV-A, B and C were placed in clusters I, II and III, respectively, although some PERV-A sequences were found to group with cluster II (Klymiuk et al., 2008).

Few studies have investigated ERV distribution and relationships in the Suidae (suids, pigs and hogs) (Patience et al., 2001; Niebert and Tönjes, 2005) and no study has determined if ERVs from this different suid hosts belong to different viral sub-populations. This is

☆ GenBank accession numbers: FJ357714–FJ357741, FJ357743–FJ357780, GQ906343–GQ906349 and HQ456136–HQ456208 for sequences generated in this study.

* Corresponding author at: R.M.C. Gunn Building, Faculty of Veterinary Science, The University of Sydney, NSW 2006, Australia. Tel.: +61 2 93513553; fax: +61 2 93512114.
*E-mail address:* christopher.moran@sydney.edu.au (C. Moran).
[1] Present address: Instituto de Microbiologia Professor Paulo de Góes, Universidade Federal do Rio de Janeiro, 21941-590, Rio de Janeiro, RJ, Brazil.

1567-1348/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.meegid.2011.01.010

partly because the host family relationships were not clarified until recently (Nascimento, 2009; Gongora et al., unpublished results) by analysis of nuclear and mitochondrial DNA sequences from Eurasian and African suid hosts. This has enabled the investigation of ERVs from Eurasian and African species to determine the relationship between LTRs and host phylogeny, and to determine the diversity of LTR groups in relation to retroviral *env* gene sequence using the terminology proposed by Klymiuk et al. (2008).

## 2. Materials and methods

### 2.1. PCR assays

The upstream (5′) and downstream (3′) LTR DNA sequences were amplified from 12 animals from 10 species and one subspecies of the family Suidae (Table 1) using one LTR primer and one primer within the *gag* or *env* gene, respectively (Table 2). In the former, the 5′ most end of the LTR is not amplified whereas the 3′ end, is not amplified in the latter. Products were amplified by PCR with an initial denaturation at 94 °C for 30 s followed by 35 cycles of denaturation at 94 °C for 30 s, annealing at 58 °C for 60 s, extension at 72 °C for 90 s, and a final extension of 72 °C for 7 min. Amplicons were resolved on a 1.5% agarose gel and fragments of expected size (Table 2) were purified following gel band excision using UltraClean™ Gel Spin DNA Purification kit (Mo Bio, Australia).

### 2.2. Cloning and sequencing

Purified PCR fragments were cloned using the TOPO TA cloning kit® (Invitrogen, Australia) and DNA from plasmids containing inserts was subsequently extracted with UltraClean™ Mini Plasmid Prep kit (Mo Bio). Different numbers of colonies for each animal (Table 1) were sequenced by the Australian Genome Research Facility Ltd. (Brisbane, Australia) using primers listed in Table 2. The retroviral origin of sequences was confirmed by blastn (Altschul et al., 1997) implemented in the NCBI homepage (http://blast.ncbi.nlm.nih.gov/Blast.cgi/). Electropherograms were then checked using BIOEDIT (version 7.0.9.0) (Hall, 1999).

### 2.3. Alignment and pairwise genetic distance estimations

We have merged 5′ and 3′ LTR sequences generated in this study with an alignment kindly provided by N. Klymiuk which was used in his paper (Klymiuk et al., 2008). A new alignment was performed with MUSCLE (version 3.6) (Edgar, 2004) and manually checked. Columns showing gaps in the majority of aligned sequences and nucleotides at the minority of sequences were manually removed. Conversely, columns showing gaps at sites where nucleotides were present in the majority of sequences were not removed, and these gaps were treated as missing data (Yang, 2006). Novel sequences were analysed along with 91 other sequences from *S. scrofa* available in GenBank.

Genetic distances were estimated using the modified Log-Det (Tamura and Kumar, 2002) implemented in the software MEGA (version 4) (Tamura et al., 2007) which has been demonstrated to estimate reliable distances for closely related sequences (Tamura and Kumar, 2002) and identical sequences were removed from the alignment.

### 2.4. Test of sub-population structure and detection of recombination

To test whether viral sequences of Eurasian and African host species belong to statistically different sub-populations, we explored the data sets with a test for population sub-division developed by Hudson et al. (1992) and adapted to study HIV populations by Achaz et al. (2004). This non-parametric test of sub-population subdivision was performed using a web-based interface (http://wwwabi.snv.jussieu.fr/achaz/hudsontest.html) with a nominal significance threshold of 0.05 for accepting or rejecting a null hypothesis for structure. Because recombinants may misplace sequences in a phylogeny (Kosiol et al., 2006; Jermiin et al., 2008), they were detected using two different softwares. Firstly, the PHI-NNet algorithm described in Salemi et al. (2008) implemented in the software SPLITS TREE 4 (version 4.8) (Huson and Bryant, 2006) was used to detect recombination between LTR sequences. It has been shown that this algorithm can efficiently detect recombination among closely related sequences (Salemi et al., 2008) as is the case for ERV sequences analysed in this study. Secondly, recombinants were also detected using the RDP 3 software (Martin et al., 2005), which implements seven different recombination programs, (1) the original RDP method (Martin and Rybicki, 2000), (2) the GENECONV method (Sawyer, 1989; Padidam et al., 1999), (3) the MaxChi method (Smith, 1992), (4) the Chimaera method (Posada and Crandall, 2001), (5) the SiScan method (Gibbs et al., 2000) and (7) 3SEQ method (Boni et al., 2007). Default settings were used in all analyses and recombination was independently detected in LTR sequences from Eurasian host species and African host species.

### 2.5. Evolutionary analyses

A preliminary network analysis using the neighbor-net (Bryant and Moulton, 2004) implemented in the SPLITS TREE 4 software was carried out with all sequences to evaluate whether the newly generated LTR sequences from Eurasian and African host species would group into different or the same clusters previously reported by Klymiuk et al. (2008).

**Table 1**
List of species and subspecies of Suidae showing the origin, source and number of sequences generated.

| Origin | Species and subspecies | Common name | Source | Tissue | N[a] |
|---|---|---|---|---|---|
| Africa | *Hylochoerus meinertzhageni* | Forest hog | Uganda | Muscle | n/a |
| | *Phacochoerus aethiopicus* | Desert warthog | Kenya | Muscle | 4 |
| | *Phacochoerus africanus* | Common warthog | Iwaba Zimbabwe | Muscle | 8 |
| | *Potamochoerus larvatus* | Bush-pig | Zimbabwe | Blood | 11 |
| | *Potamochoerus porcus* | Red river hog | Duisburg Zoo, Germany | Muscle | 10[b] |
| Eurasia | *Sus scrofa* | Wild boar | Yorkshire Farm, UK | Blood | 9 |
| Asia Pacific | *Sus barbatus oi* | Western bearded pig | Singapore Zoo, Singapore | Blood | 8 |
| | *Sus barbatus barbatus* | Bornean bearded pig | Singapore Zoo, Singapore | Blood | 8 |
| | *Sus celebensis* | Sulawesi warty pig | Sulawesi, Indonesia | Muscle | 6 |
| | *Sus verrucosus* | Javan warty pig | Poznan Zoo, Poland | Muscle | 9 |
| | *Babyrousa babyrussa* | Babirusa | Surabaya Zoo, Indonesia | Blood | n/a |

n/a = no DNA sequence was amplified.
[a] Number of LTR clones sequenced using both pair of primers.
[b] Eight sequences from one individual and two sequences from a second.