TITLE

COMPARISON BETWEEN SANGER, ILLUMINA AND NANOPORE SEQUENCING EVIDENCING INTRA-HOST VARIATION OF FELINE LEUKEMIA VIRUS THAT INFECTS DOMESTIC CATS.

KEYWORDS: Feline Leukemia Virus, Domestic cats, NGS comparison, recombination

ABSTRACT

Knowledge of Feline Leukemia Virus (FeLV) sequence variation has mainly been developed using Sanger sequencing methods. However, advances in next generation sequencing methods and their broad use in laboratories has been changing our understanding of viral genetics. FeLV sequencing has specific complications with the presence of both exogenous (exFeLV) and endogenous (enFeLV) virus with frequent recombination between them, limiting sequencing approaches. Here we report an FeLV-A and FeLV-B amplicon-based comparison between Sanger, Illumina, and Oxford Nanopore sequencing methods in Chilean domestic cats. We analysed the hypervariable envelope gene, where a higher number of variants as well as recombination with endogenous strains occurs. We detected multiple variants and viral quasispecies infecting the cats. We compared these three methods to evaluate the advantages and disadvantages between them. Although the Sanger method is highly reliable, it showed a high fail rate (many amplicons did not produce useable sequence) and the sequences obtained showed artificial sequence clustering when compared with the NGS methods. Illumina sequencing showed a lower error rate but could not discriminate between exogenous and endogenous viruses. Finally, Oxford Nanopore (MinION) sequencing could successfully detect low-abundance sequences and discriminate between FeLV-A and FeLV-B sequences, although its higher error-rate requires caution in interpretation of the results. Our results indicate advantages and disadvantages for each method, with the purpose of sequencing needing to be considered in the choice of method. Results of large viral phylogenetic trees combing sequences derived from mixed sequencing methods, such as those combining historical and contemporary sequencing need to be considered with some caution as artificial clustering by sequencing method may occur.

INTRODUCTION

Feline leukemia virus (FeLV) is an enveloped retrovirus that belongs to the family *Retroviridae*, genus *Gammaretrovirus* (Coffin et al. 2021). The FeLV genome consists of two copies of a single positive sense RNA around 8.4 kb in length. It encompasses three genes:

gag (capsid proteins), *pol* (reverse transcriptase enzymes, protease, and integrase), and *env* (envelope proteins). These are flanked, 5' and 3', by two identical un-translated regulatory sequences called LTR's (long terminal repeats). During infection, the RNA genome is copied into a DNA genome by the reverse transcriptase enzyme and is inserted as a provirus into the host genome (Willett and Hosie 2013).

FeLV has a world-wide distribution in domestic cat populations and is one of the most important pathogens for their health (Hartmann and Hofmann-Lehmann 2020). The virus causes clinical signs such as: immune and bone marrow suppression, lymphadenopathy, lymphoma, leukaemia, oral lesions and respiratory diseases (Hartmann 2012). The virus is excreted in saliva, urine, faeces, milk and nasal secretions (Torres, Mathiason, and Hoover 2005). Social activities like grooming, sharing food and water dishes, nursing and bites during fights are the main transmission routes. It can also be transmitted via blood contact (blood transfusion or contaminated instruments) or sharing of litter trays (Little et al. 2020). Another important route of transmission is vertically, where the queen cat spreads virus to her neonatal kittens, either trans-placentally, during parturition or during nursing of kittens (Hardy et al. 1976). FeLV also has importance in wild felid health and conservation causing severe disease and mortality (Cunningham et al. 2008, Meli et al. 2010a, Chiu et al. 2019). This has been notably harmful in genetically bottle-necked populations (Chiu et al. 2019). Although, the virus can persist in wild species, domestic cats have been the infection source for non-domestic animals in the majority of cases (Meli et al. 2010b, Brown et al. 2008, Petch et al. 2022).

On the other hand, endogenous FeLV (enFeLV) are retroviruses that have integrated over millions of years into the cat genome. EnFeLV were acquired during the divergence of the Felidae and are present in domestic cat chromosomes and their close relatives in the *Felis* genus. They are transmitted from parents to descendants through chromosomes (Polani et al. 2010). Exogenous and endogenous viruses share 86% similarity at the nucleotide level with the main variation points occurring in *gag* and *env* (Chiu, Hoover, and Vandewoude 2018). The enFeLV in cat populations have multiple variations, influenced by the integration site, viral transposition events and several independent integrations during evolutionary history. Even the cats gender may influence the enFeLV quantity, due to an increased load on the Y chromosome compared to the X chromosome (Roca et al. 2005, Powers et al. 2018). Recently studies have associated higher loads of enFeLV with protective factors against exFeLV infections (Powers et al. 2018, Chiu and VandeWoude 2020). Conversely the presence of enFeLV can result in recombination events with exFeLV and produce more virulent variants (Willett and Hosie 2013).

There are six FeLV groups: FeLV-A, FeLV-B, FeLV-C, FeLV-D, FeLV-E, and FeLV-T. These arise from recombination between exogenous and endogenous viruses or mutation of exogenous viruses (Willett and Hosie 2013). The majority of the variation occurs in the *env* gene, which is the binding point between the virus and host cell (Cano-Ortiz et al. 2022). The *env* gene is composed of the Surface protein (SU) and Transmembrane protein (TM). SU includes three regions from 5' to 3: Receptor binding domain (RBD), Proline rich region (PRR) and C-Domain. The RBD and C-Domain are involved in cellular tropism, therefore changes in these regions can modify cell receptor usage and so change clinical progression (Faix et al. 2002, Cano-Ortiz et al. 2022). More specifically, there are three main Variable Regions (VRA,VRB,VRC) identified, along with specific amino acid changes which are associated with changes in receptor usage (Boomer et al. 1997).

FeLV-A is the most widespread group that can be horizontally transmitted and often described as the least pathogenic subtype (Ahmad and Levy 2010, Biezus et al. 2023). All other groups arise de novo from FeLV-A mutation or by recombination with endogenous elements (Hartmann and Hofmann-Lehmann 2020). FeLV-B is the second most common subgroup, being detected in roughly half of all FeLV-A cases (Powers et al. 2018). FeLV-B is a recombinant between FeLV-A and enFeLV viruses and it is associated with higher numbers of neoplastic malignancies, like mediastinal lymphoma and multicentric lymphoma. Overall, FeLV-B is associated with worse clinical symptomatology and prognosis (Ahmad and Levy 2010). Although FeLV-B has historically been described as unable to be horizontally transmitted without FeLV-A, there are documented cases confirming horizontal transmission of FeLV-B without FeLV-A (Stewart et al. 2013, Chiu et al. 2019). FeLV-C, FeLV-E and FeLV-T arise from FeLV-A mutation, in general, these subtypes are associated with increased virulence but have diminished transmission ability. FeLV-C is an infrequent subgroup related to aplastic anaemia (Riedel et al. 1986), while FeLV-T is a T-cytopathic virus which has been related to acquired immune deficiency syndrome (FeLV-FAIDS) (Overbaugh 2000). FeLV-E is the last group described and was found in a natural thymic lymphoma in one cat (Miyake et al. 2016). FeLV-D is the most distant FeLV, because it is a recombinant between FeLV-A and another phylogenetically distinct endogenous retrovirus (not enFeLV), called domestic cat endogenous retrovirus (ERV-DC) (Anai et al. 2012).

Although FeLV does not have as high mutation rate as other retroviruses like Human Immunodeficiency Virus (HIV) or Feline Immunodeficiency Virus (Perelson et al. 1996, Bendinelli et al. 1995), multiple FeLV variants can be detected in a cat population or a single cat (Watanabe et al. 2013, Erbeck et al. 2021). To detect viral quasispecies variations, next-generation sequencing has been broadly used in HIV to evaluate virus variations related to drug-resistance (Teo et al. 2022). This is in contrast to FeLV studies, where the majority of

env gene analysis have been performed using plasmid libraries and Sanger sequencing to evaluate the viral diversity (Watanabe et al. 2013, Erbeck et al. 2021, Petch et al. 2022, Biezus et al. 2023). To our knowledge, NGS in FeLV has been used in only two studies, these were: Roche 454 sequencing of an FeLV outbreak in Iberian Lynx in Spain (Geret et al. 2011) and one study using Illumina sequencing demonstrating decreased enFeLV expression in lymphomas from exFeLV negative cats (Krunic et al. 2015).

FeLV epidemiology is related to multiple factors including: household conditions, cat population management and vaccination rates, even the purchasing power parity per capita of owners can influence prevalence rates (Ludwick and Clymer 2019). Thus, developed countries generally have a controlled situation with FeLV. The current prevalence in Europe has been described as between 0.7%-5.5% (Studer et al. 2019), with a 3.1% FeLV prevalence in North-America (Burling et al. 2017). Conversely, the situation in South America, is highly variable ranging from 3% in north-eastern Brazil (Lacerda et al. 2017) up to 59.44% in Colombia (Ortega et al. 2020), whilst in Chile, the prevalence has been described as between 20.2%-33% in rural areas (Mora et al. 2015, Sacristán et al. 2021).

This study was developed in Chilean domestic cats, with the aim of evaluating the diversity of *env* gene in FeLV provirus, using the traditional sequencing method (Sanger) compared with short and long read next generation sequencing methods (Illumina and Nanopore) in order to examine sequence diversity and recombination between exogenous and endogenous FeLV with higher precision than previously possible.

METHODS

Samples and DNA extraction

Eight domestic cat blood samples were collected in EDTA by jugular venipuncture and kept at -20°C. All procedures and handling were done by veterinarians during clinical diagnostics in the city of Concepcion under the owner's consent. Overarching ethical approval for this study was granted by the University of Nottingham School of Veterinary Medicine and Science Committee for Animal Research and Ethics (CARE) from the number 3672 220923.

Nucleic acid extractions were processed at the Haiken Laboratory, Concepcion, using a Geneaid® kit extraction following manufacturer's instructions. PCR diagnosis of FeLV was made using 3'LTR primers (Cattori et al. 2006) as a first screen at Haiken Laboratory. Genomic DNA was then shipped to the School of Veterinary Medicine and Science, University of Nottingham United Kingdom.

End-point PCR and Sample Preparation

End-point PCR using two primer pairs following methods from Erbeck *et al*, (2020) was used to amplify the envelope gene hypervariable region and 3'LTR region from FeLV-A and FeLV-B. PCR was performed using GoTaq® Long PCR Master mix (Promega) and each reaction consisted of: 12.5µl of GoTaq® Long PCR Master mix; 0.5µl (10µM) Forward primer ; 0.5µl (10µM) Reverse primer; 9.5µl of Nuclease-Free water; 2µl of template DNA to reach a final volume of 25µl per reaction. The same protocol was used for each pair of primers and was performed in triplicate to confirm results. Thermo-cycler conditions used a Gene-Touch Thermal-cycler, from Bioer Technology® following the Erbeck *et al*., (2020) protocol with some modifications according to the polymerase requirements. The protocol was: initial denaturation at 94ºC for 2 min followed by 45 cycles of denaturation at 95ºC for 30 s; annealing at 63 ºC for 30 s; extension at 72ºC for 2 mins and a final extension at 72ºC at 2 mins. The thermo-cycler conditions were tested with a gradient PCR to select the optimal annealing temperature and every set of PCR reaction included negative control with no template (DNA free water) and a positive control (FeLV-A DNA) which was kindly donated by Dr. Margaret Hosie, University of Glasgow. The PCR products were run by electrophoresis on a 1% agarose TAE (Fisher scientific®) gel at 80V, 400 mA for 60 minutes. DNA purification used the Nucleospin® extract II kit (Macherey-Nagel®), according to manufacturer's instructions.

DNA quantification was performed with a Qubit 4 Fluorometer (Thermo Fisher Scientific®). FeLV-A and FeLV-B amplicons of high enough quality and quantity were selected for Illumina and Nanopore sequencing.

Table 1 Sequencing method in Chilean domestic cats

Cat name	Sequencing Method	Amplicon sequenced and number of long sequences	Age	Sex	Genbank Accession
1. Omali	Sanger	FeLV-A (1) / FeLV-B (1)	n/a	n/a	OR513915 /OR513914
1. Maqui	Sanger	FeLV-A (1)	n/a	n/a	OR513916
2. Coco	Sanger	FeLV-A (1)	2 y	Female	OR513918
3. Leon	Sanger	FeLV-A (1)	5 y	Male	OR513917
4. Polo	Nanopore	FeLV-A (3) / FeLV-B (5)	11 y	Male	
5. Adonis	Nanopore	FeLV-A (8) / FeLV-B (2)	n/a	n/a	
6. Cuchi	Illumina / Nanopore	FeLV-A (6) / FeLV-B (3)	3 y	Male	SRA:PRJNA1014909
7. Felix	Illumina / Nanopore	FeLV-A (5) / FeLV-B (2)	6 y	Male	

n/a: Not available; y: year

Genome sequencing

Amplicons from 8 domestic cats were sequenced by three different methods detailed in Table 1 and all results sequencing were submitted to GenBank. Five amplicons were sequenced by Sanger method in a commercial laboratory (Eurofins Inc., Germany).

For Illumina sequencing, amplicons of FeLV-A and FeLV-B were pooled and sequenced using the Illumina platform (HiSeq4000) with 300-bp paired-end distances using a coverage of 30x by Novogene Europe, Cambridge, United Kingdom.

The libraries for Nanopore sequencing were processed following the Native Barcoding Kit 24 V14 protocol (Oxford Nanopore Technologies) with FeLV-A and FeLV-B amplicons individually barcoded. Library Quantification was carried out using a Qubit fluorometer and the Qubit dsDNA HS Assay kit (Thermofisher). 20 fmol of DNA library was loaded onto a R10.4.1 flow cell and sequenced using an Oxford Nanopore MinION Mk1C.

Read Processing, Assembly and Mapping

Illumina reads were trimmed and adaptors removed using FastP v0.23.1 (Chen et al. 2018). The Nanopore raw reads were base called using Guppy v6.4 and super accurate model was used. and Nanopore adaptors were eliminated using Porechop v0.2.4 (Wick et al. 2017), NanoFilt v2.6.0 (De Coster et al. 2018) was used to filter and discard reads shorter than 500 and longer than 4,000 bp with a quality of <Q10.

De Novo Assembly from Illumina reads was performed using MEGAHIT (Li et al. 2015) and for Nanopore reads, Flye (Lin et al. 2016) was used for FeLV-A barcodes. FeLV-B barcodes did not assemble reliably in Flye due to a mix of endogenous and exogenous sequences.

The contigs generated by MEGAHIT and raw reads from Nanopore, were mapped to FeLV-A FAIDS (M18247), enFeLV (M25425, LC196053), FeLV-T (M89997); FeLV-C (M14331), (MT302119, MT302153 and MT681671), Gardner-Arnstein FeLV-B (V01172) and FeLV-A and FeLV-B Chilean Sanger sequences. Illumina sequences were mapped using Bowtie2 (Langmead and Salzberg 2012). Nanopore reads were mapped using Minimap2 (Li 2018) with two preset options: the first was all raw reads by the more sensitive preset options “Oxford nanopore reads” (map-ont). The second approach was mapping to specific variations of interest (Chilean Sanger, FeLV-T, FeLV-E and US FeLV-B) with raw reads and using the preset options for “long assembly increased the specificity up to 5% sequence divergence” (asm5) to consider those under-represented sequences that could be missed by Flye. Reads were considered in the second approach if there were at least 30 reads with similar variations to generate a consensus sequence.

All contigs and consensus sequence were extracted and imported to Geneious v2023.0.4 (Biomatters Inc., Newark, NJ) to annotate and continue phylogenetic analysis.

Phylogenetic Analysis of FeLV

The dataset was constructed in Geneious (Biomatters Inc., Newark, NJ) using all available complete *env* gene FeLV related sequences and were annotated and aligned with Muscle (Edgar 2004). This was exported and used to construct phylogenetic trees and evaluate recombination between FeLV-A and enFeLV sequences.

The alignment was sent to the IQtree web server (Trifinopoulos et al. 2016), where a maximum likelihood tree was inferred using the ModelFinder (Kalyaanamoorthy et al. 2017) to select the best-fitted model nucleotide substitution. The model chosen was general time reversible (GTR+F+G4), branch support analysis was ultrafast with 5,000 bootstrap (Hoang et al. 2018). EnFeLV was used as an outgroup to root the tree and the phylogenetic tree was visualized in FigTree v1.4.4 (Rambaut 2018).

Intra-host Variation Analysis and Recombination Analysis

The intra-host single nucleotide variation (iSNV) was done using BWA-MEM (Li 2013) and iVar (Grubaugh et al. 2019). All libraries were mapped to FeLV-A reference sequence (M18247) using BWA-MEM and a consensus sequences were generated using iVar consensus command for each library. All reference sequences generated by iVar were cut at nucleotide 66 (M18247), to initiate at the same nucleotide position. The output was used to call single nucleotide variants and indels in iVar. The minimum quality parameter was <Q10 and the minimum frequency selected was 0.01 and was considered as true only those statistically significative ($p < 0.05$).

Recombination analysis was carried out in RDP v4.101 (Martin et al. 2015) using six different methods and default parameters. The methods used were RDP (Martin et al. 2015), GENECONV (Padidam, Sawyer, and Fauquet 1999), Chimaera (Posada and Crandall 2001), MaxChi (Smith 1992), Bootscan (Martin and Rybicki 2000), 3Seq (Boni, Posada, and Feldman 2007) and Siscan (Gibbs, Armstrong, and Gibbs 2000). Recombinants were considered if at least three methods detected breakpoints. The analysis included the reference sequences used previously and all sequences obtained by the three methods.

RESULTS

Amplicons for FeLV-A PCR showed lengths around the expected amplicon size (~1.9kb) with some products showing multiple bands. All FeLV-B PCR amplicons had similar lengths to the expected amplicon (~1.8 kb). Additionally, some FeLV-B PCR products showed the presence of an extra band below 1.8 kb (these were identified as truncated enFeLV in the NGS results).

Twenty Sanger sequencing (10 of FeLV-A and 10 of FeLV-B) of isolates were performed to confirm PCR product identity. Only four samples of FeLV-A were successfully Sanger sequenced (Omali, Maqui, Coco and Leon) with around 1.8 kb of sequence obtained, these DNA samples however did not have enough volume remaining to sequence with the other methods. BLASTn analysis from NCBI showed 97-99% identity with FeLV-FAIDS (M18247.1) sequence. Only one FeLV-B sample (Omali) was sequenced successfully (1,793bp) by the Sanger method, showing 98% identity with FeLV-B sequences from the USA (MT302107.1). Most amplicons failed to produce good quality results with the Sanger method, due to high interference from the presence of multiple variants.

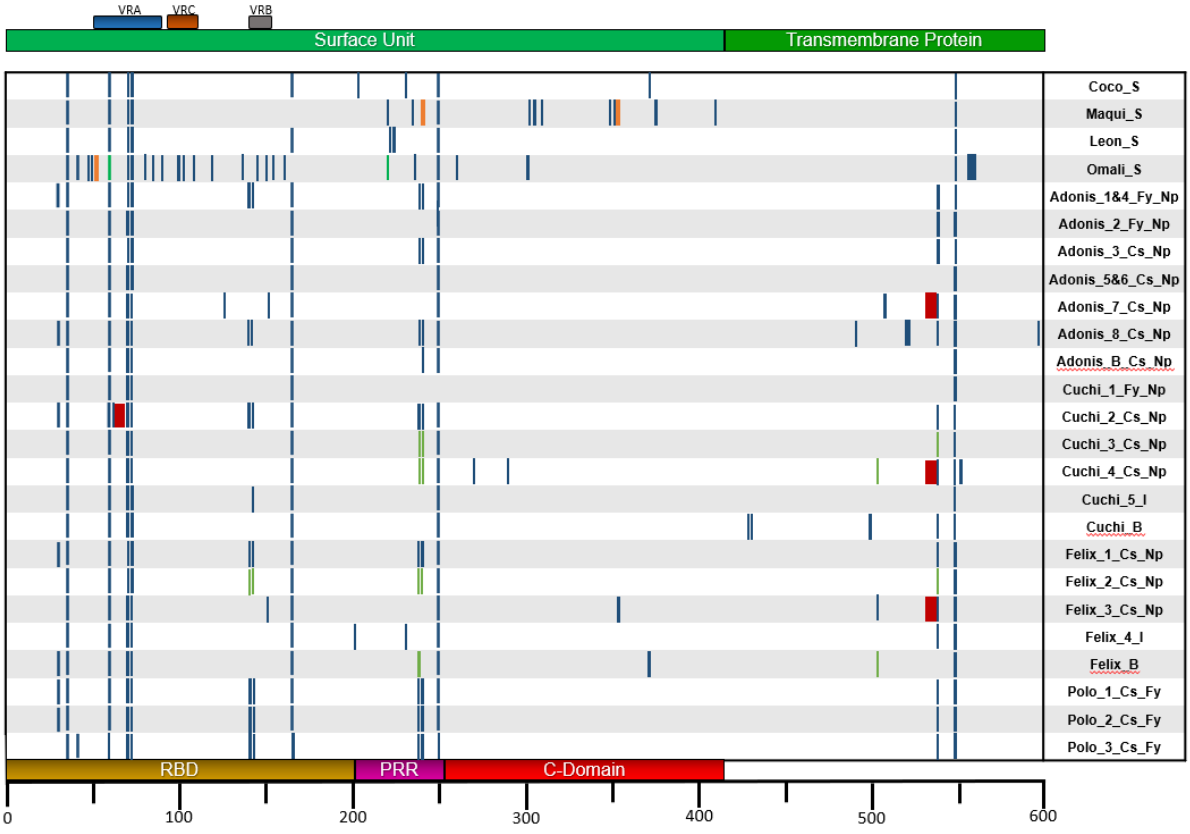


Figure 1. Scheme of alignment of envelope gene showing amino acid changes positions in FeLV-A sequences obtained from Sanger (S), Illumina (I), and Oxford Nanopore (Np) method. The coloured horizontal bars show the variable regions (VRA, VRC and VRB), the surface unit (light green), Transmembrane protein (green), Receptor binding domain (RBD) (yellow), Proline rich region (purple) and the C-Domain (red). Each blue vertical bar shows a single amino acid variation compared to the reference sequence, broader bar indicates two or more adjacent variation. Red vertical bars are deletions; Orange bars are insertion; Green bars alternate amino acid in the same position. Sequences are labelled by cat name, sequencing methods, assembly method, consensus (Cs) or De novo assembly in Flye (Fy).

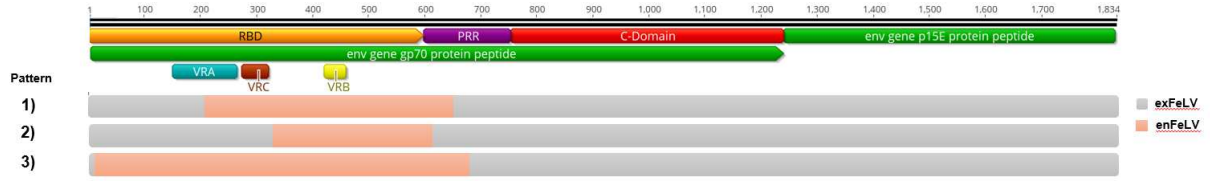
For the amplicons sequenced by Illumina, one long sequence per cat was assembled with multiple small contigs in different gene sections. Four long sequences between FeLV-A and FeLV-B were obtained and were considered in the alignment for further phylogenetic analysis. For Nanopore sequencing, two contigs per cat were obtained by the first approach (De Novo assembly using Flye). These contigs were around 1,900bp of length. Afterwards, using the second preset options in Minimap2, 13 consensus sequences were obtained (6

252 sequences from Adonis; 4 from Cuchi; 5 from Felix; 1 from Polo). These were between
253 1,764bp to 1,602bp of length, including insertions, deletions and missed nucleotides in the
254 sequence start.

255 The alignment of all these FeLV-A included 29 sequences belonging to 8 different cats. The
256 alignment shows a similarity between 100-94.8% at nucleotide level. Amino acid changes
257 were mainly identified in the Surface unit (Figure 1 and supplementary information). Seven
258 variations were present in almost all Chilean sequences, 5 variations showed 2 possible
259 amino acids (Supplementary table). Other variations were less often detected, these were
260 mainly present in Sanger sequences. The most distant sequence was a Sanger sequence
261 (Omali_S). Illumina amino acid changes were consistent with Nanopore results, except for
262 some minor changes in Felix_5_I in positions 202 and 230 (Supplementary information). The
263 main amino acid insertions were detected in Maqui (Sanger). Deletions were observed in the
264 SU region from nanopore sequences generated by consensus in Felix (5_Cs_Np), Cuchi
265 (4_Cs_Np) and Adonis (7_Cs_Np).

266 The FeLV-B amplicon in Illumina only produced one long sequence and multiple small
267 fragments concordant with FeLV-B or enFeLV. The longest sequence was obtained from
268 Cuchi.

269 From Nanopore FeLV-B amplicons, 17 sequences were generated, although 4 of them were
270 considered as FeLV-A. The other sequences were closer related to enFeLV. The longest
271 sequence was 1,764bp and the shortest was 1,474bp. These short sequences were missing
272 nucleotides at the beginning of the sequence (5' end). The percentage similarity was 99.5%
273 to 82.3%. The Illumina sequence was the most distant phylogenetically and some products
274 were highly homologue with FeLV-A.



Pattern detected	Number of events detected	Sequence Name
1	10	Omali_FeLV-B_S; Adonis_FeLV-B_2_Cs_Np; Adonis_FeLV-B_4_Cs_Np; Felix_FeLV-B_3_Cs_Np; Felix_FeLV-B_4_Cs_Np; Cuchi_FeLV-B_1_Cs_Np; Polo_FeLV-B_1_Cs_Np; Polo_FeLV-B_2_Cs_Np; Polo_FeLV-B_3_Cs_Np; Polo_FeLV-B_4_Cs_Np
2	2	Cuchi_FeLV-B_3_Cs_Np and Polo_FeLV-B_5_Cs_Np
3	1	Cuchi_FeLV-B_I

Figure 2. Scheme of recombinant patterns from RDP4 analysis. The reference sequence is the envelope gene from FeLV-A_FAIDS (M18247) sequence annotated in Geneious 2023.1.1. All sequences were considered in the analysis, but recombination was only detected in FeLV-B amplicons. Sequences from Sanger (S), Illumina (I), and Nanopore (Np) are shown. The gray bar color is sequencesrelated to exogenous FeLV (exFeLV) and orange is for endogenous FeLV (enFeLV). VRA, VRC and VRB are the variable region A, B and C. RBD is the receptor binding domain, PRR: Proline rich region. The table enumerates the pattern detected, and number of sequence where were detected the recombination and sequence name when were detected.

The recombination analysis performed in RDP4 showed three recombination breakpoint sites which were unique in the FeLV-B amplicons. All recombination breakpoints were derived from enFeLV genes with variable length inserted in the SU. The longest enFeLV insertion was in Cuchi_FeLV-B_I encompassed the whole RBD and partial PRR. While the second pattern was a smaller portion of enFeLV inserted in the end of RBD (Figure 2).

A total of 174 sequences were considered in the phylogenetical analysis, including all *env* complete sequences from GenBank (in supplementary information) and a second tree was constructed including 104 sequences (Figure 3), including phylogenetically closer sequences and representative clades. The tree is rooted in enFeLV, being subdivided into two major clades, enFeLV and FeLV-A. The Chilean FeLV-A clade is clearly identified next to the US (United States of America) sequences and the cats are homogenously represented through the Chilean clade. Sanger sequences are more related to each other than the rest of the NGS sequences. In contrast to the FeLV-A Chilean clade, the FeLV-B sequences are not closely related in a unique clade, where the biggest nanopore group are located next to sanger sequences and the Illumina sequence is closer to Gardner-Arnstein (V01172) with a third clade formed by Cuchi_FeLV-B_3_Cs_Np and Polo_FeLV-B_5_Cs_Np, closer to FeLV-A. This third clade also demonstrated the second recombination breakpoint pattern with a shorter enFeLV portion.

293

294 **Figure 3.** Phylogenetic tree of nucleotide env gene constructed with 5000 bootstrap approximation and rooted in
295 enFeLV (M25425). Eighty-four FeLV sequences were considered, including references sequences of FeLV-A,
296 FeLV-B, FeLV-C, FeLV-T and FeLV-E. Clades from the US and Japan have been collapsed for clarity (horizontal
297 triangles). The sequence name colour indicates the cat sampled (Cuchi in red; Polo in blue; Felix in green; and
298 Adonis in yellow). Sequences obtained from GenBank were called according to its reference number, type of
299 FeLV (if is known it), species (if was required), and country of origin. The sequences obtained in this study were
300 called according to name; FeLV amplicon obtained; number sequence per cat; sequence obtained by consensus
301 (Cs) or de novo assembly in Flye (Fy); and sequencing method, Illumina (I), Nanopore (Np) or Sanger (S).

302 iVar intra-host variation analysis was applied to all NGS results, comparing the two different
303 NGS methods (Cuchi and Felix) and comparing FeLV-A and B amplicons from nanopore
304 results (Figure 4). Differences between the methods and amplicons were observed
305 (Complete tabulate tables are in supplementary information). In general, the nanopore
306 results in FeLV-A amplicons detected around 400 variation events, but only a small portion
307 of these were statistically significant (less than hundred). Similarly, in FeLV-B above of 800
308 variations events were detected, but less than 200 on average were statistically significant.
309 This contrasted with the Illumina results, where less than 100 variations were observed and
310 half of them were considered as true. All Illumina variations were detected in low frequencies

and were located from the end of RBD and beginning of PRR. Nanopore results showed a longer distribution of variation through all the *env* gene from FeLV-A amplicons. The cat "Felix" FeLV-A amplicon showed high frequencies of variations in the RBD, but these were highly similar to the FeLV-B amplicon from the same cat. All FeLV-B amplicons showed a high number of variations in the first portion of RBD and a minor group of changes in TM (Figure 4).

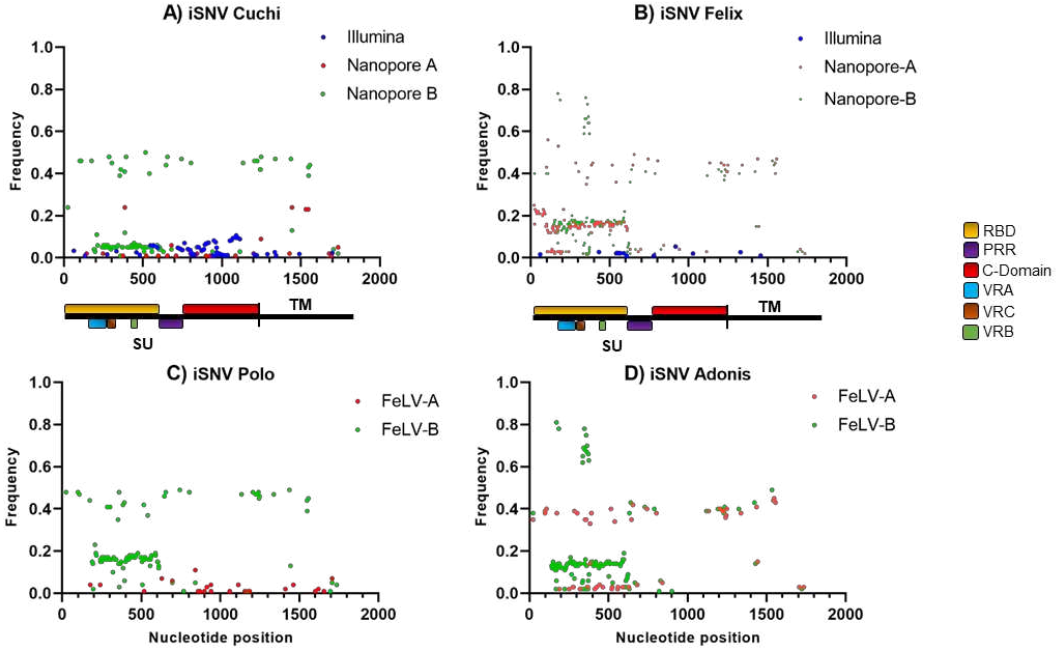


Figure 4. iSNV sites detected in env gene amplicons from NGS results. iVar software was used to generate a consensus sequences per library and to be used as reference sequence (~1,800bp). In the middle are schemes of FeLV-A (M18247) showing gene positions and main variation sites (RBD: Receptor Binding Domain; PRR: Proline Rich Region; VRA, VRC and VRB are variable region A, B, and C respectively. Each graph (A,B, C and D) represents the cat sampled and are compared frequency and nucleotide position classified by amplicon and sequencing method. The blue points correspond to variation points by the Illumina method (in A and B graph), the red points are variation points in FeLV-A amplicon and green points are FeLV-B amplicon variations (all graphs).

With regards to nucleotide variation (Figure 5) compared to the reference sequence (1765bp), its frequency is led by adenine (29.35%) and cytosine (27.20%), followed by thymine (21.53%) and guanine (21.93%). The most common variation in Cuchi and Polo (all libraries) were A→G and T→C. Similarly, Felix-I has higher variation to A→G and T→C, although nanopore results had a similar variation percentage between A→G and T→C with C→T and G→A. Adonis, in both nanopore libraries, showed an increased percentage of C→T and G→A mutations.

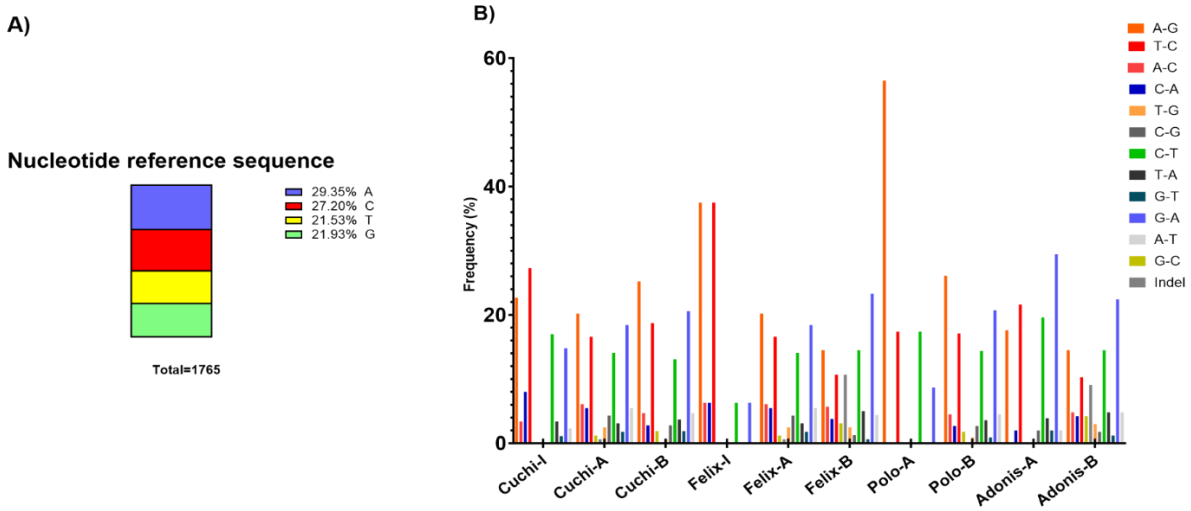


Figure 5. Nucleotide variation in sequences obtained from NGS sequencing. A) Graph of nucleotide frequencies in reference sequence. B) Bar graph showing nucleotide change frequency. Each colour represents one nucleotide change and there are grouped by cat and amplicon obtained. Nanopore results are pointed out the cat and from which amplicon was obtained, FeLV-A (A) or FeLV-B (B). Illumina results are represented with the cat name and "I".

DISCUSSION

Originally, FeLV variant characterization has been studied by viral interference assays in cell culture but genetic advances have facilitated the evaluation of specific genes and correlates with functional proteins (Chiu, Hoover, and Vandewoude 2018). The most common genetic analysis in FeLV has been based in the LTR due to its high conservation. Nonetheless, the importance of viral recombination has driven analysis of the envelope gene. Studies using molecular analysis based in *env* are available only in a few countries: Japan, the US (Erbeck et al. 2021), the UK (Stewart et al. 2013), Spain (Geret et al. 2011) and Brazil (Cano-Ortiz et al. 2022, Biezus et al. 2023). This is mainly because the work is laborious, with multiple steps such as end-point PCR or nested PCR, plasmid cloning and Sanger sequencing (Watanabe et al. 2013, Erbeck et al. 2021). Although NGS has had a rapid advancement, these techniques are still new in FeLV diagnosis, where the main problem is the high homology between enFeLV and exogenous FeLV (Chiu 2018). All our sequencing results were broadly consistent, however there were differences detected depending on the sequencing method. Here we documented the comparison of *env* gene amplicon-based sequencing using three methods and evidencing advantages and limitations of each technique.

Sanger sequencing has been the most widely used method in *env* analysis, it is highly reliable but only can generate only one single read per amplicon. To obtain clear reads in a mixed population (like a retroviral quasispecies), it is usually necessary to create plasmid

clones and sequence these individually (Watanabe et al. 2013, Erbeck et al. 2021). Our results were performed directly from the PCR amplicon because it was not the aim of this study and, as expected, most sequences did not show good quality. Only those with clear high-quality reads were considered in this methods comparison. Similarly, FeLV-B amplicons had worse quality due to the presence of FeLV-A and enFeLV in the same amplicon (as was confirmed by the NGS results). The main advantage of the Sanger method is the lower error rate, described as 0.001% (Hoff 2009), however added to this error rate should be considering the multiple steps, PCR amplifications, plasmid cloning and sequencing (Cheng, Fei, and Xiao 2023). Besides, studies based on HIV has shown that Sanger sequencing has a limited capacity to detect low-abundance intra-host variants when these are under a threshold of 20% of frequencies. This could mean missing information about less-representative but still clinically important sequences (Ávila-Ríos et al. 2020). Our FeLV-A Sanger sequences clustered closer together than sequences from the other two methods with only one sequence per cat. These sequences, however, had a higher number of SNPs and indels per sequence than contigs obtained by other methods, with the “Omali” Sanger sequence showing the highest number of amino acid variations. This likely reflects the consensus nature of the Sanger sequences, but also there was a high rate of failure for Sanger sequencing (only 5 of 20 sequences attempted showed good quality), likely reflecting the sequence diversity of the PCR amplicons, failing to produce a clean enough sequence to read.

Illumina short-read method is described as the most reliable method for detecting iSNVs, showing only slightly higher base-pair error than Sanger sequencing, between 0.26%-0.8% (van Dijk et al. 2014). However, it was not possible to assemble many long contigs from the Illumina short read data. In order to increase the number of contigs assembled, various bioinformatics modifications were tried, for instance reducing the filtering options in FastP and using SPAdes as an assembler, but this did not make a substantial difference to contig recovery. Only one Illumina FeLV-B sequence was therefore available for recombinant analysis with this sequence a different structure to those detected by the Sanger or nanopore methods. Due to a failure to generate long contigs, it was only possible evaluate SNVs using the iVar software comparing reads mapped to reference sequences of FeLV. Using iVar software a reduction in the minimum frequency threshold to include under-represented reads was required, because a high number of artifacts endogenous FeLV from the cat genome were detected as well.

Oxford Nanopore was a successful tool to discriminate between FeLV-B and FeLV-A and to determine the structure of the enFeLV and exFeLV recombinants in the FeLV-B PCRs, but it is a still developing tool and its error rate is described as between 5-15% (Rang,

Kloosterman, and de Ridder 2018). Although the successive MinION versions have been decreasing this error rate. At the iSNVs level, its results should be cautiously treated in order to minimize the error (Grubaugh et al. 2019). This is reflected in the sequences retrieved in this study by the number of sequences with X (uncertain amino acid) reads. The Nanopore long read method is however much better suited to detecting large indels in sequences and this is reflected in this study by its better success at retrieving and mapping FeLV-B recombinants. Similarly to the main FeLV-B recombination pattern previously described (Cano-Ortiz et al. 2022), the majority of sequences obtained included RBD and portion of PRR derived from enFeLV. Nonetheless, we used a conservative criterion to generate consensus sequences. Some were in fact exogenous reads with a longer portion derived from enFeLV, where the exogenous sequence had an endogenous insertion encompassing the RBD up to almost the whole C-Domain, but these were detected in a small number of reads therefore were not considered in further analysis.

Regarding amino acid and nucleotide variations, some of them were present in all Chilean reads when compared to the reference sequence (a US isolate from 1988) and are clearly features of Chilean isolates. Other iSNVs were present in only a small number of sequences at a low frequency and probably represent minor variants within the quasispecies mix. Interestingly many of these iSNVs were outside what are usually regarded as the more variable regions of the *env* gene. Different recombination patterns are also evident in the sequences. A comprehensive FeLV recombination review (Cano-Ortiz et al. 2022) showed multiple recombination patterns between exogenous and endogenous FeLV. In our data recombination's between enFeLV and exFeLV were frequent in the RBD with very few in the C-Domain. This is reflected in the lower number of SNPs in the C-Domain with most related to exFeLV variations (as few FeLV-B isolates contain enFeLV sequence in this domain. Our results confirm the previous data with the RBD (the most important region for binding to the receptor) the most variable site and the site of most frequent recombination with enFeLV. The C-Domain appears to be the most highly conserved region of the virus. This domain is known to play a role during viral infection as a secondary receptor, particularly for the FeLV-C strain, and appears to be particularly important in tropism for particular receptors/cell types (Boomer et al. 1997, Ramsey, Spibey, and Jarrett 1998, Gwynn et al. 2000, Rey, Prasad, and Tailor 2008). The lack of sequence variation in this study indicates that conservation in this domain is clearly critical to the virus.

The NGS results also showed that neither the PCR primers nor the mapping software completely discriminated between FeLV-A and FeLV-B sequences. FeLV-A sequences were generated from FeLV-B amplicons, and during the iSNV analysis, all sequences were mapped to a consensus generated from FeLV-A reads, however, in Adonis-A and Felix-A,

these showed a group of high frequency iSNVs closely related to FeLV-B. The pattern of nucleotide change (A→G, T→C, and A→T) is similar to that in a study of FeLV cats infected with tumors (Rohn et al. 1994). This type of mutation signature is commonly caused by the innate anti-viral APOBEC protein system, one of the main host responses in retrovirus infection (Stavrou and Ross 2015). These proteins generate hypermutated viral genomes with an increased number of G→A mutations to produce non-infective virions (Terry et al. 2017) and our results are consistent with this.

Each sequencing method produced different results even for the same animal and this needs to be considered when comparing large phylogenetic trees generated from amplicons derived from different sequencing methods. Each sequencing method has advantages and disadvantages, but phylogenies based on Sangar sequencing probably overestimate viral variation and may result in artificial sequence clustering when compared with the NGS methods. Moreover, other factors, like the input DNA amounts and concentrations used need to be considered in interpreting viral sequence variation. Our Illumina analysis was done using the minimum DNA requirement (around 5µg) and using 30x of coverage input, conversely, we used the maximum allowed DNA input for nanopore (around 250µg), this may have influenced the sequencing quality and under-represented sequences in Illumina as has been reported before (Illingworth et al. 2017) contrasting to the higher diversity detected in nanopore, although similarly to other reports, minority reads were masked and diluted by a greater number of sequences during consensus read construction (Mori et al. 2022).

Overall, the different methods were useful for different purposes with Sanger and Nanopore less reliable for iSNVs detection than Illumina, on the other hand, Nanopore was more successful for recombinant FeLV-B detection. With the growth in sequencing studies using different technologies, it will become increasingly important to consider which sequencing method was used in interpreting geographic and temporal variations in viral sequences, as well as designing and standardizing appropriate pipeline analysis to improve the quality of sequence, data interpretation and avoid sequence dilution, effects as has begun to be implemented in diagnostic HIV sequencing (Ávila-Ríos et al. 2020, Lee et al. 2020, Mori et al. 2022). Artificial clustering and estimation of sequencing variation due to sequencing artifact may complicate the interpretation of large viral phylogenies, particularly in viruses like FeLV where there is high virus circulation and recombination. Due to this being the first study conducted with NGS sequencing strategies for, there are no previous comparable data. Future works should include a greater number of animals and geographically distant FeLV, to analyse the viral diversity in domestic and non-domestic felids.

ACKNOWLEDGEMENTS

The first author gratefully acknowledges ANID (Agencia Nacional de Investigacion y Desarrollo, Chile) – Scholarship ID 72210211 for the support to this research.

REFERENCES

Ahmad, S. and Levy, L.S. (2010) 'The Frequency of Occurrence and Nature of Recombinant Feline Leukemia Viruses in the Induction of Multicentric Lymphoma by Infection of the Domestic Cat with FeLV-945'. *Virology* [online] 403 (2), 103–110. available from <http://dx.doi.org/10.1016/j.virol.2010.04.011>

Anai, Y., Ochi, H., Watanabe, S., Nakagawa, S., Kawamura, M., Gojobori, T., and Nishigaki, K. (2012) 'Infectious Endogenous Retroviruses in Cats and Emergence of Recombinant Viruses'. *Journal of Virology* 86 (16), 8634–8644

Ávila-Ríos, S., Parkin, N., Swanstrom, R., Paredes, R., Shafer, R., Ji, H., and Kantor, R. (2020) 'Next-Generation Sequencing for HIV Drug Resistance Testing: Laboratory, Clinical, and Implementation Considerations'. *Viruses* 12 (Viruses), 1–15

Bendinelli, M., Pistello, M., Lombardi, S., Poli, A., Garzelli, C., Matteucci, D., Ceccherini-Nelli, L., Malvaldi, G., and Tozzini, F. (1995) 'Feline Immunodeficiency Virus: An Interesting Model for AIDS Studies and an Important Cat Pathogen'. *Clinical Microbiology Reviews* 8 (1), 87–112

Biezus, G., Grima de Cristo, T., Bassi das Neves, G., da Silva Casa, M., Barros Brizola, P., Silvestre Sombrio, M., Miletti, L.C., and Assis Casagrande, R. (2023) 'Phylogenetic Identification of Feline Leukemia Virus A and B in Cats with Progressive Infection Developing into Lymphoma and Leukemia'. *Virus Research* [online] 329 (October 2022), 199093. available from <https://doi.org/10.1016/j.virusres.2023.199093>

Boni, M.F., Posada, D., and Feldman, M.W. (2007) 'An Exact Nonparametric Method for Inferring Mosaic Structure in Sequence Triplets'. *Genetics* 176 (2), 1035–1047

Boomer, S., Eiden, M., Burns, C.C., and Overbaugh, J. (1997) 'Three Distinct Envelope Domains, Variably Present in Subgroup B Feline Leukemia Virus Recombinants, Mediate Pit1 and Pit2 Receptor Recognition'. *Journal of Virology* 71 (11), 8116–8123

Brown, M.A., Cunningham, M.W., Roca, A.L., Troyer, J.L., Johnson, W.E., and Brien, S.J.O. (2008) 'Characterization of FeLV from Florida Panthers'. *Emerging Infectious Diseases* [online] 14 (2). available from <www.cdc.gov/eid>

Burling, A.N., Levy, J.K., Scott, H.M., Crandall, M.M., Tucker, S.J., Wood, E.G., and Foster, J.D. (2017) 'And Feline Immunodeficiency Virus Infection and Risk Factors for Seropositivity'. *Small Animal* 251 (2), 187–194

Cano-Ortiz, L., Tochetto, C., Roehe, P.M., Franco, A.C., and Junqueira, D.M. (2022) 'Could Phylogenetic Analysis Be Used for Feline Leukemia Virus (FeLV) Classification?' *Viruses* 14 (2)

Cattori, V., Tandon, R., Pepin, A., Lutz, H., and Hofmann-Lehmann, R. (2006) 'Rapid Detection of Feline Leukemia Virus Provirus Integration into Feline Genomic DNA'. *Molecular and Cellular Probes* 20 (3–4), 172–181

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018) 'Fastp: An Ultra-Fast All-in-One FASTQ Preprocessor'. *Bioinformatics* 34 (17), i884–i890

Cheng, C., Fei, Z., and Xiao, P. (2023) 'Methods to Improve the Accuracy of Next-Generation Sequencing'. *Frontiers in Bioengineering and Biotechnology* 11 (January), 1–13

Chiu, E.S., Hoover, E.A., and Vandewoude, S. (2018) 'A Retrospective Examination of Feline Leukemia Subgroup Characterization: Viral Interference Assays to Deep Sequencing'. *Viruses* 10 (1)

Chiu, E.S., Kraberger, S., Cunningham, M., Cusack, L., Roelke, M., and Vandewoude, S. (2019) 'Multiple Introductions of Domestic Cat Feline Leukemia Virus in Endangered Florida Panthers 1'. *Emerging Infectious Diseases* 25 (1), 92–101

Chiu, E.S. and VandeWoude, S. (2020) 'Presence of Endogenous Viral Elements Negatively Correlates with Feline Leukemia Virus Susceptibility in Puma And'. *Journal of Virology* 94 (21)

Coffin, J., Blomberg, J., Fan, H., Gifford, R., Hatziioannou, T., Lindemann, D., Mayer, J., Stoye, J., Tristem, M., and Johnson, W. (2021) 'ICTV Virus Taxonomy Profile: Retroviridae 2021'. *Journal of General Virology* 102 (12), 1–2

De Coster, W., D'Hert, S., Schultz, D.T., Cruts, M., and Van Broeckhoven, C. (2018) 'NanoPack: Visualizing and Processing Long-Read Sequencing Data'. *Bioinformatics* 34 (15), 2666–2669

Cunningham, M.W., Brown, M.A., Shindle, D.B., Terrell, S.P., Hayes, K.A., Ferree, B.C., McBride, R.T., Blankenship, E.L., Jansen, D., Citino, S.B., Roelke, M.E., Kiltie, R.A., Troyer, J.L., and O'Brien, S.J. (2008) 'Epizootiology and Management of Feline Leukemia Virus in the Florida Puma'. *Journal of Wildlife Diseases* 44 (3), 537–552

van Dijk, E.L., Auger, H., Jaszczyszyn, Y., and Thermes, C. (2014) 'Ten Years of Next-Generation Sequencing Technology'. *Trends in Genetics : TIG* 30 (9), 418–426

Edgar, R.C. (2004) 'MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput'. *Nucleic Acids Research* 32 (5), 1792–1797

Erbeck, K., Gagne, R.B., Kraberger, S., Chiu, E.S., Roelke-Parker, M., and VandeWoude, S. (2021) 'Feline Leukemia Virus (FeLV) Endogenous and Exogenous Recombination Events Result in

534 Multiple FeLV-B Subtypes during Natural Infection'. *Journal of Virology* 95 (18)

535 Faix, P.H., Feldman, S.A., Overbaugh, J., and Eiden, M. V. (2002) 'Host Range and Receptor Binding
536 Properties of Vectors Bearing Feline Leukemia Virus Subgroup B Envelopes Can Be Modulated
537 by Envelope Sequences Outside of the Receptor Binding Domain'. *Journal of Virology* 76 (23),
538 12369–12375

539 Geret, C.P., Cattori, V., Meli, M.L., Riond, B., Martínez, F., López, G., Vargas, A., Simón, M.A.,
540 López-Bao, J. V., Hofmann-Lehmann, R., and Lutz, H. (2011) 'Feline Leukemia Virus Outbreak
541 in the Critically Endangered Iberian Lynx (Lynx Pardinus): High-Throughput Sequencing of
542 Envelope Variable Region A and Experimental Transmission'. *Archives of Virology* 156 (5),
543 839–854

544 Gibbs, M.J., Armstrong, J.S., and Gibbs, A.J. (2000) 'Sister-Scanning: A Monte Carlo Procedure for
545 Assessing Signals in Rebombinant Sequences'. *Bioinformatics* 16 (7), 573–582

546 Grubaugh, N.D., Gangavarapu, K., Quick, J., Matteson, N.L., Jesus, J.G. De, Main, B.J., Tan, A.L.,
547 Paul, L.M., Brackney, D.E., Grewal, S., Gurfield, N., Rompay, K.K.A. Van, Isern, S., Michael,
548 S.F., Coffey, L.L., Loman, N.J., and Andersen, K.G. (2019) *An Amplicon-Based Sequencing*
549 *Framework for Accurately Measuring Intrahost Virus Diversity Using PrimalSeq and IVar*. 1–
550 19

551 Gwynn, S.R., Hankenson, F.C., Lauring, A.S., Rohn, J.L., and Overbaugh, J. (2000) 'Feline Leukemia
552 Virus Envelope Sequences That Affect T-Cell Tropism and Syncytium Formation Are Not Part
553 of Known Receptor-Binding Domains'. *Journal of Virology* 74 (13), 5754–5761

554 Hardy, W.D., MacEwen, E.G., McClelland, A.J., Zuckerman, E.E., Myron, M., and Essex, E. (1976)
555 'Biology of Feline Leukemia Virus in the Natural Environment'. *Cancer Research* 36
556 (February), 582–588

557 Hartmann, K. (2012) 'Feline Leukemia Virus Infection'. in *Infectious Diseases of the Dog and Cat*.
558 Fourth Edi. ed. by Greene, C.E. St Louis, Missouri: Elsevier, 108–136

559 Hartmann, K. and Hofmann-Lehmann, R. (2020) 'What's New in Feline Leukemia Virus Infection'.
560 *Veterinary Clinics of North America - Small Animal Practice* 50 (5), 1013–1036

561 Hoang, D.T., Chernomor, O., Von Haeseler, A., Minh, B.Q., and Vinh, L.S. (2018) 'UFBoot2:
562 Improving the Ultrafast Bootstrap Approximation'. *Molecular Biology and Evolution* 35 (2),
563 518–522

564 Hoff, K.J. (2009) 'The Effect of Sequencing Errors on Metagenomic Gene Prediction'. *BMC*
565 *Genomics* 10, 1–9

566 Illingworth, C.J.R., Roy, S., Beale, M.A., Tutill, H., Williams, R., and Breuer, J. (2017) 'On the

Effective Depth of Viral Sequence Data'. *Virus Evolution* 3 (2), 1–9

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., and Jermiin, L.S. (2017) 'ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates'. *Nature Methods* [online] 14, 587–589. available from <https://doi.org/10.1038/nmeth.4285>

Krunic, M., Ertl, R., Hagen, B., Sedlazeck, F.J., Hofmann-Lehmann, R., von Haeseler, A., and Klein, D. (2015) 'Decreased Expression of Endogenous Feline Leukemia Virus in Cat Lymphomas: A Case Control Study'. *BMC Veterinary Research* [online] 11 (1), 1–7. available from <???>

Lacerda, L.C., Silva, A.N., Freitas, J.S., Cruz, R.D.S., Said, R.A., and Munhoz, A.D. (2017) 'Feline Immunodeficiency Virus and Feline Leukemia Virus: Frequency and Associated Factors in Cats in Northeastern Brazil'. *Genetics and Molecular Research* 16 (2)

Langmead, B. and Salzberg, S.L. (2012) 'Fast Gapped-Read Alignment with Bowtie 2'. *Nature Methods* 9 (4), 357–359

Lee, E.R., Parkin, N., Jennings, C., Brumme, C.J., Enns, E., Casadellà, M., Howison, M., Coetzer, M., Avila-Rios, S., Capina, R., Marinier, E., Van Domselaar, G., Noguera-Julian, M., Kirkby, D., Knaggs, J., Harrigan, R., Quiñones-Mateu, M., Paredes, R., Kantor, R., Sandstrom, P., and Ji, H. (2020) 'Performance Comparison of next Generation Sequencing Analysis Pipelines for HIV-1 Drug Resistance Testing'. *Scientific Reports* 10 (1), 1–10

Li, D., Liu, C.M., Luo, R., Sadakane, K., and Lam, T.W. (2015) 'MEGAHIT: An Ultra-Fast Single-Node Solution for Large and Complex Metagenomics Assembly via Succinct de Bruijn Graph'. *Bioinformatics* 31 (10), 1674–1676

Li, H. (2018) 'Minimap2: Pairwise Alignment for Nucleotide Sequences'. *Bioinformatics* 34 (18), 3094–3100

Li, H. (2013) *Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM*. [online] 00 (00), 1–3. available from <http://arxiv.org/abs/1303.3997>

Lin, Y., Yuan, J., Kolmogorov, M., Shen, M.W., Chaisson, M., and Pevzner, P.A. (2016) 'Assembly of Long Error-Prone Reads Using de Bruijn Graphs'. *Proceedings of the National Academy of Sciences of the United States of America* 113 (52), E8396–E8405

Little, S., Levy, J., Hartmann, K., Hofmann-Lehmann, R., Hosie, M., Olah, G., and Denis, K.S. (2020) '2020 AAFP Feline Retrovirus Testing and Management Guidelines'. *Journal of Feline Medicine and Surgery* 22 (1), 5–30

Ludwick, K. and Clymer, J.W. (2019) 'Comparative Meta-Analysis of Feline Leukemia Virus and Feline Immunodeficiency Virus Seroprevalence Correlated with GDP per Capita around the Globe'. *Research in Veterinary Science* 125, 89–93

600 Martin, D. and Rybicki, E. (2000) 'A Modified Bootscan Algorithm for Automated Identification of
601 Recombinant Sequences and Recombination Breakpoints'. *Bioinformatics* [online] 16 (6), 562–
602 563. available from <https://www.liebertpub.com/doi/10.1089/aid.2005.21.98?url_ver=Z39.88-
603 2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub 0pubmed>

604 Martin, D.P., Murrell, B., Golden, M., Khoosal, A., and Muhire, B. (2015) 'RDP4: Detection and
605 Analysis of Recombination Patterns in Virus Genomes'. *Virus Evolution* 1 (1), 1–5

606 Meli, M., Cattori, V., Martınez, F., López, G., Vargas, A., Palomares, Francisco López-Bao, J. V.,
607 Hofmann-lehmann, R., and Lutz, H. (2010a) *Feline Leukemia Virus Infection : A Threat for the*
608 *Survival of the Critically Endangered Iberian Lynx (Lynx Pardinus)*. 134, 61–67

609 Meli, M., Simmler, P., Cattori, V., Martínez, F., Vargas, A., Palomares, F., López-Bao, J. V., Simón,
610 M.A., López, G., León-Vizcaino, L., Hofmann-Lehmann, R., and Lutz, H. (2010b) 'Importance
611 of Canine Distemper Virus (CDV) Infection in Free-Ranging Iberian Lynxes (Lynx Pardinus)'.
612 *Veterinary Microbiology* 146 (1–2), 132–137

613 Miyake, A., Watanabe, S., Hiratsuka, T., Ito, J., Ngo, M.H., Makundi, I., Kawasaki, J., Endo, Y.,
614 Tsujimoto, H., and Nishigaki, K. (2016) ' Novel Feline Leukemia Virus Interference Group
615 Based on the Env Gene '. *Journal of Virology* 90 (9), 4832–4837

616 Mora, M., Napolitano, C., Ortega, R., Poulin, E., and Pizarro-Lucero, J. (2015) 'Feline
617 Immunodeficiency Virus and Feline Leukemia Virus Infection in Free-Ranging Guignas
618 (Leopardus Guigna) and Sympatric Domestic Cats in Human Perturbed Landscapes on ChiloÉ
619 Island, Chile'. *Journal of Wildlife Diseases* 51 (1), 199–208

620 Mori, M., Ode, H., Kubota, M., Nakata, Y., Kasahara, T., Shigemi, U., Okazaki, R., Matsuda, M.,
621 Matsuoka, K., Sugimoto, A., Hachiya, A., Imahashi, M., Yokomaku, Y., and Iwatani, Y. (2022)
622 'Nanopore Sequencing for Characterization of HIV-1 Recombinant Forms'. *Microbiology*
623 *Spectrum* 10 (4), 1–12

624 Ortega, C., Valencia, A.C., Duque-Valencia, J., and Ruiz-Saenz, J. (2020) 'Prevalence and Genomic
625 Diversity of Feline Leukemia Virus in Privately Owned and Shelter Cats in Aburrá Valley,
626 Colombia'. *Viruses* 12 (4), 1–13

627 Overbaugh, J. (2000) 'Identification of a Cellular Cofactor Required for Infection by Feline Leukemia
628 Virus'. *Science* 287 (5459), 1828–1830

629 Padidam, M., Sawyer, S., and Fauquet, C.M. (1999) 'Possible Emergence of New Geminiviruses by
630 Frequent Recombination'. *Virology* 265 (2), 218–225

631 Perelson, A.S., Neumann, A.U., Markowitz, M., Leonard, J.M., and Ho, D.D. (1996) 'HIV-1
632 Dynamics in Vivo: Virion Clearance Rate, Infected Cell Life-Span, and Viral Generation Time'.

633 *Science* 271 (5255), 1582–1586

634 Petch, R.J., Gagne, R.B., Chiu, E., Mankowski, C., Rudd, J., Roelke-Parker, M., Vickers, T.W.,
635 Logan, K.A., Alldredge, M., Clifford, D., Cunningham, M.W., Onorato, D., and VandeWoude,
636 S. (2022) 'Feline Leukemia Virus Frequently Spills Over from Domestic Cats to North
637 American Pumas'. *Journal of Virology*

638 Polani, S., Roca, A.L., Rosensteel, B.B., Kolokotronis, S.O., and Bar-Gal, G.K. (2010) 'Evolutionary
639 Dynamics of Endogenous Feline Leukemia Virus Proliferation among Species of the Domestic
640 Cat Lineage'. *Virology* [online] 405 (2), 397–407. available from
641 <http://dx.doi.org/10.1016/j.virol.2010.06.010>

642 Posada, D. and Crandall, K.A. (2001) 'Evaluation of Methods for Detecting Recombination from
643 DNA Sequences: Computer Simulations'. *Proceedings of the National Academy of Sciences of*
644 *the United States of America* 98 (24), 13757–13762

645 Powers, J.A., Chiu, E.S., Kraberger, S.J., Roelke-parker, M., Lowery, I., Erbeck, K., Troyer, R.,
646 Carver, S., and VandeWoude, S. (2018) 'Crossm Feline Leukemia Virus (FeLV) Disease
647 Outcomes in a Domestic Cat Breeding Colony : Relationship to Endogenous FeLV and Other
648 Chronic Viral Infections'. *Journal of Virology* 92 (18), 1–16

649 Rambaut, A. (2018) *FigTree v1.4.4*. available from <http://tree.bio.ed.ac.uk/software/figtree/>

650 Ramsey, I.K., Spibey, N., and Jarrett, O. (1998) 'The Receptor Binding Site of Feline Leukemia Virus
651 Surface Glycoprotein Is Distinct from the Site Involved in Virus Neutralization'. *Journal of*
652 *Virology* 72 (4), 3268–3277

653 Rang, F.J., Kloosterman, W.P., and de Ridder, J. (2018) 'From Squiggle to Basepair: Computational
654 Approaches for Improving Nanopore Sequencing Read Accuracy'. *Genome Biology* 19 (1), 1–
655 11

656 Rey, M.A., Prasad, R., and Tailor, C.S. (2008) 'The C Domain in the Surface Envelope Glycoprotein
657 of Subgroup C Feline Leukemia Virus Is a Second Receptor-Binding Domain'. *Virology* 370
658 (2), 273–284

659 Riedel, N., Hoover, E.A., Gasper, P.W., Nicolson, M.O., and Mullins, J.I. (1986) 'Molecular
660 Analysis'. *Journal of Virology* 60 (1), 242–250

661 Roca, A.L., Nash, W.G., Menninger, J.C., Murphy, W.J., and O'Brien, S.J. (2005) 'Insertional
662 Polymorphisms of Endogenous Feline Leukemia Viruses'. *Journal of Virology* 79 (7), 3979–
663 3986

664 Rohn, J.L., Linenberger, M.L., Hoover, E.A., and Overbaugh, J. (1994) 'Evolution of Feline
665 Leukemia Virus Variant Genomes with Insertions, Deletions, and Defective Envelope Genes in

Infected Cats with Tumors'. *Journal of Virology* 68 (4), 2458–2467

Sacristán, I., Acuña, F., Aguilar, E., García, S., José López, M., Cabello, J., Hidalgo-Hermoso, E., Sanderson, J., Terio, K.A., Barrs, V., Beatty, J., Johnson, W.E., Millán, J., Poulin, E., and Napolitano, C. (2021) 'Cross-Species Transmission of Retroviruses among Domestic and Wild Felids in Human-Occupied Landscapes in Chile'. *Evolutionary Applications* 14 (4), 1070–1082

Smith, J.M. (1992) 'Analyzing the Mosaic Structure of Genes'. *Journal of Molecular Evolution* 34, 126–129

Stavrou, S. and Ross, S.R. (2015) 'APOBEC3 Proteins in Viral Immunity'. *The Journal of Immunology* 195 (10), 4565–4570

Stewart, H., Jarrett, O., Hosie, M.J., and Willett, B.J. (2013) 'Complete Genome Sequences of Two Feline Leukemia Virus Subgroup B Isolates with Novel Recombination Sites'. *Genome Announcements* 1 (1), 1–2

Studer, N., Lutz, H., Saegerman, C., Gönczi, E., Meli, M.L., Boo, G., Hartmann, K., Hosie, M.J., Moestl, K., Tasker, S., Belák, S., Lloret, A., Boucraut-Baralon, C., Egberink, H.F., Pennisi, M.-G., Truyen, U., Frymus, T., Thiry, E., Marsilio, F., Addie, D., Hochleithner, M., Tkalec, F., Vizi, Z., Brunetti, A., Georgiev, B., Ludwig-Begall, L.F., Tschuor, F., Mooney, C.T., Eliasson, C., Orro, J., Johansen, H., Juuti, K., Krampl, I., Kovalenko, K., Šengaut, J., Sobral, C., Borska, P., Kovaříková, S., and Hofmann-Lehmann, R. (2019) 'Pan-European Study on the Prevalence of the Feline Leukaemia Virus Infection-Reported by the European Advisory Board on Cat Diseases (ABCD Europe)'. *Viruses* [online] 11, 993. available from <www.mdpi.com/journal/viruses>

Teo, C.H.Y., Norhisham, N.H.B., Lee, O.F., Png, S., Chai, C.N., Yan, G., Tang, J.W.T., and Lee, C.K. (2022) 'Towards Next-Generation Sequencing for HIV-1 Drug Resistance Testing in a Clinical Setting'. *Viruses* 14 (10), 1–16

Terry, A., Kilbey, A., Naseer, A., Levy, L.S., Ahmad, S., Watts, C., Mackay, N., Cameron, E., Wilson, S., and Neil, J.C. (2017) 'Barriers to Infection of Human Cells by Feline Leukemia Virus: Insights into Resistance to Zoonosis'. *Journal of Virology* 91 (5)

Torres, A.N., Mathiason, C.K., and Hoover, E.A. (2005) 'Re-Examination of Feline Leukemia Virus: Host Relationships Using Real-Time PCR'. *Virology* 332 (1), 272–283

Trifinopoulos, J., Nguyen, L.T., von Haeseler, A., and Minh, B.Q. (2016) 'W-IQ-TREE: A Fast Online Phylogenetic Tool for Maximum Likelihood Analysis'. *Nucleic Acids Research* 44 (W1), W232–W235

Watanabe, S., Kawamura, M., Odahara, Y., Anai, Y., Ochi, H., Nakagawa, S., Endo, Y., Tsujimoto,

H., and Nishigaki, K. (2013) 'Phylogenetic and Structural Diversity in the Feline'. *PLoS ONE* 8 (4), 1–16

Wick, R.R., Judd, L.M., Gorrie, C.L., and Holt, K.E. (2017) 'Completing Bacterial Genome Assemblies with Multiplex MinION Sequencing'. *Microbial Genomics* 3 (10), 1–7

Willett, B.J. and Hosie, M.J. (2013) 'Feline Leukaemia Virus: Half a Century since Its Discovery'. *Veterinary Journal* [online] 195 (1), 16–23. available from <http://dx.doi.org/10.1016/j.tvjl.2012.07.004>