JGV Papers in Press. Published December 22, 2014 as doi:10.1099/vir.0.000035

Journal of General Virology

Rapid evolution of the env gene leader sequence in cats naturally infected with feline immunodeficiency virus (FIV)

--Manuscript Draft--

Manuscript Number:	JGV-D-14-00142R1
Full Title:	Rapid evolution of the env gene leader sequence in cats naturally infected with feline immunodeficiency virus (FIV)
Short Title:	FIV evolution
Article Type:	Standard
Section/Category:	Animal - Retroviruses
Corresponding Author:	Pawel M Beczkowski University of Glasgow Centre for Virus Research Glasgow, UNITED KINGDOM
First Author:	Pawel M Beczkowski
Order of Authors:	Pawel M Beczkowski
	Joseph Hughes
	Roman Biek
	Annette Litster
	Brian J Willett
	Margaret J Hosie
Abstract:	Analysing the evolution of FIV on the intra-host level is important, in order to address whether the diversity and composition of viral quasispecies affects disease progression. We examined the intra-host diversity and the evolutionary rates of the entire env and structural fragments of the env sequences obtained from sequential blood samples in 43 naturally infected domestic cats that displayed different clinical outcomes. We observed in the majority of cats that FIV env showed very low levels of intra-host diversity. We estimated that env evolved at the rate of 1.16 x 10-3 substitutions per site per year and demonstrated that recombinant sequences evolved faster than non-recombinant sequences. It was evident that the V3-V5 fragment of FIV env displayed higher evolutionary rates in healthy cats than in those with terminal illness. Our study provided the first evidence that the leader sequence of env, rather than the V3-V5 sequence, had the highest intra-host diversity and the highest evolutionary rate of all env fragments, consistent with this region being under a strong selective pressure for genetic variation. Overall, FIV env displayed relatively low intra-host diversity and evolved slowly in naturally infected cats. The maximal evolutionary rate was observed in the leader sequence of env. Although genetic stability is not necessarily a prerequisite for clinical stability, the higher genetic stability of FIV compared to HIV might explain why many naturally infected cats do not progress to AIDS rapidly.

Rapid evolution of the *env* gene leader sequence in cats naturally infected with feline immunodeficiency virus (FIV)

Running title: FIV evolution (standard paper)

Paweł M Bęczkowski[1,2§], Joseph Hughes[1], Roman Biek[1,3], Annette Litster[4], Brian J Willett [1], Margaret J Hosie[1]

[1]MRC University of Glasgow Centre for Virus Research, University of Glasgow, Glasgow, United Kingdom

[2]Small Animal Hospital, University of Glasgow, Glasgow, United Kingdom

[3]Boyd Orr Centre for Population and Ecosystem Health & Institute of Biodiversity, Animal Health & Comparative Medicine, University of Glasgow, Glasgow, United Kingdom

[4]Department of Veterinary Clinical Sciences, Purdue University, West Lafayette, IN 47907, USA

[§]Corresponding author: Paweł M Bęczkowski

Email: pawel.beczkowski@glasgow.ac.uk

Tel: +44 0141 330 6068

Word count: Word count: Abstract: 233, Text: 4721, Tables: 1, Figures: 5, Additional files: 8

The GenBank accession numbers for the novel FIV *env* sequences generated in this study are: KP264257 - KP264562

Keywords: FIV, Quasispecies, Evolutionary rate, BEAST, Intra-host diversity, Leader, V3-V5, Natural infection.

Abstract

Analysing the evolution of FIV on the intra-host level is important, in order to address whether the diversity and composition of viral quasispecies affects disease progression.

We examined the intra-host diversity and the evolutionary rates of the entire *env* and structural fragments of the *env* sequences obtained from sequential blood samples in 43 naturally infected domestic cats that displayed different clinical outcomes. We observed in the majority of cats that FIV *env* showed very low levels of intra-host diversity. We estimated that *env* evolved at the rate of 1.16×10^{-3} substitutions per site per year and demonstrated that recombinant sequences evolved faster than non-recombinant sequences. It was evident that the V3-V5 fragment of FIV *env* displayed higher evolutionary rates in healthy cats than in those with terminal illness. Our study provided the first evidence that the leader sequence of *env*, rather than the V3-V5 sequence, had the highest intra-host diversity and the highest evolutionary rate of all *env* fragments, consistent with this region being under a strong selective pressure for genetic variation.

Overall, FIV *env* displayed relatively low intra-host diversity and evolved slowly in naturally infected cats. The maximal evolutionary rate was observed in the leader sequence of *env.* Although genetic stability is not necessarily a prerequisite for clinical stability, the higher genetic stability of FIV compared to HIV might explain why many naturally infected cats do not progress to AIDS rapidly.

Introduction

Deciphering the evolution of feline immunodeficiency virus (FIV) is essential for understanding why some infected cats have normal lifespans while others progress rapidly to AIDS. The generation of polymorphic populations (Eigen, 1993) and the potentially high intra-host diversity of RNA viruses are features attributable to their biology as well as the unique characteristics of retroviral reverse transcription (Domingo, 2000). Improved understanding of the dynamics of viral populations on the intra-host level will lead to the development of more effective diagnostic tests, control strategies and, ultimately, will inform the design of the next generation of FIV vaccines, since the limited efficacy of the current commercial vaccine (Dunham *et al.*, 2006) is likely attributable to the high genetic diversity of naturally occurring viruses. The majority of FIV phylogenetic analyses have focused on sequence variation at the population level (Carpenter *et al.*, 1998; Olmsted *et al.*, 1992; Pistello *et al.*, 1997; Samman *et al.*, 2011; Teixeira *et al.*, 2010) and highlighted the high variability of the V3-V5 region of the *env* gene, which has been used to classify FIV into 6 distinct subtypes (A, B, C, D, E and putative subtype F) (Marcola *et al.*, 2013; Sodora *et al.*, 1994). Employing GARD (Kosakovsky Pond *et al.*, 2006a) and jpHMM (Schultz *et al.*, 2006) recombination detection methods, we recently observed an abundance of recombinant *envs* in natural FIV infection, emphasizing the important role of recombination in generating viral diversity at the population level and highlighting the limitations of the current phylogenetic classification of FIV (Bęczkowski *et al.*, 2014a).

Mutations also contribute to viral diversity, although little is known about the evolutionary rate and the within-host viral diversity of the entire FIV *env* in natural infection (Table 1). In HIV-1 infected individuals, quasispecies composition varies

71 greatly, with some studies reporting 10% diversity in long-term infected individuals
72 (Delwart *et al.*, 1997). It has also been proposed that intra-host diversity could
73 influence the outcome of disease, and that the rate of viral evolution decreases during
74 disease progression (Delwart *et al.*, 1997; Shankarappa *et al.*, 1999).

75 Selection measures have been applied previously to assess forces acting on the
76 evolution of different retroviral genes and at different regions within the same gene
77 (Choisy *et al.*, 2004). For example, in HIV-1 infection, *gag* is under negative selection
78 pressure (Kils-Hutten *et al.*, 2001), while *env* evolution is generally shaped by
79 positive selection pressure (Wolfs *et al.*, 1990). Since *gag* encodes structural proteins,
80 its conserved nature is essential to maintain viral integrity. In contrast, HIV-1 *env*
81 encodes a heavily glycosylated envelope glycoprotein in which mutations in N-linked
82 glycosylation sites are important for immune evasion (Yamaguchi & Gojobori, 1997).
83 By analogy with HIV, we predict that FIV *env* is under a similar positive selection
84 pressure as its human counterpart, but relatively little is known about the
85 glycosylation pattern or selection forces acting on FIV Env in naturally infected cats.

86 To our knowledge, there are no data available concerning rates of evolution of the
87 entire ORF of FIV *env* which, in contrast to primate retroviruses, contains an
88 unusually long leader/signal region (Verschoor *et al.*, 1993). Evolutionary studies of
89 HIV-1 infection (Korber *et al.*, 2000; Leitner & Albert, 1999; Lukashov *et al.*, 1995;
90 Shankarappa *et al.*, 1999; Zhang *et al.*, 1997) suggest that within-host evolution is
91 influenced by both viral- and host-dependent factors. Whereas some studies reported
92 increased evolutionary rates during the early stages of infection, declining towards
93 terminal stage disease (Delwart *et al.*, 1997; Shankarappa *et al.*, 1999), others did not
94 observe a similar pattern (Mikhail *et al.*, 2005). Nevertheless, the assumption that the
95 virus is under constant pressure in immunocompetent individuals and evolves at a

high rate in response to a changing environment, provides a plausible explanation for differences in evolutionary rates at different stages of disease (Lukashov *et al.*, 1995). It is apparent that rates of evolution are dependent on the combined effects of host and viral factors, demonstrated by comparisons of the evolutionary tempo of different HIV-1 subtypes (Abecasis *et al.*, 2009) and differences in replication rates of HIV-1 and HIV-2 (MacNeil *et al.*, 2007b).

Our ability to understand immune evasion, and to predict the outcome of FIV infection, depends on understanding how the intra-host evolution of FIV differs within and among cats. For this reason, we examined *env* sequences collected serially over 12 months from 2 distinct cohorts of naturally infected cats that developed different clinical outcomes. Our specific aims were to: 1) quantify viral diversity, selection, and rates of evolution in naturally infected cats and 2) determine whether the evolutionary rate correlated with the clinical outcomes of FIV infection.

Results

Phylogenetic inference

The Maximum Likelihood (ML) tree constructed using the entire data set (Additional file 1) revealed high genetic diversity in FIV *env* at the population level (overall mean pairwise distance of 14.1%). Although more than 41% of cats were infected with circulating recombinant viruses, we found no evidence of within host recombination (Bęczkowski *et al.*, 2014a). On the intra-host level, sequential sequences amplified from 93% (40/43) of cats clustered together in monophyletic groups, while sequences from 3 cats (M5, P8 and P21), amplified at different time points, showed incongruent phylogenetic assignment (Fig. 1). Sequences from a parent-offspring pair (cat M1 was an offspring of female M11) were classified as non-recombinant clade B and clustered closely together, with an overall mean pairwise distance of 0.3% (Additional file 2-Fig. S1).

Intra-host diversity

Having excluded sequences from the 3 cats that formed non-monophyletic groups, the mean pairwise distances for *env* within cats ranged from 0% to 2.9% (median 0.13), 0% to 1.9% (median 0.11) and 0% to 1.9% (median 0.17), for early (A), intermediate (B) and late (C) time points, respectively (Additional file 3-Table S1). When we compared the diversity of structural fragments of the *env*, although not statistically significant (p=0.29), the leader region displayed the highest variation at the intra-host level; the median overall mean pairwise distances for each time point were 0.2%, 0.2% and 0.27%, respectively (Fig. 2; Additional file 3-Table S1).

Selection

The mean ω value (dN/dS ratio) determined by SLAC for the *env* sequences from both cohorts was 0.34. Examining the entire data set, the SLAC analysis indicated 14 (1.6%) positively and 201 (23%) negatively selected sites. FEL reported 22 (2.5%)

positively and 391 (44.8%) negatively selected sites, while IFEL indicated 24 (2.7%) and 377 (43.2%) sites under positive and negative selection respectively. Positively selected sites, consistently identified by the three methods (Additional file 4-Table S2), were restricted to specific regions of *env*, with a high proportion (5/10, 50%; sites 8, 62, 96, 156, 157) lying within the leader region (Fig. 3). The selection hot spots at positions 400 and 750 were located within B cell epitopes identified in previous studies (Lombardi *et al.*, 1993; Pancino *et al.*, 1993) (Fig. 3). An investigation of the global distribution of potential N-linked glycosylation sites (PNGS) for the presence of fixed and shifting sequons revealed that, regardless of the high genetic variability of the analysed sequences (n=329), conserved PNGS did exist (Fig. 3). However, none of the predicted PNGS fell within sites identified as being under positive selection.

Rates of molecular evolution

For all cats in the study, *env* was estimated to evolve at a mean rate of 1.16×10^{-3} with 0.7-1.67 95% HPD substitutions per site per year under a relaxed log normal clock model (Additional file 5 - Table S3) with the first and second codon positions evolving at a lower rate than the third codon position (0.95 with 0.90-0.98 95% HPD versus 1.1 with 1.02-1.19 95% HPD). To determine whether specific regions of the *env* gene evolved at different rates, we estimated the clock rate for the leader, the V3-V5 and the non-variable (NV) regions. The leader region evolved at the highest rate (3.43×10^{-3} with 1.55-5.6 95% HPD substitutions per site per year under a relaxed log normal clock model) in comparison to other fragments, while the evolutionary rate of the V3-V5 region, commonly regarded as highly variable, was over 3 times lower (1.08×10^{-3} with 0.5-1.74, 95% HPD substitutions per site per year under a relaxed log normal clock model) (Fig. 4; Additional file 5 -Table S3).

Since the individually housed cats from the Chicago cohort were generally healthy, while the majority of group housed cats from the Memphis cohort had high morbidity and 63% mortality over the study period, we compared the evolutionary rates between and within the cohorts. Median rates of evolution of the entire *env* and its fragments were higher for the sequences from cats in the Chicago cohort (Fig. 4). We hypothesised that *env* evolution in healthy cats that remained alive throughout the study period would be faster than in terminally ill cats that died during the study period. Fig. 5 illustrates a comparison of the evolutionary rates for sequences from i) 11 deceased cats and ii) 6 cats that remained alive and generally free of clinical signs during the study. The entire *env*, leader and NV fragment evolved at higher rates in terminally ill animals. This relationship, however, was reversed for the V3-V5 region, which evolved at a higher rate in cats that remained alive during the study period. Recombinant *env* sequences (n=123) evolved 3 times faster than sequences without any detectable prior history of recombination (n=184), (Additional file 5-Table S3).

Discussion

Analyses of serial *env* sequences collected over 12 months from naturally infected cats offered a rare opportunity to examine the intra-host dynamics and evolution of FIV. Despite considerable variation among individuals, the overall intra-host diversity of *env* (up to 3%) was low compared to primate immunodeficiency viruses (commonly >5%) (Rambaut *et al.*, 2004; Salemi, 2013). Furthermore, the overall mean distance between sequences obtained from 2 closely related animals (M1 was the offspring of M11), following at least 1 postulated transmission event, was remarkably low. These results suggest that the *env* sequences were stably maintained and evolved slowly following vertical transmission. This is consistent with an earlier study examining V3-V5 *env* sequence diversity of the Aomori-2 strain of FIV following vertical transmission; 100% homology was observed between viruses isolated from a queen and her kitten 48 weeks post transmission (Motokawa *et al.*, 2005).

The apparent discrepancy between low within-host viral diversity and relatively high diversity of FIV at the population level is intriguing, especially in light of differences in evolutionary rates demonstrated at the within- and epidemiological levels in HIV infection (Alizon & Fraser, 2013; Lythgoe & Fraser, 2012). Analysis of FIV evolutionary rates at the epidemiological level are constrained by the low number of longitudinal full length FIV *env* sequences that are available and comprehensive data sets would be needed to quantify these reliably.

Low diversity and rates of evolution in lentiviruses are commonly associated with lower pathogenicity. Although HIV-1 and HIV-2 achieve similar proviral loads in infected individuals, HIV-2 replicates to lower titres in the plasma (Popper *et al.*, 2000), displays a lower rate of sequence evolution (MacNeil *et al.*, 2007a) and lower

pathogenicity, reflected by a slower decline in CD4+ T cell numbers and slower disease progression (Marlink *et al.*, 1988; Marlink *et al.*, 1994; Popper *et al.*, 1999; Whittle *et al.*, 1994). Similarly, bovine immunodeficiency virus (BIV), which has low pathogenicity, is more closely related to FIV than to primate lentiviruses (Olmsted *et al.*, 1989) and also exhibits very little sequence variation (Carpenter *et al.*, 2000).

The genetic stability of some lentiviruses could be associated with the fidelity of their reverse transcriptase (Lewis *et al.*, 1999). Mutations in the catalytic YMDD motif of HIV-1 and the V148S mutation in SIV (Huisman *et al.*, 2008) have been implicated in increasing RT fidelity (Olmsted *et al.*, 1989; Operario *et al.*, 2005). It is likely that the RT of FIV has enhanced fidelity compared to the RT of HIV-1 (Huisman *et al.*, 2008). The presence of almost identical sequences in our study might further indicate that FIV exhibits a low replication rate during persistent infection, or that most infections occur via cell-to-cell transfer owing to feline tetherin/BST-2 (Dietrich *et al.*, 2011). It is also possible that viral evolution is constrained due to the high fitness cost associated with divergence from the parental virus (Domingo & Holland, 1997). Whatever the specific mechanism, the low sequence variation of FIV in comparison to HIV-1 could be indicative of co-adaptation between FIV and its host, which are thought to have co-existed together for longer than HIV-1 and humans (Pecon-Slattery *et al.*, 2008). Furthermore, the analysis of viral sequence divergence and disease pathogenesis amongst the Felidae suggest different periods of virus-host co-evolution (Troyer *et al.*, 2008). In general, the least pathogenic viruses appear to have co-evolved with their hosts for longer periods of time, whereas more virulent strains have more recent origins. This pattern is reflected in the disease inducing potential of different species-specific FIVs, with FIV-Pco (FIV of the puma) being the least pathogenic while FIV-Fca (FIV of the domestic cat) is the most pathogenic. Although

FIV infection in pumas and lions was previously reported to be asymptomatic, recent evidence suggests that infected lions do indeed exhibit mild clinical sings of immunodeficiency, subsequently affecting the lifespans of infected animals (Roelke *et al.*, 2009). Similar observations have been made recently in SIV-infected chimpanzees, previously regarded as asymptomatic virus carriers (Terio *et al.*, 2011).

It has been suggested that the slow rate of evolution documented in terminal HIV-1 infection might be linked to the weakened host immune system (Delwart *et al.*, 1997; Shankarappa *et al.*, 1999). Consistent with this theory, we found that the V3-V5 fragment, which contains neutralisation epitopes and is under immune system surveillance (Lombardi *et al.*, 1993; Pancino *et al.*, 1993), evolved faster in healthy cats than in terminally ill cats. In addition, the rates of evolution across all data sets from the individually housed cats from Chicago (which were generally in good health) were consistently higher than those estimated for sequences from the cats from Memphis (which displayed more clinical signs of illness).

Recombination is a potential driver of rapid evolutionary change and plays a significant role in generating viral diversity among retroviruses (Onafuwa-Nuga & Telesnitsky, 2009). In the present study, viruses with a prior history of recombination (Bęczkowski *et al.*, 2014a) exhibited higher evolutionary rates, which could be the result of strong immune selection pressure acting on the recombinant Envs compared to (presumably) more host adapted, evolutionary older non-recombinants. Nonetheless, the mechanisms responsible for the faster evolution of recombinants might not be related to the immune system, and could be attributed to their higher replication tempo and increased infectivity (Tebit *et al.*, 2007).

This is the first study to demonstrate that against a background of low diversity and slow evolution for *env* as a whole, the leader region exhibited the highest proportion

of positively selected sites and the highest evolutionary rate among all fragments examined, but more data is required to confirm this. The N-terminal signal peptide of FIV plays an important role in post-translational targeting of the Env precursor and its translocation through the endoplasmic reticulum (ER) (Verschoor *et al.*, 1993). Studies of viral leader sequences highlighted their enormous complexity, suggesting a role not only in post-translational events but also in post-cleavage events (Hegde & Bernstein, 2006; Lemberg & Martoglio, 2002). Leader encoded peptides are involved in self-antigen presentation (Borrego *et al.*, 1998), a property exploited by human cytomegalovirus. This virus has a signal peptide with 9 residues identical to the MHC I signal peptides (Tomasec, 2000), which helps it to evade detection by NK cells presenting a virus-encoded MHC I molecule (Ulbrecht *et al.*, 2000). Furthermore, sequence variation among the signalling peptides of acute and chronic isolates of HIV has also been demonstrated (Gnanakaran *et al.*, 2011). These findings suggest that the leader may play an important role in regulating *env* expression on the viral surface (Hegde & Bernstein, 2006) and therefore viral infectivity. This hypothesis warrants further research and validation in functional studies.

By comparing amino acid sequences and global patterns of PNGS, 19 sites were identified which were conserved in over 97% of the examined sequences, with 6 PNGS displaying 100% fixed cross-clade pattern. This striking consistency at the population level suggests that glycosylation at these sites is likely to play an essential role in the folding and integrity of the viral Env. Patterns of fixed N-linked glycosylation sites have also been noted in infections with SIV (Chackerian *et al.*, 1997) and HIV-1 group M subtypes A through G (Gao *et al.*, 1996).

Statistical characterization of a relationship between within-host evolutionary dynamics and the clinical outcome of retroviral infection remains problematic.

Although HPDs permit a comparison of evolutionary rates between groups of patients, the estimate uncertainty tends to be high due to limited signal within intra-host data sets (Carvajal-Rodriguez *et al.*, 2008). We have considered hierarchical phylogenetic models (HPMs) to test for rate differences between cats, but preliminary analyses suggested limited power to detect effects in our data using these parameter-rich models as also suggested by others (Edo-Matas *et al.*, 2011).

The conclusions drawn from the intra-host diversity and rates of evolution are restricted by 1) the 63% mortality rate within the Memphis cohort, 2) the relatively low number of sequences available from some cats at selected time-points, and 3) the short 12-month study period (Seo *et al.*, 2002) and indeed we identified almost identical sequences over the 3 different sampling times. However, despite this short time frame and the potential limitations of the methodology used, we reliably quantified within-host evolution. The presence of identical amplicons in cats at 3 different sampling times suggest that factors which tend to generate false diversity (polymerase template switching, PCR induced errors) did not affect our results. Furthermore, the results presented here are in agreement with previous FIV studies employing "bulk" PCR (Ikeda *et al.*, 2004; Motokawa *et al.*, 2005), end-point dilution proviral DNA PCR (Kraase *et al.*, 2010) as well as a study examining sequences from plasma viral RNA (Huisman *et al.*, 2008) and we observed low sequence variation and high genetic stability of FIV in a relatively large sample of naturally infected cats.

Conclusion

We observed relatively low intra-host diversity and a low rate of evolution of the entire *env* in cats naturally infected with FIV. The greater overall genetic stability of FIV compared to HIV-1 might explain why many naturally infected cats do not progress rapidly to AIDS. This is the first study to demonstrate that the leader

sequence is the fastest evolving region of *env*. Since the majority of previous studies focused on V3-V5 region, these results indicate that the role of the unusually long leader sequence of FIV *env* in viral infectivity and immune evasion might have been underestimated.

Methods

Cats, FIV *env* sequences and alignments

Forty four privately-owned, neutered domestic cats, of various ages, breeds and conditions of health, were enrolled into the study (Bęczkowski, 2013) based on a history of a positive FIV antibody test result (SNAP® FIV/FeLV Combo Test, IDEXX Laboratories). All cats were FeLV antigen negative and their FIV positive status was confirmed by virus isolation (Hosie *et al.*, 2009). In the study group, 27 cats lived together in a large multi-cat household in Memphis, TN, USA, where FIV-positive and FIV-negative cats were housed indoors with unrestricted access to one another. The remaining 17 cats lived in individual households in Chicago, IL, USA with exception of 5 cats: 2 cats (P7 and P4) had been rehomed together and were living in the same household; 1 cat (P9) had been rehomed with another FIV-positive cat not enrolled in the study; 1 cat (P13) had been rehomed with another FIV-negative cat; and 1 cat (P21) was housed with another 2 FIV-positive cats in the rehoming centre. Cats were classified, based on a 6-monthly clinical examination by a registered specialist in feline medicine (AL), into 2 groups: 1) healthy - cats with no abnormalities found on clinical examination, and 2) unhealthy – cats with any abnormalities detected on clinical examination. At the time of enrolment, there were 10 healthy (59%) and 7 not healthy (41%) cats in the Chicago cohort. In the Memphis cohort, 12 cats were classified as healthy (44%) and 15 cats were classified as unhealthy (56%) at the time of enrolment. During the study period, there was a 63% mortality rate in the Memphis cohort (17/27), while only 1 cat (5.9 %) from Chicago cohort died during the same time frame (Bęczkowski, 2013).

Multiple full length FIV *env* genes (~2500 bp) were amplified directly from whole blood collected at 6 monthly intervals starting in January and May 2010 for Memphis

and Chicago respectively (time points A, B and C), using nested PCR protocol (Additional file 6-Table S4). First round PCR products were amplified directly from blood, without genomic DNA extraction by Phusion® Blood Direct II Polymerase (Thermo Fisher Scientific) followed by direct nucleic acid sequence determination. Phusion® Blood Direct II Polymerase is a proofreading polymerase with 25 fold greater accuracy than Taq DNA polymerase, determined with a modified lacI-based method (Frey & Suppmann, 1995). The nucleic acid sequence of the first-round PCR product informed the primer design for the second round PCR, which was performed using High Fidelity PCR Master (Roche), which amplifies with 3 fold greater accuracy than Taq DNA polymerase (Roche). Strain-specific primers for the second round PCR incorporated restriction sites. Amplified *envs* were cloned into the eukaryotic expression vector VR1012 (Hartikka *et al.*, 1996) and transformed into *E. coli* MAX Efficiency® DH5α™ Competent Cells (Invitrogen). Thus constructed VR1012 plasmids expressing FIV *env* were sequenced (Additional file 7-Table S5) using Big Dye Terminator v1.1 kit (Applied Biosystems) for the purpose of the present study, before being assessed in functional studies (Bęczkowski *et al.*, 2014b; Bęczkowski et al., 2014c). Special measures were taken to avoid the possibility of contamination, in both the clinical and laboratory settings: cats were double identified prior to blood sampling, PCR reactions were prepared in a designated UV treated room, and fresh, unopened reagents were used at each separate time point throughout the study.

There were 355 serial *env* sequences from 43 cats available for analysis from the 2 cohorts (Additional file 7-Table S5). The number of sequences varied according to the availability of follow-up samples, being largely influenced by the 63% mortality rate within the Memphis cohort. Multiple sequence alignments were conducted in

MEGA5 (Tamura *et al.*, 2011) and curated manually to ensure homology of gaps in sequences of variable length. Analyses were performed using the entire *env* DNA sequences and the 3 structural fragments: 1) leader/signal region (approx. 509 bp in length); 2) variable V3-V5 region (approx. 630bp); and 3) remaining concatenated fragments of the entire *env* after exclusion of the V3-V5 region, denoted NV (approx. 1900bp). The temporal signal and 'clocklikeness' of the *env* phylogenies were tested in Path-o-Gen (http://tree.bio.ed.ac.uk), (Additional file 8-Table S6).

Phylogenetic trees

Maximum likelihood (ML) trees were constructed in MEGA5 (Tamura *et al.*, 2011) under the HKY nucleotide substitution model, selected through jMODELTEST analysis (Posada, 2008). Statistical support for ML tree was estimated using bootstrapping analysis with 1000 replicates (Efron *et al.*, 1996). The ML tree constructed using the entire data set was carefully examined for evidence of non-monophyletic clustering of multiple sequences amplified from each individual. Sequences from 3 animals (M5, P8 and P21), amplified at different time points did not form monophyletic groups and were excluded from further intra-host and evolutionary rate analyses.

Intra-host diversity

Intra-host sequence variation among the *env* sequences and 3 fragments of the gene: 1) leader, 2) V3-V5 and 3) NV at each time point were calculated as mean and highest pairwise distances under HKY nucleotide substitution model in PAUP 4.0 b10 (Wilgenbusch & Swofford, 2002). To explore the variation in mean pairwise distances of 3 structural fragments of the *env* a General Linear Model (GLM) (Khan, 2013) was used with "structural fragment of the *env"* and "time point" as fixed effects

and “cat” as a random effect. No evidence of recombination within hosts was found using rigorous five-fold recombination testing (Bęczkowski *et al.*, 2014a).

Selection and PNGS

Nucleotide sites under diversifying or purifying selection were identified and dN/dS ratios (ω) for every codon in the alignment, were estimated using three different methods: 1) single likelihood ancestor counting (SLAC) (Kosakovsky Pond & Frost, 2005) at $p<0.01$, 2) fixed effects likelihood (FEL) (Kosakovsky Pond & Frost, 2005) at $p<0.1$, and 3) internal fixed effects likelihood (IFEL) (Kosakovsky Pond *et al.*, 2006b) at $p<0.1$. To focus on the most strongly supported positively selected sites, we reported those sites which were consistently identified by all three methods.

Potential N-linked glycosylation sites and their position in the protein alignment were identified by N-GlycoSite tool available at Los Alamos National Laboratory web server (http://www.hiv.lanl.gov/) and the number of sequences with glycosylation sites was counted. The analysis included reference sequences from GenBank representing clades A, B, C and D: Aomori 1 [GenBank:D37816], Aomori 2 [GenBank:D37817.1], FIV C [GenBank:AF474246.1], Dixon [GenBank:L00608.1], Dutch [GenBank:X60725], Fukuoka [GenBank:D37815.1], Sendai 1 [GenBank:D37813.1], Shizuoka [GenBank:D37811.1], UK2 [GenBank:X69494.1], UK8 [GenBank:X69496.1], USIL2489 [GenBank:U11820.1], Yokohama [GenBank:D37812.1], Petaluma [GenBank:M25381.1], PPR [GenBank:M36968.1] and sequences from all cats excluding identical, duplicate sequences (n=329). Alignments were numbered according to positions in the M49C C76 sequence.

Rate of molecular evolution

Rates of evolution of the entire *env* gene and individual fragments from each Memphis and Chicago cohorts and all cats were estimated using Bayesian

Evolutionary Analysis Sampling Trees (BEAST) version v.1.7.1 (Drummond & Rambaut, 2007), based on sampling date information. Additionally, rate comparisons were made between non-recombinant or recombinant sequences, sequences from healthy and sick cats, and cats from Memphis that remained alive and those that died during the study period.

Sequence alignments from animals from which data were available from more than 1 time point were included in the analysis: 17 cats from Memphis (197 sequences) and 12 cats from Chicago (108 sequences). The analysis of differences in evolutionary rate between sequences from alive (n=6) and deceased cats (n=11), was based on sequences from Memphis cohort. The HKY evolutionary model of substitution with four category gamma distribution was selected with codon positions (1+2) and 3 as partitions (Shapiro *et al.*, 2006). The analyses in BEAST were performed estimating independent trees for each cat, with a linked clock rate parameter across individual cats, under strict and relaxed lognormal clock models with uniform distribution clock rate priors, informed by previous estimates of evolutionary rates for FIV and HIV (Table 1). The length of MCMC chain was set for 200,000,000 iterations with 20,000,000 burn-in and was run until convergence and effective sample sizes >100 were obtained.

The BEAST generated output log file was analysed in TRACER v1.5. The coefficient of variation in rates among branches was used to determine whether a relaxed molecular clock was more appropriate for these data. This was assumed to be the case if the estimate for the coefficient of variation excluded zero.

The statistical support is provided in the form of parameter estimates and their highest posterior densities (HPD).

The study and its aims were reviewed and approved by the University of Glasgow Ethics Committee and the Purdue Animal Care and Use Committee. Cat owners provided written informed consent for their participation in the study.

List of abbreviations

FIV: Feline immunodeficiency virus

BEAST: Bayesian evolutionary analysis sampling trees

MCMC: Markov chain Monte Carlo

ML: Maximum likelihood

SLAC: Single likelihood ancestor counting

FEL: Fixed effects likelihood

IFEL: Internal fixed effects likelihood

PNGS: Potential N-linked glycosylation sites

bp: Base pair

HPM: Hierarchical phylogenetic models

HPD: Highest posterior density

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

PB carried out amplifications, cloning, sequencing, alignments, data analyses, coordinated the study and drafted the manuscript. JH and RB contributed to the design, implementation of the data analysis and draft of the manuscript. AL carried out clinical examinations, collected the blood samples and coordinated logistics. MJH and BJW conceived the study, participated in its design and coordination and contributed to the draft of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This study was supported by The Wellcome Trust. Blood specimen collections were supported by the Purdue Maddie's Shelter Medicine Program underwritten by a grant from Maddie's Fund, The Pet Rescue Foundation (www.maddiesfund.org), helping to fund the creation of a no-kill nation. RB is supported by the RAPIDD programme of the Science and Technology Directorate of the Department of Homeland Security and NIH Fogarty International Center. We thank Professor Dominic Mellor for statistical advice. We thank Kristen Hall CVT, Dr Jui Ming Lin, Dr Christian Leutenegger, PAWS Chicago, Drennan Animal Hospital, the Fitzhugh B. Crews FIV Cat Sanctuary and participating cat owners for their assistance with the study.

References

Abecasis, A. B., Vandamme, A. M. & Lemey, P. (2009). Quantifying differences in the tempo of human immunodeficiency virus type 1 subtype evolution. *J Virol* **83**, 12917-12924.

Alizon, S. & Fraser, C. (2013). Within-host and between-host evolutionary rates across the HIV-1 genome. Retrovirology **10**, 49.

Bęczkowski P. M. (2013). Virus evolution in the progression of natural feline immunodeficiency virus infection. PhD Thesis, University of Glasgow, Centre for Virus Research.

Bęczkowski, P. M., Hughes, J., Biek, R., Litster, A., Willett, B. J. & Hosie, M. J. (2014a). Feline immunodeficiency virus (FIV) env recombinants are common in natural infections. *Retrovirology* **11**, 80.

Bęczkowski, P. M., Techakriengkrai, N., Logan, N., McMonagle, E., Litster, A., Willett, B. J. & Hosie, M. J. (2014b). Emergence of CD134 cysteine-rich domain 2 (CRD2)-independent strains of feline immunodeficiency virus (FIV) is associated with disease progression in naturally infected cats. Retrovirology **11**, 95.

Bęczkowski, P. M., Logan, N., McMonagle, E., Litster, A., Willett, B. J. & Hosie, M. J. (2014c). An investigation of the breadth of neutralising antibody response in cats naturally infected with feline immunodeficiency virus (FIV). J Gen Virol. (in press).

Biek, R., Rodrigo, A. G., Holley, D., Drummond, A., Anderson, C. R., Ross, H. A. & Poss, M. (2003). Epidemiology, Genetic Diversity, and Evolution of Endemic Feline Immunodeficiency Virus in a Population of Wild Cougars. *J Virol* **77**, 9578-9589.

Borrego, F., Ulbrecht, M., Weiss, E. H., Coligan, J. E. & Brooks, A. G. (1998). Recognition of human histocompatibility leukocyte antigen (HLA)-E complexed with HLA class I signal sequence-derived peptides by CD94/NKG2 confers protection from natural killer cell-mediated lysis. *Journal of Experimental Medicine* **187**, 813-818.

Carpenter, M. A., Brown, E. W., MacDonald, D. W. & O'Brien S, J. (1998). Phylogeographic patterns of feline immunodeficiency virus genetic diversity in the domestic cat. *Virology* **251**, 234-243.

Carpenter, S., Vaughn, E. M., Yang, J., Baccam, P., Roth, J. A. & Wannemuehler, Y. (2000). Antigenic and genetic stability of bovine immunodeficiency virus during long-term persistence in cattle experimentally infected with the BIVR29 isolate. *Journal of General Virology* **81**, 1463-1472.

Carvajal-Rodriguez, A., Posada, D., Perez-Losada, M., Keller, E., Abrams, E. J., Viscidi, R. P. & Crandall, K. A. (2008). Disease progression and evolution of the HIV-1 env gene in 24 infected infants. *Infect Genet Evol* **8**, 110-120.

Chackerian, B., Rudensey, L. M. & Overbaugh, J. (1997). Specific N-linked and O-linked glycosylation modifications in the envelope V1 domain of simian immunodeficiency virus variants that evolve in the host alter recognition by neutralizing antibodies. *J Virol* **71**, 7719-7727.

Choisy, M., Woelk, C. H., Guégan, J.-F. & Robertson, D. L. (2004). Comparative study of adaptive molecular evolution in different human immunodeficiency virus groups and subtypes. *J Virol* **78**, 1962-1970.

Delwart, E. L., Pan, H., Sheppard, H. W., Wolpert, D., Neumann, A. U., Korber, B. & Mullins, J. I. (1997). Slower evolution of human immunodeficiency virus type 1 quasispecies during progression to AIDS. *J Virol* **71**, 7498-7508.

Dietrich, I., McMonagle, E. L., Petit, S. J., Vijayakrishnan, S., Logan, N., Chan, C. N., Towers, G. J., Hosie, M. J. & Willett, B. J. (2011). Feline tetherin efficiently restricts release of feline immunodeficiency virus but not spreading of infection. *J Virol* **85**, 5840-5852.

Domingo, E. (2000). Viruses at the edge of adaptation. *Virology* **270**, 251-253.

Domingo, E. & Holland, J. J. (1997). RNA virus mutations and fitness for survival, pp. 151-178.

Drummond, A. J. & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *Bmc Evolutionary Biology* **7**.

Dunham, S. P., Bruce, J., MacKay, S., Golder, M., Jarrett, O. & Neil, J. C. (2006). Limited efficacy of an inactivated feline immunodeficiency virus vaccine. *Vet Rec* **158**, 561-562.

Edo-Matas, D., Lemey, P., Tom, J. A., Serna-Bolea, C., van den Blink, A. E., van 't Wout, A. B., Schuitemaker, H. & Suchard, M. A. (2011). Impact of CCR5delta32 host genetic background and disease progression on HIV-1 intrahost evolutionary processes: efficient hypothesis testing through hierarchical phylogenetic models. *Molecular Biology and Evolution* **28**, 1605-1616.

Efron, B., Halloran, E. & Holmes, S. (1996). Bootstrap confidence levels for phylogenetic trees. *Proc Natl Acad Sci U S A* **93**, 13429-13434.

Eigen, M. (1993). Viral quasispecies. *Sci Am* **269**, 42-49.

Frey, B. & Suppmann, B. (1995). Demonstration of the expend PCR systems greater fidelity and higher yields with a lacIbased fidelity assay. Biochemica 2, 34-35.

Gao, F., Morrison, S. G., Robertson, D. L., Thornton, C. L., Craig, S., Karlsson, G., Sodroski, J., Morgado, M., Galvao-Castro, B. & other authors (1996).

Molecular cloning and analysis of functional envelope genes from human immunodeficiency virus type 1 sequence subtypes A through G. The WHO and NIAID Networks for HIV Isolation and Characterization. *J Virol* **70**, 1651-1667.

Gnanakaran, S., Bhattacharya, T., Daniels, M., Keele, B. F., Hraber, P. T., Lapedes, A. S., Shen, T., Gaschen, B., Krishnamoorthy, M. & other authors (2011). Recurrent signature patterns in HIV-1 B clade envelope glycoproteins associated with either early or chronic infections. *PLoS Pathog* **7**, e1002209.

Greene, W. K., Meers, J., del Fierro, G., Carnegie, P. R. & Robinson, W. F. (1993). Extensive sequence variation of feline immunodeficiency virus env genes in isolates from naturally infected cats. *Arch Virol* **133**, 51-62.

Hartikka, J., Sawdey, M., CornefertJensen, F., Margalith, M., Barnhart, K., Nolasco, M., Vahlsing, H. L., Meek, J., Marquet, M. & other authors (1996). An improved plasmid DNA expression vector for direct injection into skeletal muscle. *Hum Gene Ther* **7**, 1205-1217.

Hayward, J. J. & Rodrigo, A. G. (2010). Molecular epidemiology of feline immunodeficiency virus in the domestic cat (Felis catus). *Vet Immunol Immunopathol* **134**, 68-74.

Hegde, R. S. & Bernstein, H. D. (2006). The surprising complexity of signal sequences. *Trends Biochem Sci* **31**, 563-571.

Hosie, M. J., Addie, D., Belák, S., Boucraut-Baralon, C., Egberink, H., Frymus, T., Gruffydd-Jones, T., Hartmann, K., Lloret, A. & other authors (2009). Feline immunodeficiency. ABCD guidelines on prevention and management. *Journal of Feline Medicine & Surgery* **11**, 575-584.

Huisman, W., Schrauwen, E. J. A., Rimmelzwaan, G. F. & Osterhaus, A. D. M. E. (2008). Intrahost evolution of envelope glycoprotein and OrfA sequences after experimental infection of cats with a molecular clone and a biological isolate of feline immunodeficiency virus. *Virus Research* **137**, 24-32.

Ikeda, Y., Miyazawa, T., Nishimura, Y., Nakamura, K., Tohya, Y. & Mikami, T. (2004). High genetic stability of TM1 and TM2 strains of subtype B feline immunodeficiency virus in long-term infection. *Journal of Veterinary Medical Science* **66**, 287-289.

Khan, R. M. (2013). Analysis of Variance. In *Problem Solving and Data Analysis using Minitab*, pp. 150-208: John Wiley & Sons, Ltd.

Kils-Hutten, L., Cheynier, R., Wain-Hobson, S. & Meyerhans, A. (2001). Phylogenetic reconstruction of intrapatient evolution of human immunodeficiency virus type 1: predominance of drift and purifying selection. *J Gen Virol* **82**, 1621-1627.

Korber, B., Muldoon, M., Theiler, J., Gao, F., Gupta, R., Lapedes, A., Hahn, B. H., Wolinsky, S. & Bhattacharya, T. (2000). Timing the ancestor of the HIV-1 pandemic strains. *Science* **288**, 1789-1796.

Kosakovsky Pond, S. L. & Frost, S. D. (2005). Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Molecular Biology and Evolution* **22**, 1208-1222.

Kosakovsky Pond, S. L., Posada, D., Gravenor, M. B., Woelk, C. H. & Frost, S. D. (2006a). Automated phylogenetic detection of recombination using a genetic algorithm. *Molecular Biology and Evolution* **23**, 1891-1901.

Kosakovsky Pond, S. L., Frost, S. D. W., Grossman, Z., Gravenor, M. B., Richman, D. D. & Brown, A. J. L. (2006b). Adaptation to different human

populations by HIV-1 revealed by codon-based analyses. *PLoS Comput Biol* **2**, e62.

Kraase, M., Sloan, R., Klein, D., Logan, N., McMonagle, L., Biek, R., Willett, B. J. & Hosie, M. J. (2010). Feline immunodeficiency virus env gene evolution in experimentally infected cats. *Vet Immunol Immunopathol* **134**, 96-106.

Leitner, T. & Albert, J. (1999). The molecular clock of HIV-1 unveiled through analysis of a known transmission history. *Proceedings of the National Academy of Sciences* **96**, 10752-10757.

Lemberg, M. K. & Martoglio, B. (2002). Requirements for signal peptide peptidase-catalyzed intramembrane proteolysis. *Molecular Cell* **10**, 735-744.

Lewis, D. A., Bebenek, K., Beard, W. A., Wilson, S. H. & Kunkel, T. A. (1999). Uniquely altered DNA replication fidelity conferred by an amino acid change in the nucleotide binding pocket of human immunodeficiency virus type 1 reverse transcriptase. *Journal of Biological Chemistry* **274**, 32924-32930.

Lombardi, S., Garzelli, C., La Rosa, C., Zaccaro, L., Specter, S., Malvaldi, G., Tozzini, F., Esposito, F. & Bendinelli, M. (1993). Identification of a linear neutralization site within the third variable region of the feline immunodeficiency virus envelope. *J Virol* **67**, 4742-4749.

Lukashov, V. V., Kuiken, C. L. & Goudsmit, J. (1995). Intrahost human immunodeficiency virus type 1 evolution is related to length of the immunocompetent period. *J Virol* **69**, 6911-6916.

Lythgoe, K. A. & Fraser, C. (2012). New insights into the evolutionary rate of HIV-1 at the within-host and epidemiological levels. *Proc R Soc B-Biol Sci* **279**, 3367-3375.

MacNeil, A., Sankale, J. L., Meloni, S. T., Sarr, A. D., Mboup, S. & Kanki, P. (2007a). Long-term intrapatient viral evolution during HIV-2 infection. *J Infect Dis* **195**, 726-733.

MacNeil, A., Sarr, A. D., Sankalé, J. L., Meloni, S. T., Mboup, S. & Kanki, P. (2007b). Direct evidence of lower viral replication rates in vivo in human immunodeficiency virus type 2 (HIV-2) infection than in HIV-1 infection. *J Virol* **81**, 5325-5330.

Marcola, T. G., Gomes, C. P., Silva, P. A., Fernandes, G. R., Paludo, G. R. & Pereira, R. W. (2013). Identification of a novel subtype of feline immunodeficiency virus in a population of naturally infected felines in the Brazilian Federal District. Virus Genes 46, 546-550.

Marlink, R., Kanki, P., Thior, I., Travers, K., Eisen, G., Siby, T., Traore, I., Hsieh, C. C., Dia, M. C. & other authors (1994). Reduced rate of disease development after HIV-2 infection as compared to HIV-1. *Science* **265**, 1587-1590.

Marlink, R. G., Ricard, D., M'Boup, S., Kanki, P. J., Romet-Lemonne, J. L., N'Doye, I., Diop, K., Simpson, M. A., Greco, F. & other authors (1988). Clinical, hematologic, and immunologic cross-sectional evaluation of individuals exposed to human immunodeficiency virus type-2 (HIV-2). *Aids Res Hum Retrovir* **4**, 137-148.

Mikhail, M., Wang, B., Lemey, P., Beckthold, B., Vandamme, A. M., Gill, M. J. & Saksena, N. K. (2005). Role of viral evolutionary rate in HIV-1 disease progression in a linked cohort. *Retrovirology* **2**, 41.

Motokawa, K., Hohdatsu, T., Imori, A., Arai, S. & Koyama, H. (2005). Mutations in feline immunodeficiency (FIV) virus envelope gene V3-V5 regions in FIV-infected cats. *Vet Microbiol* **106**, 33-40.

Olmsted, R. A., Hirsch, V. M., Purcell, R. H. & Johnson, P. R. (1989). Nucleotide sequence analysis of feline immunodeficiency virus: Genome organization and relationship to other lentiviruses. *Proc Natl Acad Sci U S A* **86**, 8088-8092.

Olmsted, R. A., Langley, R., Roelke, M. E., Goeken, R. M., Adger-Johnson, D., Goff, J. P., Albert, J. P., Packer, C., Laurenson, M. K. & other authors (1992). Worldwide prevalence of lentivirus infection in wild feline species: epidemiologic and phylogenetic aspects. *J Virol* **66**, 6008-6018.

Onafuwa-Nuga, A. & Telesnitsky, A. (2009). The remarkable frequency of human immunodeficiency virus type 1 genetic recombination. *Microbiology and Molecular Biology Reviews* **73**, 451-480.

Operario, D. J., Reynolds, H. M. & Kim, B. (2005). Comparison of DNA polymerase activities between recombinant feline immunodeficiency and leukemia virus reverse transcriptases. *Virology* **335**, 106-121.

Pancino, G., Chappey, C., Saurin, W. & Sonigo, P. (1993). B epitopes and selection pressures in feline immunodeficiency virus envelope glycoproteins. *J Virol* **67**, 664-672.

Pecon-Slattery, J., Troyer, J. L., Johnson, W. E. & O'Brien, S. J. (2008). Evolution of feline immunodeficiency virus in Felidae: implications for human health and wildlife ecology. *Vet Immunol Immunopathol* **123**, 32-44.

Pistello, M., Cammarota, G., Nicoletti, E., Matteucci, D., Curcio, M., Del Mauro, D. & Bendinelli, M. (1997). Analysis of the genetic diversity and phylogenetic relationship of Italian isolates of feline immunodeficiency virus

indicates a high prevalence and heterogeneity of subtype B. *Journal of General Virology* **78**, 2247-2257.

Popper, S. J., Sarr, A. D., Guèye-Ndiaye, A., Mboup, S., Essex, M. E. & Kanki, P. J. (2000). Low plasma human immunodeficiency virus type 2 viral load is independent of proviral load: low virus production in vivo. *J Virol* **74**, 1554-1557.

Popper, S. J., Sarr, A. D., Travers, K. U., Guèye-Ndiaye, A., Mboup, S., Essex, M. E. & Kanki, P. J. (1999). Lower human immunodeficiency virus (HIV) type 2 viral load reflects the difference in pathogenicity of HIV-1 and HIV-2. *Journal of Infectious Diseases* **180**, 1116-1121.

Posada, D. (2008). jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution* **25**, 1253-1256.

Rambaut, A., Posada, D., Crandall, K. A. & Holmes, E. C. (2004). The causes and consequences of HIV evolution. *Nature reviews Genetics* **5**, 52-61.

Roelke, M. E., Brown, M. A., Troyer, J. L., Winterbach, H., Winterbach, C., Hemson, G., Smith, D., Johnson, R. C., Pecon-Slattery, J. & other authors (2009). Pathological manifestations of feline immunodeficiency virus (FIV) infection in wild African lions. Virology **390**, 1-12.

Salemi, M. (2013). The intra-host evolutionary and population dynamics of human immunodeficiency virus type 1: a phylogenetic perspective. *Infect Dis Rep* **5**, e3.

Samman, A., McMonagle, E. L., Logan, N., Willett, B. J., Biek, R. & Hosie, M. J. (2011). Phylogenetic characterisation of naturally occurring feline immunodeficiency virus in the United Kingdom. *Vet Microbiol* **150**, 239-247.

Schultz, A. K., Zhang, M., Leitner, T., Kuiken, C., Korber, B., Morgenstern, B. & Stanke, M. (2006). A jumping profile Hidden Markov Model and applications to recombination sites in HIV and HCV genomes. *BMC Bioinformatics* **7**, 265.

Seo, T.-K., Thorne, J. L., Hasegawa, M. & Kishino, H. (2002). A viral sampling design for testing the molecular clock and for estimating evolutionary rates and divergence times. *Bioinformatics* **18**, 115-123.

Shankarappa, R., Margolick, J. B., Gange, S. J., Rodrigo, A. G., Upchurch, D., Farzadegan, H., Gupta, P., Rinaldo, C. R., Learn, G. H. & other authors (1999). Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. *J Virol* **73**, 10489-10502.

Shapiro, B., Rambaut, A. & Drummond, A. J. (2006). Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. *Molecular Biology and Evolution* **23**, 7-9.

Sodora, D. L., Shpaer, E. G., Kitchell, B. E., Dow, S. W., Hoover, E. A. & Mullins, J. I. (1994). Identification of three feline immunodeficiency virus (FIV) env gene subtypes and comparison of the FIV and human immunodeficiency virus type 1 evolutionary patterns. *J Virol* **68**, 2230-2238.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* **28**, 2731-2739.

Tebit, D. M., Nankya, I., Arts, E. J. & Gao, Y. (2007). HIV diversity, recombination and disease progression: how does fitness "fit" into the puzzle? *AIDS Rev* **9**, 75-87.

Teixeira, B. M., Logan, N., Cruz, J. C. M., Reis, J. K. P., Brandao, P. E., Richtzenhain, L. J., Hagiwara, M. K., Willett, B. J. & Hosie, M. J. (2010). Genetic diversity of Brazilian isolates of feline immunodeficiency virus. *Arch Virol* **155**, 379-384.

Terio, K. A., Kinsel, M. J., Raphael, J., Mlengeya, T., Lipende, I., Kirchhoff, C. A., Gilagiza, B., Wilson, M. L., Kamenya, S. & other authors (2011). Pathologic lesions in chimpanzees (Pan trogylodytes schweinfurthii) from Gombe National Park, Tanzania, 2004-2010. J Zoo Wildl Med **42**, 597-607.

Tomasec, P. (2000). Surface expression of HLA-E, an inhibitor of natural killer cells, enhanced by human cytomegalovirus gpUL40. *Science* **287**, 1031-1033.

Troyer, J. L., Vandewoude, S., Pecon-Slattery, J., McIntosh, C., Franklin, S., Antunes, A., Johnson, W. & O'Brien, S. J. (2008). FIV cross-species transmission: an evolutionary prospective. Vet Immunol Immunopathol **123**, 159-166.

Ulbrecht, M., Martinozzi, S., Grzeschik, M., Hengel, H., Ellwart, J. W., Pla, M. & Weiss, E. H. (2000). Cutting edge: The human cytomegalovirus UL40 gene product contains a ligand for HLA-E and prevents NK cell-mediated lysis. *J Immunol* **164**, 5019-5022.

Verschoor, E. J., Hulskotte, E. G., Ederveen, J., Koolen, M. J., Horzinek, M. C. & Rottier, P. J. (1993). Post-translational processing of the feline immunodeficiency virus envelope precursor protein. *Virology* **193**, 433-438.

Whittle, H., Morris, J., Todd, J., Corrah, T., Sabally, S., Bangali, J., Ngom, P. T., Rolfe, M. & Wilkins, A. (1994). HIV-2-infected patients survive longer than HlV-1-infected patients. *Aids* **8**, 1617-1620.

Wilgenbusch, J. C. & Swofford, D. (2002). Inferring evolutionary trees with PAUP*. In *Current Protocols in Bioinformatics*: John Wiley & Sons, Inc.

Wolfs, T. F., de Jong, J. J., Van den Berg, H., Tijnagel, J. M., Krone, W. J. & Goudsmit, J. (1990). Evolution of sequences encoding the principal neutralization epitope of human immunodeficiency virus 1 is host dependent, rapid, and continuous. *Proc Natl Acad Sci U S A* **87**, 9938-9942.

Yamaguchi, Y. & Gojobori, T. (1997). Evolutionary mechanisms and population dynamics of the third variable envelope region of HIV within single hosts. *Proc Natl Acad Sci U S A* **94**, 1264-1269.

Zhang, L., Diaz, R. S., Ho, D. D., Mosley, J. W., Busch, M. P. & Mayer, A. (1997). Host-specific driving force in human immunodeficiency virus type 1 evolution in vivo. *J Virol* **71**, 2555-2561.

Tables

Table 1 - Comparison of evolutionary rates of the *env* gene of HIV-1 and fragments of the *env* of FIV-Fca (FIV strain of domestic cat) and FIV-Pco (FIV strain of cougar).

Virus	Evolutionary rate (x10^{-3} substitutions per site per year)	Study
HIV-1 gp160 *env*	2.4 (1.8-2.8)*	(Korber *et al.*, 2000)
FIV-Fca V3-V6 *env*	3.1-6.6 (2.2–4.1, 5.0–8.4, 3.3–9.2)*[1]	(Hayward & Rodrigo, 2010)
FIV-Pco V4-V5 *env*	1.54 (0.89-2.23)*	(Biek *et al.*, 2003)
FIV-Fca V1-V2 *env*	3.4[ζ]	(Greene *et al.*, 1993)
FIV-Fca V1-V9 *env*	0.9[ɬ] (0.18-1.59)*-6.7[§] (1.72-13.35)*	(Kraase *et al.*, 2010)

*95% lower and upper highest posterior density (HPD),[1]95 % HPD for each of three cats (Hayward & Rodrigo, 2010), [ɬ] 322 weeks post infection, [§]12 weeks post infection (Kraase *et al.*, 2010), [ζ] evolutionary rate estimated using method of Gojobori and Yokoyama (Greene *et al.*, 1993).

Figures

Fig. 1 - Maximum likelihood (ML) tree, based on the HKY model, rooted on a clade C reference FIV *env*. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The ML phylogeny includes 47 entire *env* nucleotide sequences (representative of a total of 355 sequences from Chicago and Memphis), 15 entire *env* sequences derived from GenBank; accession numbers: Aomori 1 [GenBank:D37816], Aomori 2 [GenBank:D37817.1], FIV C [GenBank:AF474246.1], Dixon [GenBank:L00608.1], Dutch [GenBank:X60725], Fukuoka [GenBank:D37815.1], Sendai 1 [GenBank:D37813.1], Shizuoka [GenBank:D37811.1], UK2 [GenBank:X69494.1], UK8 [GenBank:X69496.1], USIL2489 [GenBank:U11820.1], Yokohama [GenBank:D37812.1], Petaluma [GenBank:M25381.1], PPR [GenBank:M36968.1], Leviano [GenBank:FJ374696.1], 3 V3-V5 region sequences representing Clade E: LP3 [GenBank:D84496], LP20 [GenBank:D84498], LP24 [GenBank:D84500] and 1 shorter 504 bp in length RUS14 [GenBank:EF447297] sequence. Taxa with inconsistent clade assignment are represented with an asterisk (P8, P21). Non-monophyletic taxa from cat M5 are marked with triangle. Only bootstrap values above 80 are shown.

Fig. 2 - Comparison of mean pairwise distances between *env* fragments and entire *env* sequences amplified from 3 samplings (A, B and C on the *x* axis)**.** Median values for the leader, V3-V5, NV and *env* respectively for: 1) time point A: (0.20%, 0.11%, 0.16% and 0.13%), 2) time point B: (0.20%, 0.12%, 0.12% and 0.11%), 3) time point C: (0.27%, 0.19%, 0.14% and 0.17%).

Fig. 3 - Frequency and location of predicted PNGS (a) and sites under positive selection (b) with corresponding dN/dS values. PNGS at positions 303, 335, 347, 536, 553 and 738 represent fixed sequons among all clades of FIV from different geographic origins (n=329). Sequons at positions 263, 279, 423, 427, 504, 734 and 341, 523, 746 were also fixed and were present in >99% of sequences. Positively selected sites (n=10) were consistently identified by three different detection methods (for detail see Additional file 4-Table S2). Leader sequence and B cell epitope regions identified in previous studies are highlighted in grey. The putative leader sequence cleavage site is located at the position 528 bp from the start codon of the *env* ORF of reference FIV Petaluma [GenBank: M25381.1]. Location of V1-V8 regions is highlighted below the graphs.

Fig. 4 - Distribution of the evolutionary rates calculated for the: 1) leader, 2) V3V5 fragment, 3) NV fragment and 4) entire *env* gene. Evolutionary rates are compared between sequences amplified from Memphis (M), Chicago (C) and all cats (SG). Rate estimates were either based on strict or relaxed clock models (for details see Additional file 5-Table S3).

Fig. 5 - Comparison of evolutionary rates of the leader, V3-V5, NV fragments and the entire *env* genes from 6 alive (healthy, A) and 11 deceased (sick, D) cats from Memphis cohort during the 12 month observation period. Rate estimates were either based on strict or relaxed clock models (for details see Additional file 5-Table S3).

Additional files

Additional file 1 – The Maximum Likelihood (ML) tree constructed under HKY substitution model using the entire data set. The tree is saved in the NEWICK format and can be opened in FigTree v 1.3.1 (http://tree.bio.ed.ac.uk/).

Additional file 2 – Fig. S1 Maximum Likelihood tree of 35 *env* sequences amplified from male cat M10 (in red) and 2 closely related cats (M1 (in blue) and M11 (in green)) rooted on reference clade B *env* sequence.

Additional file 3 – Table S1 Mean and highest pairwise distances (percentages) calculated for consecutive *env* sequences and its fragments, obtained from 3 blood samples (A, B and C) collected at 6 monthly intervals starting in January and May 2010 for Memphis and Chicago respectively.

Additional file 4 - Table S2 Positively selected sites identified by three different detection methods: 1) SLAC (p<0.01), 2) IFEL (p<0.1) and 3) FEL (p<0.1).

Additional file 5 - Table S3 Evolutionary rates of the entire *env*, and structural fragments of the *env* estimated under relaxed and strict clock models for Memphis and Chicago cohorts and the entire study group.

Additional file 6 – Table S4 Primers.

Additional file 7 – Table S5 Number of sequences detected at each time point from the US cats.

Additional file 8 - Table S6 Results of investigation of the temporal signal and 'clocklikeness' of *env* phylogenies, from Memphis and Chicago cats.

Figure 1

Click here to download Figure: Figure 1.docx

Figure 1

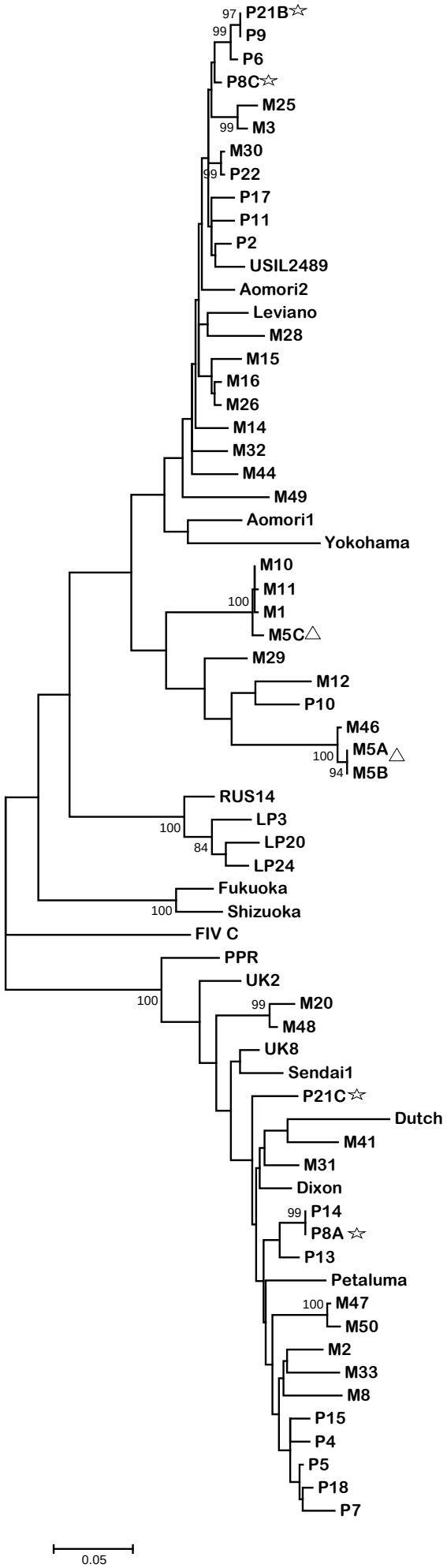


Figure 2
Click here to download Figure: Figure 2.docx

Figure 2

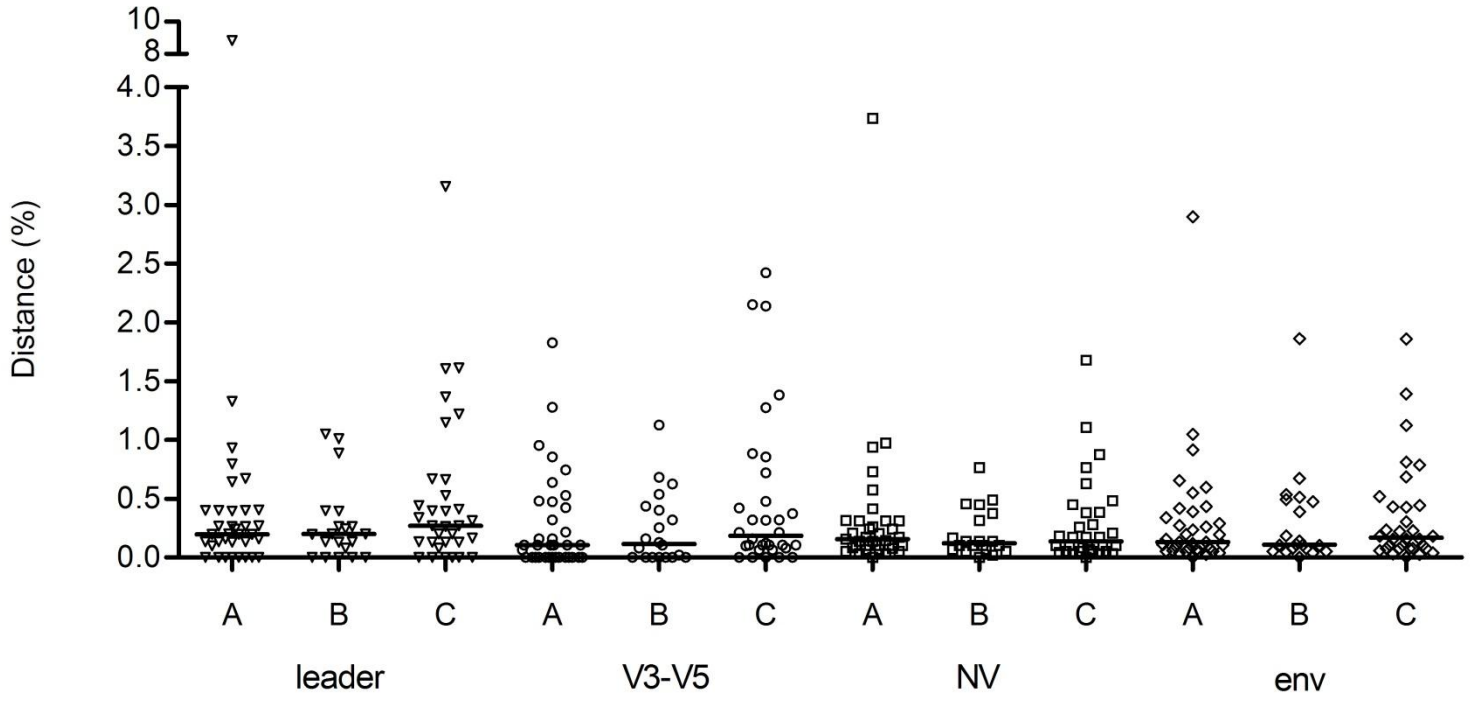


Figure 3
Click here to download Figure: Figure 3.docx

Figure 3

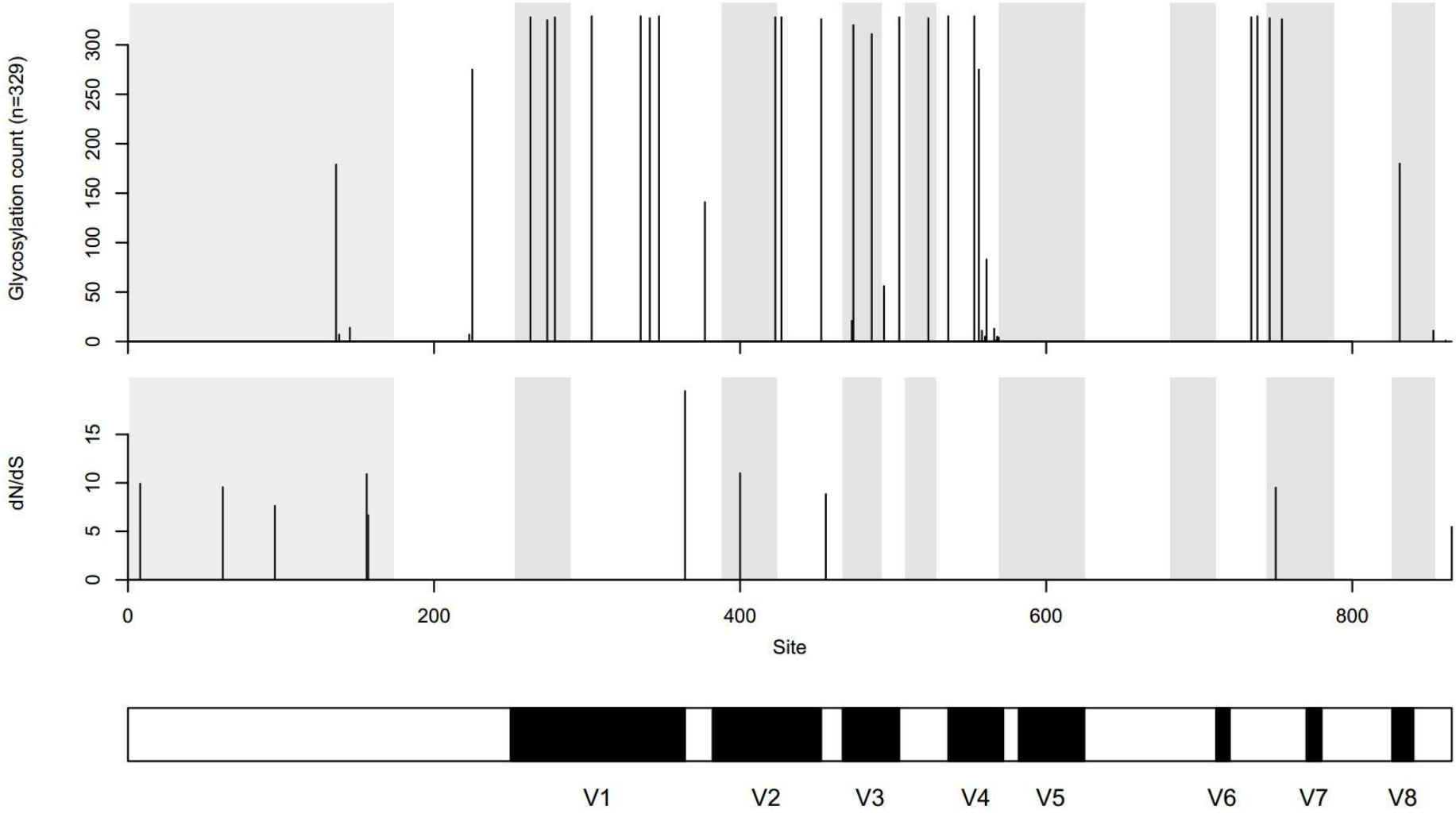


Figure 4
Click here to download Figure: Figure 4.docx

Figure 4

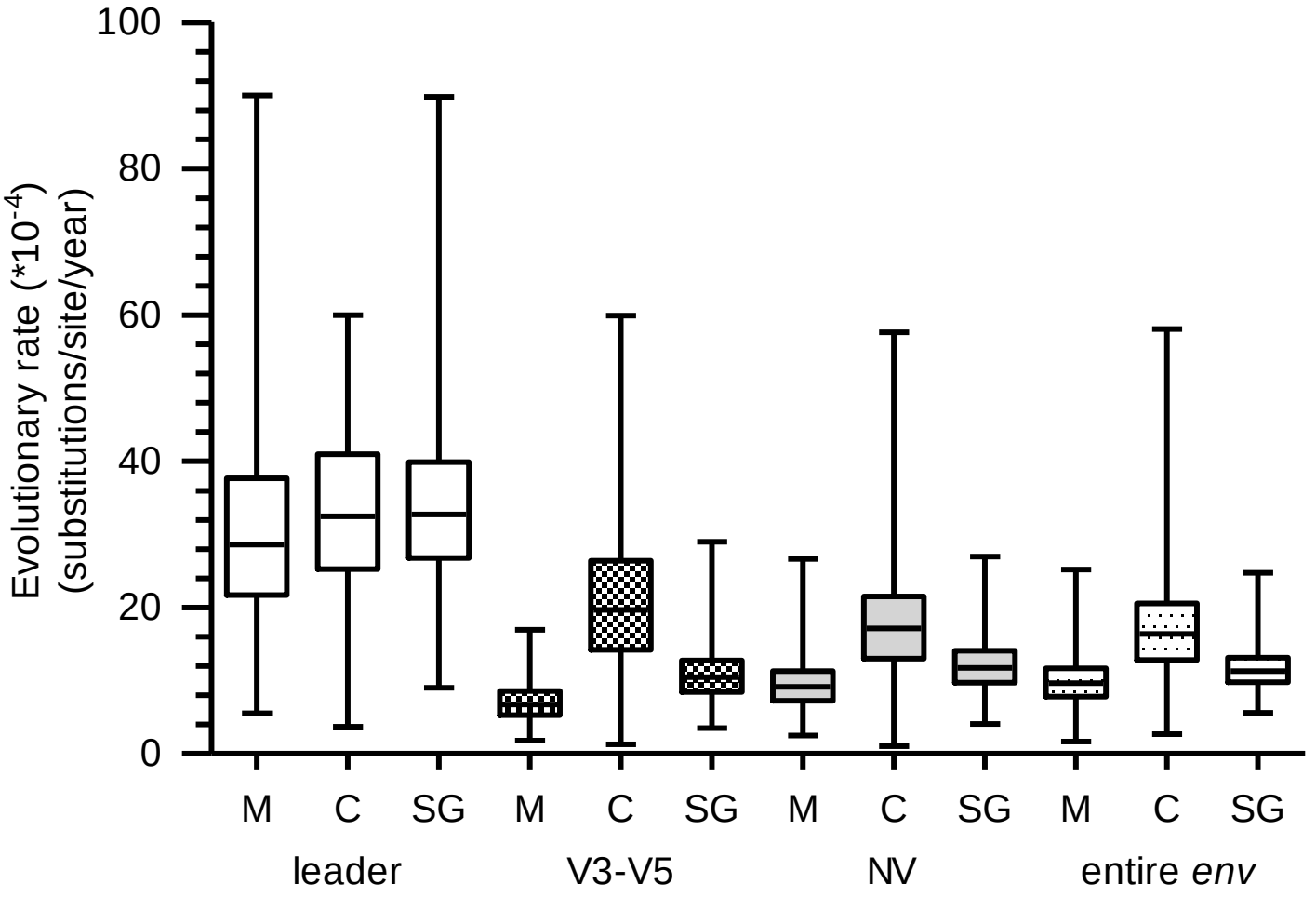
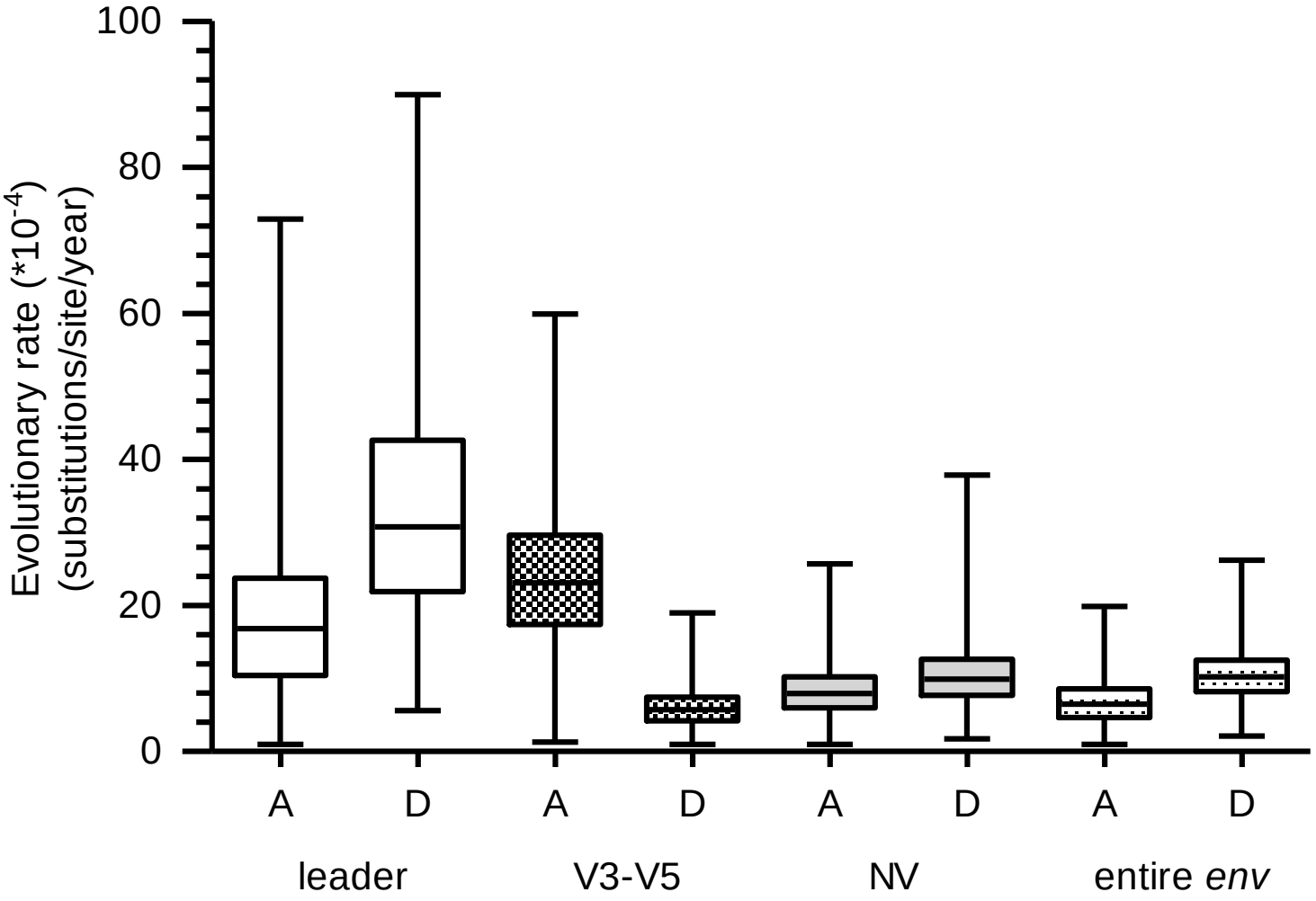


Figure 5
Click here to download Figure: Figure 5.docx

Figure 5



Supplementary tables and figures

Click here to download Supplementary Material Files: Supplementary material.pdf

Supplementary phylogenetic tree (in NEWICK format)
Click here to download Supplementary Material Files: Additional file 1.nwk