Virus Research 270 (2019) 197645

Contents lists available at ScienceDirect

Virus Research

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

The evolution of endogenous retroviral envelope genes in bats and their potential contribution to host biology

Emilia Cecylia Skirmuntt, Aris Katzourakis*

Department of Zoology, University of Oxford, Peter Medawar Building for Pathogen Research, South Parks Road, Oxford OX1 3SY, UK

ARTICLE INFO

Keywords:
Bat
EDI
ERV
Retrovirus
Envelope
Syncytin

ABSTRACT

Bats are the primary reservoirs and carriers of a wide range of viruses of unknown infectivity and pathogenic potential. Some of those if transmitted to other species can cause enormous economic losses in agriculture, and mortality in humans. Bats can be persistently infected with viruses while not showing any symptoms of disease, despite having high virus titre levels in their tissues and shedding virions for months or years after primary infection. It has been suggested that the lack of symptoms of viral infections and low mortality rate in bats might be due to immune adaptations that result from their long-term co-evolution with viruses. In this study, we screened all publicly available bat genomes from six bat families within which we have identified several envelope sequences of retroviral origin (gammaretroviruses). We analysed the identified sequences with Bayesian methods and maximum-likelihood inference to generate a phylogenetic tree with additional reference sequences of known endogenous and exogenous viral envelope genes. We also identified groups of orthologous viral envelopes and analysed them to determine if any of them might be an EVE (endogenous virus element) with an EDI (EVE- derived immunity) function or a candidate for a bat *syncytin* gene, which is an endogenized viral envelope, mostly known from its function in placentation in animals. Our study shows that bat genomes contain a substantial number of large, intact envelopes with open reading frames, which were found clustering closely on a phylogenetic tree reconstruction with *syncytin* sequences of other species. That might indicate that such sequences are good candidates for further bat-*syncytin*/EDI search.

1. Introduction

Viral genes in the host genome arise when viruses become permanently integrated into a host cell's genome, either through an obligatory step in their replication process (as with retroviruses) or chance events such as interaction with cellular retroelements or illegitimate recombination with host genomic DNA (Katzourakis and Gifford, 2010). If this integration occurs in germline cells, the virus can be transmitted vertically between generations in the form of an endogenous viral element (EVE[1]) or in the case of retroviruses, endogenous retrovirus (ERV[2]) (Katzourakis and Gifford, 2010; Patel et al., 2011). ERVs arising from ancient viral infections can persist and accumulate in the genomes of their hosts, constituting, for example, ~8% of the genome in humans (Lander et al., 2001), around 10% in mice (Chinwalla et al., 2002), and around 5% in bats (Zhuo et al., 2013).

The retroviral envelope (*env*) plays a crucial role in viral infection, controlling attachment to the receptors and cell entry (Lucas, 2010), as well as host immune evasion (Alcami and Koszinowski, 2000; Finlay and McFadden, 2006; Ploegh, 1998). The protein it encodes contains two main domains: surface (SU[3]) and transmembrane (TM[4]) - the SU subunit interacts with different receptors on the cell surface during the first part of the infection process. The TM subunit regulates fusogenicity - fusion of viral and cellular membranes during viral entry (Januszeski et al., 1997; Mangeney et al., 2007), and also contains an

Abbreviations: EVE, endogenous viral element; ERV, endogenous retrovirus; LTRs, long terminal repeats; EDI, EVE-derived immunity; SU, surface domain of envelope gene; TM, transmembrane domain of envelope gene; ISD, immunosuppressive region of TM domain; ORF, open reading frame; TE, transposable element

* Corresponding author.
E-mail address: aris.katzourakis@zoo.ox.ac.uk (A. Katzourakis).

[1] EVE -endogenous viral element
[2] ERV -endogenous retrovirus
[3] SU – surface domain of *envelope* gene
[4] TM –transmembrane domain of *envelope* gene

https://doi.org/10.1016/j.virusres.2019.197645
Received 19 February 2019; Received in revised form 26 June 2019; Accepted 27 June 2019
Available online 01 July 2019
0168-1702/

immunosuppressive region (ISD[5]): 17–20 amino-acids which help retroviruses in evasion of the host immune system (Blaise et al., 2001; Cianciolo et al., 1985; Mangeney and Heidmann, 1998; Mangeney et al., 2001; Morozov et al., 2012; Schlecht-Louf et al., 2010, 2014; Sonigo et al., 1986) (Supplementary materials – Fig. 2).

Multiple retroviral *env* genes exist in vertebrate genomes in the form of ERVs, some of which have a demonstrable function in the biology of the host, notably in the context of EVE-derived immunity (Aswad and Katzourakis, 2012). For example, sheep endogenous betaretroviruses both protect sheep against the Jaagsiekte sheep retrovirus through receptor interference, and are required in sheep conception and placental development (Arnaud et al., 2007a,b; Dunlap et al., 2006). One of the most well studied co-opted ERVs is an *env* with a TM domain and an ISD region - *syncytin*. *Syncytins* are retroviral *env* glycoprotein genes independently captured and endogenized by different animal lineages (Cornelis et al., 2014, 2015, 2017; Henzy et al., 2016; Lavialle et al., 2013). They can influence myoblast fusion and sexual dimorphism in some species (Redelsperger et al., 2016), and there is a possibility that some of them might be immunosuppressive and prevent mother-foetus conflict, but formal evidence is still missing (Haig, 2012; Malik, 2012; Pötgens et al., 2002). Syncytins have not been found yet in some of the orders of mammals like whales and dolphins (odd-toed ungulates), pigs (even-toed ungulates), elephants (proboscidean), or bats (chiroptera). Bats can be an interesting model for placentation and *syncytin* studies, because they have two types of placenta unevenly divided between both *Chiroptera* suborders. *Syncytins* are homologous to each other, but not necessarily orthologous (Lavialle et al., 2013), and types of placentation might differ according to the type of ancestral viral sequence.

Bats (*Chiroptera*) are the second largest order of mammals, representing about 20% of all classified species. *Chiroptera* can be divided into two sub-orders: Yinpterachiroptera and Yangochiroptera. Bats are also major reservoir for ˜100 different viruses, including viruses able to induce zoonotic diseases – infectious diseases of animals that can be transmitted to humans (Calisher et al., 2006; Omatsu et al., 2007; Plowright et al., 2015; Smith and Wang, 2013; Wang et al., 2011) The mechanism for how bats can be asymptomatic hosts for diseases deadly to other species (e.g. Ebola and rabies) is however not yet fully understood. It has been suggested that the lack of symptoms of viral infections and low mortality rate in bats might be due to immune adaptations that result from their long-term co-evolution with viruses (Baer and Bales, 1967; Chua et al., 2002; Chu et al., 2008; Cui and Wang, 2015; Field et al., 2001; Hanna et al., 2000; Johnson, 2003; Leroy et al., 2005; Mackenzie et al., 2003; Middleton et al., 2007; Taylor et al., 2010; Sulkin and Allen, 1974; Zhuo and Feschotte, 2015).

In this article, we describe the results of an extensive search for ERV envelopes, run on all 14 publicly available bat genomes (*Myotis lucifugus*, *Myotis brandtii*, *Myotis davidii*, *Miniopterus natalensis*, *Eptesicus fuscus*, *Pteropus alecto*, *Pteropus vampyrus*, *Roussetus aegyptiacus*, *Eidolon helvum*, *Megaderma lyra*, *Pteronotus parnelli*, *Hipposiderios armiger*, *Rhinolopus sinicus*, and *Rhinolopus ferrumequinum*). We looked for *envs* inside open reading frames (ORFs[6]), searched for orthology between identified sequences, performed selection analysis and reconstructed a phylogenetic tree for all the detected *envs* and reference sequences. To the best of our knowledge, a single comprehensive analysis of all of these genomes has not been previously conducted.

2. Materials and methods

2.1. Database search and sequence analysis

Retroviral endogenous envelope-like sequences were searched for by local BLAST (Altschul et al., 1990) in 14 bat genomes (*Myotis lucifugus* [GCA_000147115.1], *Myotis brandtii* [GCA_000412655.1], *Myotis davidii* [GCA_000327345.1], *Miniopterus natalensis* [GCA_001595765.1], *Eptesicus fuscus* [GCA_000308155.1], *Pteropus alecto* [GCA_000325575.1], *Pteropus vampyrus* [GCA_000151845.2], *Roussetus aegyptiacus* [GCA_001466805.2], *Eidolon helvum* [GCA_000465285.1], *Megaderma lyra* [GCA_000465345.1], *Pteronotus parnelli* [GCA_000465405.1], *Hipposiderios armiger* [GCA_001890085.1], *Rhinolopus sinicus* [GCA_001888835.1] and *Rhinolopus ferrumequinum* [GCA_000465495.1]) with 9 probes of known exogenous and endogenous retroviral envelope glycoproteins (Supplementary materials-probes).

From local BLAST search results, using Python (ver. 2.7) custom scripts we selected sequences of length ≥100aa with an e-value ≤0.09. We extracted contigs containing selected sequences from assembled genomes and searched them for open reading frames (ORF) ≥200aa in length (from one stop codon to the next stop codon). Use of the Bedtools package (Quinlan, 2014) allowed us to select 1447 unique sequences from all bat genomes complying with the above requirements. We ran reciprocal BLAST against the chosen sequences and discarded all which did not return results for gammaretroviruses with the first hits. We manually inspected sequences and discarded ones which did not contain the highly conserved C-X_{-6}-C-C motif or were duplicates. We used NCBI's Conserved Domain search tool (Marchler-Bauer et al., 2014) to determine the exact placement of the *env* sequences and what other viral genes and domains can be found in the closest genomic neighbourhood.

2.2. Sequence alignments and phylogenetic tree reconstruction

We manually aligned the 737 chosen *env* TM regions using the Aliview software (Larsson, 2014). Putative *env* sequences with accompanying genes were sorted into categories based on the presence of conserved TM domains such as a cleavage site, hydrophobic domain, and C-X_{-6}-C-C motif. We aligned sequences according to the generated consensus sequence consisting of the most conserved motifs (78 aa) – R-X-K-R-[X_{-46-64}]-S-Q-T-L-V-Q-D-Q-V-D-S-L-S-E-V-V-L-Q-N-R-R-G-L-D-L-L-T-A-E-K-G-G-L-C-A-A-L-G-E-E-C-C-F-Y-V-N-Q-S-G-L-V-R-D-[X_{-27-46}]-I-L-P-F-L-G-P-L-I-V-L-L-M-I-L-L-F-G-P-C-I-L-[X_{-2}]-L-L-L-X-L-X-F-G-P-C-X-[X_{-17}]. In the final alignment, we included only those sequences, which had conserved motifs present at least in 50% (≥ 38 aa).

To reconstruct a phylogenetic tree of retroviral *env* sequences in 14 bat genomes, known mammalian *syncytin* sequences, and a selection of the closest viral matches for both endogenous and exogenous viruses [*Supplementary materials- Reference sequences*], we first ran Prottest (Abascal et al., 2005) to determine which model to use with the RAxML ver. 8 package (Stamatakis, 2014), with the best model being the JTT + G model under AICc scoring criteria. We used MrBayes software (Huelsenbeck and Ronquist, 2001) with the SRD06 substitution model (Shapiro et al., 2005) for reconstructing a phylogenetic tree of all identified orthologous sets with reference sequences.

2.3. Orthologous sequence search and annotation

We selected all sequences ≥ 400 aa in length. Sequences long enough and containing the smallest number of gaps were further extended to generate flanking regions, which we subsequently ran in BLAST searches against the Refseq and WGS *Chiroptera* databases. We then compared the results to the accession numbers on the phylogenetic tree. From the bat genome databases, we extracted all the sequences marked as orthologous and manually annotated them in Geneious (version 10.1.3) (http://www.geneious.com, Kearse et al., 2012).

2.4. PAML analysis of orthologous sequences

We tested 31 sequence alignments with long envelope sequences using the CODEML algorithm from the PAML: Phylogenetic Analysis by

[5] ISD – immunosuppressive region of TM domain
[6] ORF- open reading frame

Maximum Likelihood package (Yang, 1997). We ran several different models for each set of orthologs. The one-ratio test (M_0 – single ω ratio for whole tree) was run for all sequence sets. If all sequences from the set contained stop codons or frame shifts we analysed only part of the sequence starting from the 5′ end stop codon or frame shift to the next stop codon/frame shift, or the end of the sequence. In the case of sets that contained only two orthologs we ran M_0 in pairwise mode. To check for the possibility of neutrality of the whole sequence we ran M_0 with ω fixed to 1 and compared log-likelihood ratios (LRTs) for M_0 ω = 1 vs. M_0. In sets with three or more sequences we ran additional analyses for M_1 (separate dN/dS ratios for each branch of the tree) and M_2 (giving two or more dN/dS ratios to appointed branches). LRTs calculated for M_0 vs. M_1 and M_0 vs. M_2 showed that the one-ratio model is the best fit for all the sequence sets. For sets of three or more sequences we ran NSsite models - M1: neutral (two-state, ω > 1 disallowed), M2: positive selection (two states, ω > 1 allowed), M3: discrete, M7: beta (beta distribution, ω > 1 disallowed), M8: beta& ω > 1 (beta distribution, ω > 1 allowed). We compared log-likelihood ratios for our data: M1 vs. M2, M0 vs. M3 and M7 vs. M8. The level of significance for LRTs was calculated using a $\chi 2$ table, with twice the difference of log likelihood between the models (2ΔlnL) compared to a $\chi 2$ distribution. The number of degrees of freedom (df) corresponds to the difference in number of parameters between the models [*Supplementary information-dNdS PAML*].

3. Results

3.1. Search for gammaretroviral envs

Through local BLAST searches with 9 retroviral envelope glycoprotein probes [*Supplementary materials-probes*], we initially identified 51,245 potential envelope sequences in the 14 bat genomes. We selected sequences with an *env* length of ≥400 amino acids (aa), which were in ORFs. We manually checked sequences for the presence of an ISD and ran reciprocal BLAST on them to attest if the sequence is of gammaretroviral origin, choosing all the sequences with a gammaretrovirus as first hit in the BLAST search. Sequences with long flanking regions (˜500 base pairs on both ends) were extended and run in BLAST against the Refseq and WGS *Chiroptera* database to find orthologs. The quality of the data in some contigs prevented us from using some portions of the chosen sequences because of the presence of large gaps. Low quality data with too many gaps or missing flanking regions means that a number of orthologous sequences likely remain undetected. In general, species with the highest number of identified sequences were also the genomes with the highest data quality and the lowest number of gaps (*P. vamprus, M. lucifugus, H. armiger, R. aegyptiacus*), species with the fewest number of sequences in the global alignment were from

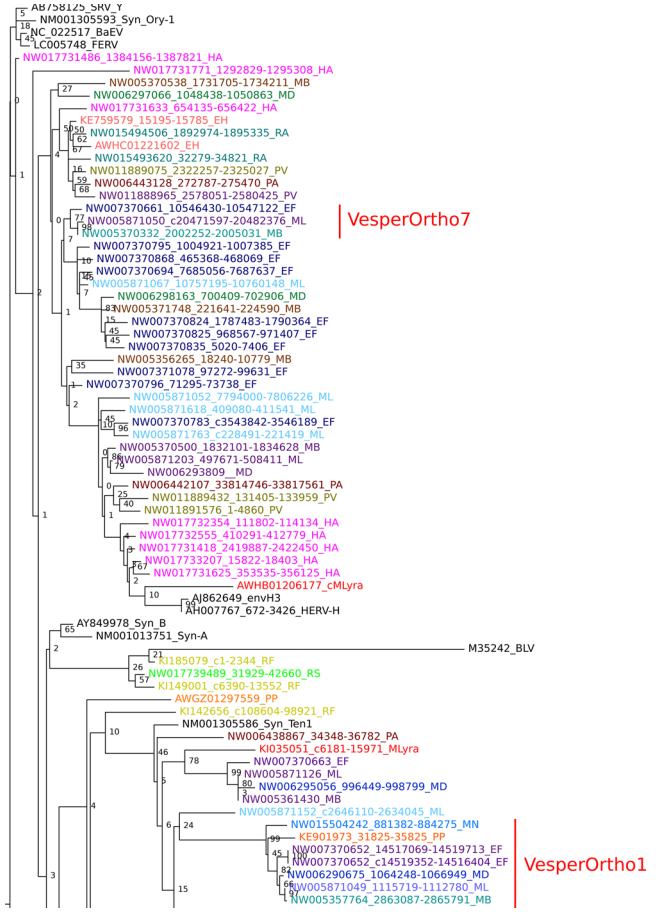

Fig. 2. Reconstructed phylogenetic tree of orthologous gammaretroviral envelope sequences found in 14 species of bats (marked with different colours). Branches containing single species or large clusters of single family sequences collapsed for clarity (list of all the sequences in *Supplementary materials – Sequence List*).

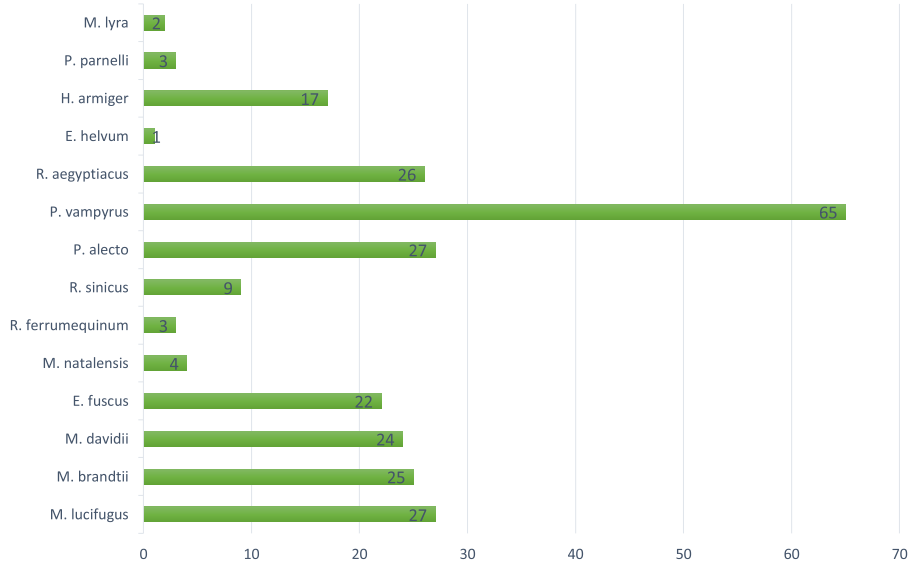

Fig. 1. Number of envelope sequences in each bat species with *env* of length ≥400 aa.

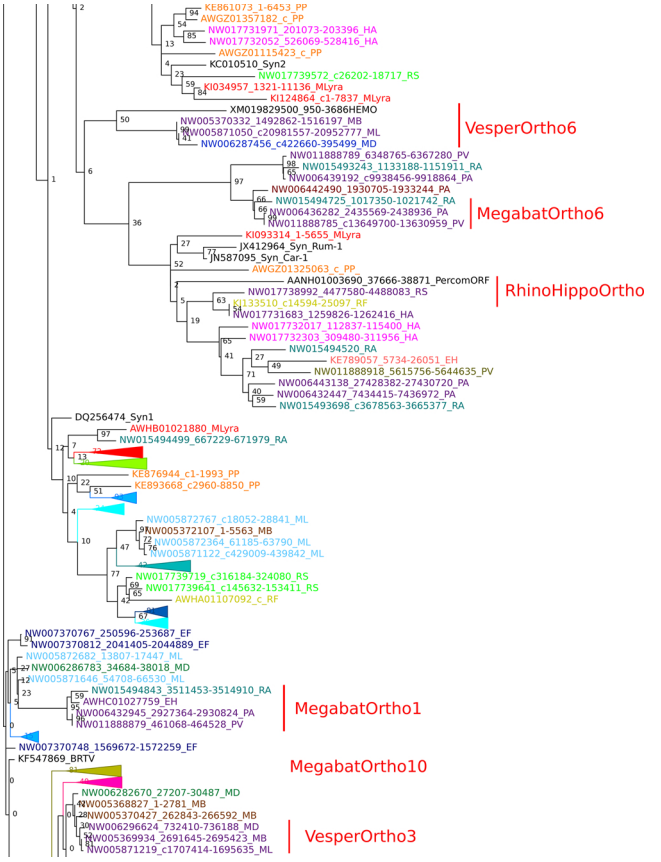

Fig. 2. (*continued*)

the assemblies with the lowest data quality (Fig. 1) [*P. parnelli, E. helvum, M. natalensis*) [*Supplementary materials- Assemblies stats*].

3.2. Retroviral envelope phylogenetic tree reconstruction

All the sequences found by us shared several features of gammaretroviral *env,* such as putative furin/PACE cleavage sites between SU and TM subunits, hydrophobic domains (involved in fusogenic activity of *env* –anchoring it to the plasma cell membrane), and the highly conserved disulphide bond (C-X_6 -C-C) characteristic of gammaretroviruses (Blaise et al., 2001; Sinha and Johnson, 2017).

Despite poor data quality in some genomes, we reconstructed a phylogenetic tree of bat gammaretroviral envelope sequences, and based on that we searched and found several examples of orthologous sequences (Fig. 2). On the reconstruction of the phylogenetic tree most of the branches contained clusters of sequences from the same bat species or same families.

3.3. Orthologous sequence identification

From all of the identified envelope sequences containing ISDs, we were able to identify 34 orthologous sets, although with 3 sets containing TM too short for further analysis. We did not find any orthologous sequences shared between Yinpterachiroptera and Yangochiroptera, but that may be due to the poor data quality. The split between Yinpterachiroptera and Yangochiroptera is dated to ~58.9 Mya, and divisions between families within the two suborders occurred around 54 and 53.8 Mya respectively. Based on our results, we assume none of the orthologs found by us are older than the splits between suborder families. Most identified sequences were in the vesper bat species or the *Pteropodidae* species and were not coding competent with a no significant evidence of selection. These sequences date back no later than to the vesper family split (~15.5 Mya), and the split between *Pteropodidae* species (~22.6 Mya). We detected two pairs of sets that are shared between families, one with *Rhinolophidae* - *Hipposideridae* sequences (~36.9 Mya), and one which was found in representatives of three Yangochiroptera families- *Vespertilionidae, Mormoopidae* and *Miniopteridae,* which sets the suspected date of endogenization no later than the suborder family division (~53.8 Mya) (Fig. 3) (Agnarsson et al., 2011).

We also reconstructed a phylogenetic tree that shows all identified true orthologs along with reference sequences (Fig. 4). All sets of orthologs were marked with different colours to distinguish them from each other. Some of the sequences do not cluster with their orthologs due to too short sequences in the alignment (seen in sets MegabatOrtho7, 11, 15, 17, 19) but further analysis on the rest of the coding sequences and flanking regions confirmed that those are still orthologous. The overall architecture of both trees and placement of orthologues sequences in relation to reference sequences correspond broadly with each other.

RhinoHippoOrtho along with MegabatOrtho4 and 12 all contain three sequences and cluster together with ruminant and carnivore *syncytins* as well as *syncytin* –like sequences found in spiny-rayed fish. The only other Yinpterachiroptera sequence clustering together with reference sequences is MegabatOrtho19 where *P. alecto* branch with reference sequences of exogenous viruses. This is probably caused by the one of orthologs being too short within the fragment used for an alignment to cluster together with its true orthologous sequence. VesperOrtho1 and VesperOrtho5 are the largest sets of orthologs we found with VesperOrtho1 containing sequences from three different bat families: *Vespertilionidae*, *Miniopteridae* and *Mormoopidae*. They both cluster together but also with tenrec *syncytin* and primate *syncytin* 2. VesperOrtho6 contains 3 vesper bat sequences and lays on the tree close to newly found *syncytin*-like sequence expressed in human placenta –HEMO. VesperOrtho7 (2 sequences) and VesperOrtho8 (3 sequences) branch together with endogenous viral sequences HERV-H and envH3.

3.4. dN/dS analysis for orthologous sequences

dN/dS analysis (ω ratio) is a calculation of the number of non-synonymous substitutions (alteration of the amino acid sequence) to synonymous substitution (changes in genetic code which do not alter the translated amino acid) in a sequence. It gives an indication of how strong the effect of natural selection is on a given protein. A higher number of non-synonymous mutations than synonymous ones suggests that the sequence is under positive selection ($\omega > 1$), while a higher number of synonymous mutations than non-synonymous ones indicates negative purifying selection ($\omega < 1$), while $\omega = 1$ is consistent with neutral evolution. Non-synonymous changes can alter the function of the gene, and are more likely to be deleterious, hence in coding sequences their number tends to be lower than synonymous ones. Functionally conserved sequences usually will have a higher number of synonymous mutations than non-synonymous ones. In our orthologues, we found representatives for all three types of ω ratio (Table 1**).** We also determined statistically whether the whole sequence is consistent with neutrality. To do that we calculated the likelihood ratio test (LRT), a statistical test which compares the fit of two statistical models, for the one-ratio model (null model) and the one-ratio model with ω fixed to 1 (alternative model). That allowed us to reject neutrality in some of our sequences, with the degrees of freedom equal to 1 for all the tests.

We tested 31 ortholog sequence alignments using the PAML: Phylogenetic Analysis by Maximum Likelihood package (Yang, 1997). PAML analyses are highly accurate in detecting adaptive molecular evolution, as well as positive selection acting on specific sequence sites (Anisimova et al., 2001,2011). We ran a one-ratio model (M_0) test for all ortholog sets – a strict model that allows a single dN/dS ratio for all branches. According to this model, the majority of the tested sequences showed evidence of purifying selection. Purifying selection works by

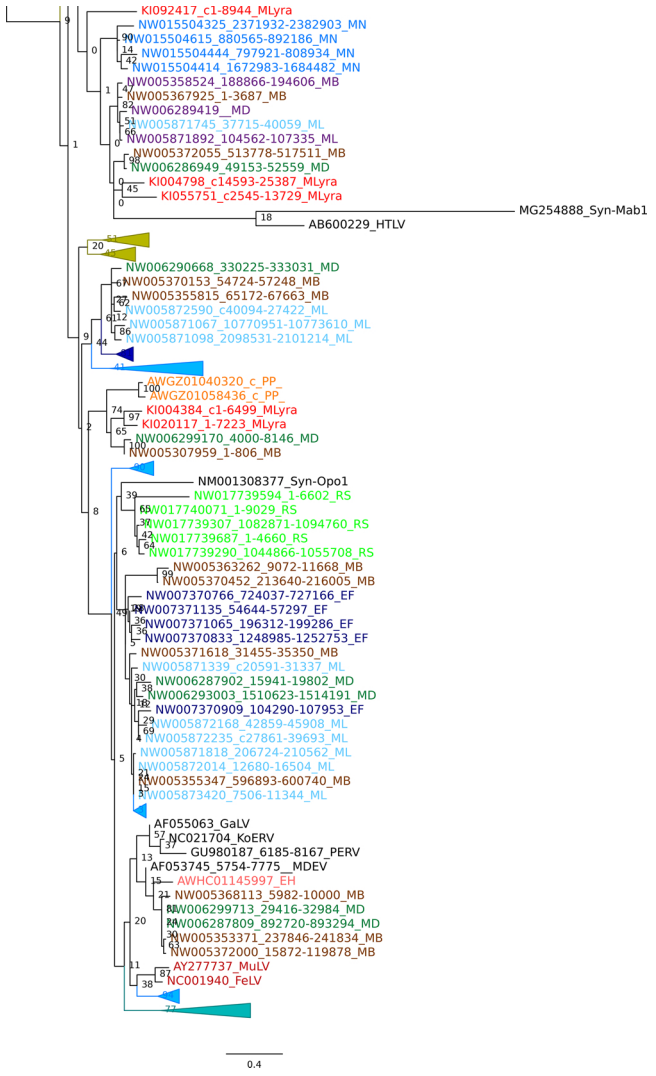

Fig. 2. (*continued*)

maintaining the function of the gene, meaning that most coding genes will have a dN/dS below 1. The rest of the tested sets show evidence of positive selection – indicating continuous adaptive evolution, and some show evidence of neutral evolution.

The likelihood-ratio test for neutrality of the whole sequence allowed us to reject neutrality in 6 out of 31 sequence sets (VesperOrtho1 and 9, MegabatOrtho3, 6, 19, and Rhino&HippoOrtho). All these alignments show evidence of purifying selection.

For all ortholog sets with the number of sequences greater than 2 (15 sets) we ran additional site models (M_0, M_1 and M_7) and compared neutral models to the selection ones (M_2 and M_8). We also ran M_3 - a discrete model, which allows two classes of sites to have dN/dS > 1. From these we compared one-ratio model (M_0) with the discrete model (M_3), neutral model (M_1) with positive selection model (M_2), and beta model (M_7) with beta&ω > 1 (M_8). All values and results are listed in *Supplementary information-dNdS PAML*.

3.5. Selected ortholog sequence analysis

We selected ortholog sets which contained full *env* (from start to stop codon) or were branching closely to reference sequences. This gave us six sets: VesperOrtho1, 3, 7 and MegabatOrtho1, 6 and 10. On further analysis, we also selected VesperOrtho6 which branches closely to the HEMO reference sequence and RhinoHippoOrtho which contains representatives of the two families. Six alignments were annotated using Geneious (ver.10.1.3) (Kearse et al., 2012). All of these sequences contain all canonical motifs found in gammaretroviral *env* proteins: a hydrophobic signal peptide in SU region, furin cleavage site between SU and TM subunits (R/K-X-R/K-R), fusion peptide, immunosuppressive domain, C-X_{6}-C-C motif, and transmembrane domain.

VesperOrtho1 is a set of six orthologs from *M. lucifugus, M. davidii, M. brandtii, E. fuscus, M. natalensis* and *P. parnelli* bats, spanning three Yangochiroptera families (*Vespertilionidae, Miniopteridae, Mormoopidae*). The *env* alignment is 480 aa in length and 57.2% of the positions in these sequences are identical to each other (Supplementary materials – Fig. 3). We observed the highest similarities between vesper bat species, with most notable aa changes in *P. parnelli* and *M. natalensis* in both of the canonical motifs and the rest of the sequence. *M. natalensis*has a 5 aa gap (position 269–274) before the fusion peptide and lacking threonine/methionine on position 409, *M. davidii* lacks glutamine on position 364. There are no frameshifts or stop codons in the middle of the *env* sequence, but *P. parnelli* does not contain a stop codon at the end of the coding sequence, instead having a lysine. In close genomic proximity to the sequence (˜300 aa) we found a fragmented *pol* region - with reverse transcriptase, RNase and integrase in *M. brandtii*, reverse transcriptase and RNase in *M. natalensis*, and RNase in *M. lucifugus, M. davidii* and *E. fuscus*. In *P. parnelli* we detected only partial RNase but that is probably due to the genomic scaffold being truncated on one end. We could not detect LTRs on both ends of the sequences although this might be caused by poor data quality in some of the sequences. This *env* is under purifying selection (*Supplementary materials – dNdS*).

VesperOrtho3 is an ortholog containing 3 sequences from the *Myotis* genus: *M. brandtii, davidii* and *lucifugus* (Supplementary materials – Fig. 4). The alignment contains 90.5% aa that are identical and contains all canonical *env* features with no frame shifts or stop codons and only single amino acid changes throughout the 593 aa length of the sequences. We detected the presence of *gag* and *pol* proteins as well as a zinc finger in close proximity to *env*. The dN/dS for this gene is 0.9698 which suggests neutral evolution, and we could not reject neutrality statistically (*Supplementary materials – dNdS*). All three sequences are flanked by LTRs. The dN/dS score and the presence of complete retroviral proteins between LTRs is consistent with neutrality.

VesperOrtho7 is a set of two *Myotis* bat sequences (*M. brandtii* and *M.lucifugus*), which are almost identical to each other (98.1%) with only 10 single aa substitutions and no frame shifts or stop codons (Supplementary materials – Fig. 6). The full *env* (537 aa) contains all canonical motifs and in its closest proximity we detected fragmented *pol* and *gag* proteins, containing frame shifts and stop codons. All glycoproteins are flanked by LTRs. The dN/dS for those sequences is 0.4374, which implies purifying selection. However the neutrality test for the whole gene was inconclusive with a borderline result (Supplementary materials – *dNdS*). On the 3′ flanking region, ˜1000aa from the *env* sequences, we detected the presence of the transposable element.

MegabatOrtho1 contains full *env* sequences (550 aa) from 4 megabats: *P. alecto*, *P.vampyrus*, and *R. aegyptiacus* and *E. helvum* (Supplementary materials – Fig. 7). These sequences are 81.6% identical to each other. All *env* canonical motifs are present, and there are no frameshifts or stop codons in the alignment. We observed a gap in position 192 instead of arginine in *E.helvum* and *R.aegyptiacus*, and the biggest differences in aa changes in those two bats. All bats have *gag* and *pol* sequences present and LTRs on both ends of the viral protein sequences, apart from *E.helvum* due to a truncated contig at the 5′ end, which contains only a partial *pol.* According to dN/dS analysis the sequence seems to be under purifying selection, as we could reject the possibility of neutrality for the whole gene (*Supplementary materials – dNdS*).

MegabatOrtho6 is an alignment containing full *env* sequences of three Yinpterachiroptera bats: *P. alecto*, *P. vampyrus*, and *R. aegyptiacus*, from the family *Pteropodidae* (Supplementary materials – Fig. 8). The *env* sequences are 625 aa long, contain all canonical *env* regions, and

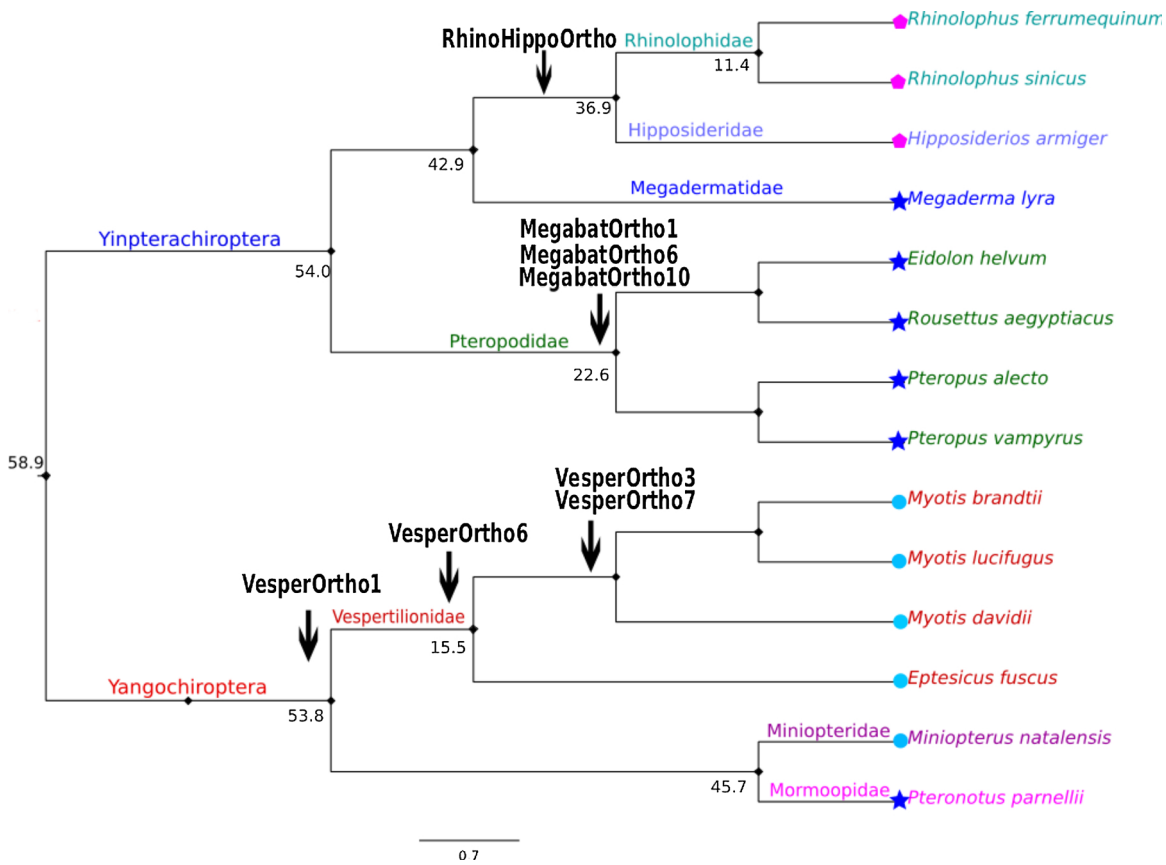

Fig. 3. Simplified cladogram of the phylogeny of analysed bat species, with estimated dates (numbers at nodes) of family splits (after Agnarsson et al., 2011) and types of placentation in each of the species. Dates are all in millions of years ago (Mya). All analysed bats have discoidal type placentation. Bats marked with dark blue dots have mesometrial haemochorial type placentation, light blue dots indicate antimesometrial haemochorial type placentation, and pink dots indicate mesometrial endotheliochorial type placentation.

show 89.1% identity to each other. We observed just minor single aa changes across the whole sequence, including in the region of fusion peptide, ISD and TM region. These were mostly seen in *R. aegyptiacus*. There are no frame shifts or stop codons across the whole coding *env*. Apart from an envelope length in the range of ˜500aa, we found fragmented *gag* and *pol* glycoproteins, as well as short LTR on both ends of all viral glycoproteins. dN/dS shows purifying selection with neutrality being rejected (*Supplementary materials – dNdS*).

MegabatOrtho10 is another ortholog set containing only 2 sequences (*P. alecto* and *P. vampyrus*) (Supplementary materials – Fig. 9). A possible insertion between the same flanking regions was detected also in *R.aegyptiacus* but containing only the pol glycoprotein. In both sequences containing a full *env* sequence there is also fragmented and frame shifted *gag* and *pol*, along with a zinc finger domain. All viral proteins are between long but fragmented LTRs. *Env* is 516 aa long, 97.9% identical for both sequences with only 11 single aa substitutions. One of the substitutions is present in the C-X_6-C-C motif of ISD, where in *P. vampyrus* one of the cysteines is substituted by tyrosine. The sequence is under positive selection according to dN/dS result, although the neutrality test for the whole sequence is inconclusive with a borderline value (*Supplementary materials – dNdS*).

We identified vesper ortholog sequences branching closely with the HEMO reference sequence, which is an endogenous retroviral envelope protein seen in the blood of pregnant women, pluripotent stem cells, and tumours, but with unusual immunosuppressive domain not seen in other retroviruses (Heidmann et al., 2017). Assigned as ortholog set VesperOrtho6, it contains 3 sequences from all 3 available *Myotis* bat genomes. We found a partial sequence in *E. fuscus,* but it is too fragmented to align with the others, containing large strings of undefined nucleotides. Close to the *env* sequence there is also a fragmented partial *pol* gene. The *env* sequence itself can be identified in *M. brandtii* and *M. davidii,* but it is very short (138 aa) and it contains only the 3′ end part of the TM domain with an ISD. The *M. lucifugus* genomic region is too fragmented to align with other *env* sequences but it has a *pol* region. All sequences are flanked by LTRs. As a reference in our alignment we used cat HEMO sequence ISD (LQN-HHLDLSAAQQGRT), as their HEMO is most closely related to chiropteran HEMO (Heidmann et al., 2017). The bat sequence branching with the reference showed similarities both in aa comparison (Supplementary materials – Fig. 5) with 8 aa same as in the 17 aa in cat HEMO ISD, and they grouped together in a phylogenetic tree reconstruction from the nucleotide alignment.

RhinoHippoOrtho is one of two ortholog sets containing sequences from two different bat families (*Hipposideridae* and *Rhinolophidae*). *Env* sequences from three bats are 405 aa long but end prematurely on the 5′end. It contains a stop codon in *H.armiger* on the 5′ end and 3′ end in *R.sinicus* and *R.ferrumequinum.* Fusion peptide, ISD and transmembrane region are present in all three sequences. Despite a short sequence dN/dS is low suggesting purifying selection, with neutrality rejected for the whole sequence [Supplementary materials – *dNdS*]. Close to the env we also detected a partial pol glycoprotein, with both being flanked by LTRs. Low dN/dS with only partial *env* sequence recovered from the data might be a result of this sequence having previously been functional, with the function lost over time, and remnants of the sequence left as a viral fossil.

The chosen ortholog sets present two different forms of immunosuppressive domains [*Suppelementary materials* – Figs. 3–9]. MegabatOrtho1, VesperOrtho 3 and 7 show variants with high levels of similarity to the 17 amino acids of the consensus immunosuppressive domain (e.g. FELV, PERV, MULV) with all of the aforementioned ISDs similar to the consensus in 12–13 aa (Bénit et al., 2001). VesperOrtho1

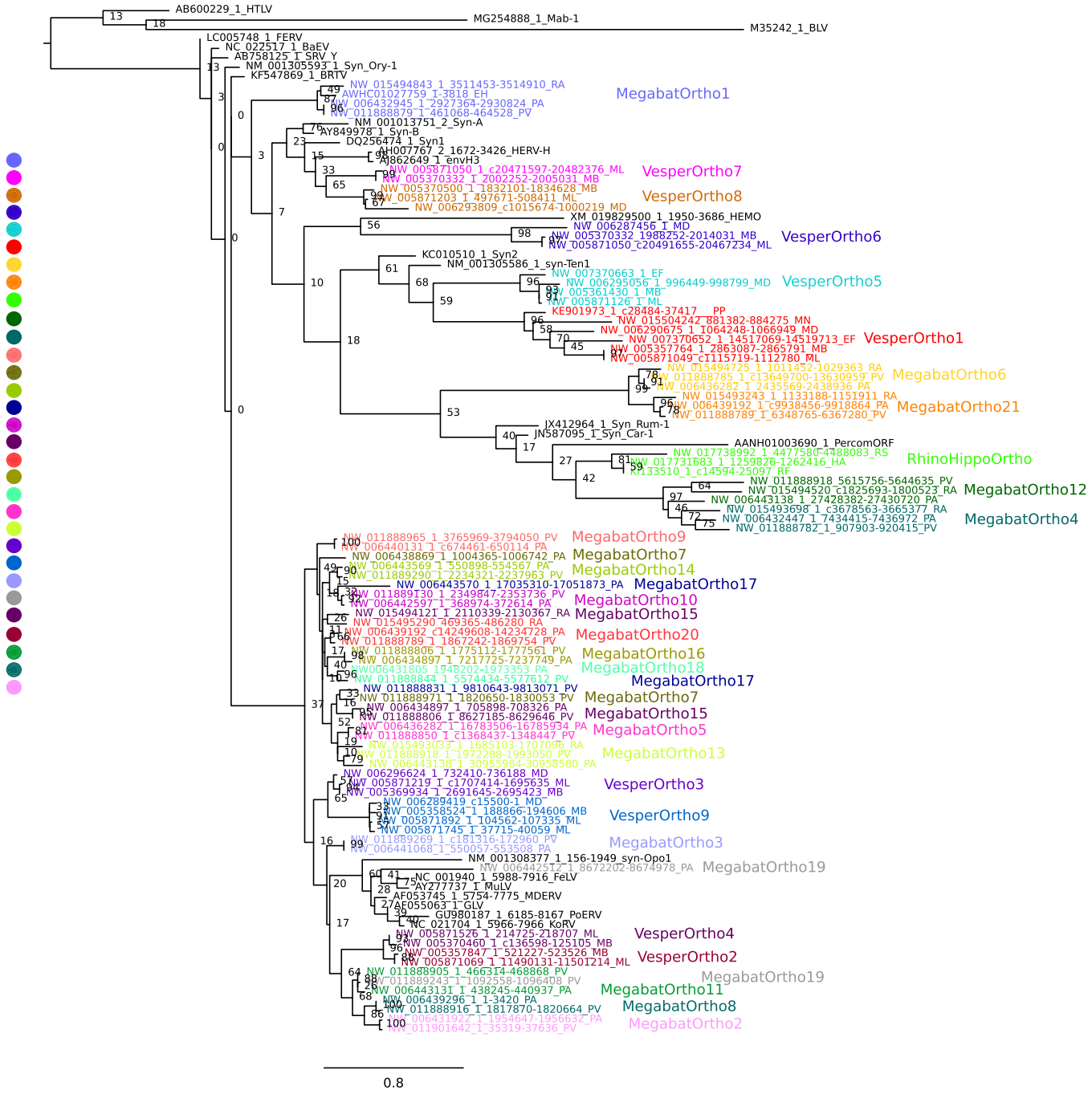

Fig. 4. Reconstructed phylogenetic tree of orthologous gammaretroviral envelope sequences found in 14 species of bats and reference sequences. *Syncytin* sequences and other reference sequences are marked with black colour. Orthologous sequences are marked with one of 31 colours – numbers in the legend correspond to the positions on the tree.

and MegabatOrtho6 contain immunosuppressive domains that are more similar to those seen in squirrels and woodchucks (Env-Mar1) (Redelsperger et al., 2014a, b) or HERV4-1 or ERV-3 (Bénit et al., 2001; Bustamante Rivera et al., 2018; Mangeney et al., 2007) with levels of similarity going up to 16/17 aa when comparing MegabatOrtho6/Env-Mar1 and to 11/17 aa when comparing VesperOrtho1/Env-Mar1 ISD. Both kinds of ISD can be seen in species with endoteliochorial (e.g. cats, dogs) (Cornelis et al., 2014; Leiser and Koob, 1993) and hemochorial placentation (e.g. humans, mice, rabbits, rats) (Blaise et al., 2003; Denner, 2016; Dupressoir et al., 2005; Esnault et al., 2013; Heidmann et al., 2009).

4. Discussion

Despite poor data quality in the case of some of the analysed genomes our study shows that bat genomes contain at least 32 sets of viral derived sequences of unknown function that could potentially be *syncytins* and/or EDIs. Large, intact *envs* in open reading frames were found clustering closely on a phylogenetic tree with *syncytin* sequences of other species, which might indicate that such sequences are good candidates for further bat *syncytin*/EDI research.

We detected 1447 long ORFs overlapping BLAST matches to *env* sequences (Supplementary figures – Fig. 1). Most of these are probably a result of relatively recent retroviral integrations, but could also represent preserved ancient coding sequences. After entering the host

Table 1
Ortholog sets with dN/dS results of one-ratio model.

Ortholog set	n = #	dN/dS (one-ratio)	Sequence size (aa)	2ΔlnL (for df = 1/p-value = 0.05)
MegabatOrtho1	3	0.7799	550*	2.23
MegabatOrtho2	2	0.8264	107	0.040764
MegabatOrtho3	2	0.4561♣	561	7.23193†
MegabatOrtho4	3	1.1498	186	0.278162
MegabatOrtho5	2	0.3708♣	102	1.326184
MegabatOrtho6	3	0.5345♣	626*	10.974878†
MegabatOrtho7	2	1.0502	125	0.011084
MegabatOrtho8	2	2.4412	231	0.374048
MegabatOrtho9	2	2.7831	185**	1.135522
MegabatOrtho10	2	4.5747	517*	3.27657
MegabatOrtho11	2	0.6593	109**	0.302754
MegabatOrtho12	3	1.0074	161	0.000512
MegabatOrtho13	3	1.2891	202	0.53744
MegabatOrtho14	2	0.9991	476***	0.000008
MegabatOrtho15	3	0.6208	170	2.04
MegabatOrtho16	2	1.7135	156	0.706314
MegabatOrtho17	2	0.5471♣	190**	2.84
MegabatOrtho18	2	1.9117	113	0.389932
MegabatOrtho19	2	0.3652♣	434	8.32724†
MegabatOrtho20	3	0.8718	183**	0.179014
MegabatOrtho21	3	2.6365	117	2.289654
VesperOrtho1	6	0.4790♣	480*	61.389886†
VesperOrtho2	2	2.1077	82	0.49425
VesperOrtho3	3	0.9698	594*	0.015048
VesperOrtho4	2	0.7137	452***	0.845898
VesperOrtho5	4	0.7001	415***	2.04
VesperOrtho6	3	0.9430	138	0.11324
VesperOrtho7	3	0.4374♣	538*	3.22
VesperOrtho8	3	1.2085	246	0.23843
VesperOrtho9	3	0.4714♣	182	4.93258†
Rhino&HippoOrtho	3	0.5920♣	405	11.009006†

* full envelope from start to stop codon.
** only TM domain analysed.
*** stop codon inside TM domain.
♣ dN/dS value significantly below 1.0.
† significant result.

genome through vertical transfer, endogenous viruses start evolving at the host rate of mutation rather than that of their viral progenitors (Gilbert and Feschotte, 2010). This leads to two destinies for such sequences - accumulation of mutations, with partial loss of the coding sequence and a slow 'decay' if its function is harmful or non-advantageous for the host's survival; or fixation and preservation in genomes of the host population if the sequence provides a selective advantage for the host (Feschotte and Gilbert, 2012). It is unclear what proportion of ERVs seen today might have gone through either of these routes. Seeing large numbers of viral sequences in ORFs most probably means that these are recent integrations in the genome that have not started or have just begun deactivating and decaying. However, since the probability of maintaining an intact ORF for an extended period of time without purifying selection is low (Katzourakis and Gifford, 2010), several of the identified *envs* in ORFs could be under purifying selection. ORFs of length higher than 1k aa are most probably due to sequencing mistakes in a genome and large gaps of unidentified bases in the middle of the sequence, as this is longer than any known retroviral *env* (Bénit et al., 2001; Henzy and Johnson, 2013). Of the all identified *envs* in ORFs, 737 were suitable for further analysis, containing almost full envelope sequences which allows them to be included in the final alignment and phylogenetic analysis.

On our reconstructed phylogenetic tree, we observed many branches with the sequences restricted to a single bat species (Fig. 2). This suggests that the given sequences came from a recent integration event that occurred after speciation. The same is true for clusters containing sequences of only *Pteropodid* bats (*P. alecto, P. vampyrus, R. aegyptiacus, E.helvum* - which split from other bat families around 22.6 Mya) or only vesper bats (*M. lucifugus, M. davidii, M. brandtii, E. fuscus* - which split from other Yangochiroptera around 53.8 Mya) (Agnarsson et al., 2011). These types of subtrees have been collapsed for clarity in the tree. Branches containing different bat species and reference retroviral or *syncytin* sequences are of particular interest; they might contain sequences of similar architecture to known ERVs and EDIs dispersed amongst different families, indicating integrations predating speciation events and preserved in a potentially coding state for a long period of time, which might mean that they have a recent function in the host.

Of all the identified *envs* in ORFs, 84 candidates in 32 sets were considered further as potentially functional host genes on the basis of presence of all gammaretroviral canonical motfis and the immunosuppressive domain, as well as being present in more than one analysed bat genome. From all of those we want to distinguish two specific sets which display the most promising features for further analysis. VesperOrtho1 is a set of 6 sequences from 3 different Yangochiroptera families, containing a full envelope sequence, and displaying full canonical features of gammaretriviral *env*. dN/dS analysis showed that the sequences are under purifying selection. The other candidate is the MegabatOrtho6 set of 3 Yinpterachiroptera full envelope sequences showing all the gammaretroviral features and evidence of purifying selection. In close proximity to both of these sets we found fragmented *gag* and *pol* genes, a feature that has also seen before in case of many of the known *syncytin* genes. Because orthologs from Yinpterachiroptera and Yangochiroptera cluster with reference *syncytins* from different types of mammal, we speculate that those potential candidates might be responsible for the presence of different types of placentation seen in bat species. We found several orthologous sequences in different bat families, suggesting some of them predate family splits and remain conserved, which may indicate function in a

genome.

Interestingly, MegabatOrtho6 sequences contain a canonical receptor binding motif in their SU region, but do not cluster closely with BaEV/SRV-Y viruses, which also contain that motif. In our analysis we did not use the SU region in the alignment, since it is very poorly conserved, thus unsuitable for phylogenetic analysis. We suspect that it is possible that the surprising positioning of those orthologs on the tree might be caused by recombination or deletion in those sequences, which made them cluster with candidates other than the expected one.

Yangochiroptera have haemochorial placentation, while different Yinpterachiroptera utilise either endoteliochorial or haemochorial placentation (Fig. 3) (Gopalakrishna and Karim, 1979; Teeling et al., 2005). Because of this, bats are a good model to show how different viral progenitors of *syncytin* are responsible for different placenta types in mammals. In contrast to most other mammals, Yangochiroptera females are larger than males both in weight, size, and wingspan (Lisón et al., 2014; Myers, 1978; O'Mara et al., 2016; Stevens et al., 2013). *Syncytins* also seem to be involved in muscle development as they have been found expressed in the muscles of rodents, primates, carnivores and ruminants (Redelsperger et al., 2016). Levels of expression of potential bat *syncytins* in muscles, as well as correlations of expression with sex, could shed additional light on the evolution of placental mammals.

The clustering of the VesperOrtho6 sequences with HEMO might mean that this is an old and unused ancestor of a sequence similar to HEMO, which could have shared a similar function. The HEMO gene was introduced into the genome of a mammalian ancestor around 100-120 Mya (before the split of Laurasiatherians and Euarchontoglires) and it can be detected in conserved full-length form in simians and cats (Heidmann et al., 2017). The presence of VesperOrtho6 in all analysed *Vespertilionidae* bats might suggest that this sequence is specific for that family, which might suggest that other families could have HEMO homologoues placed in different genomic regions compared to vesper bats. We ran BLAST analysis comparing sequences found in bats to other mammalian sequences and found similarities in 5′ end flanking region which further confirms that sequences in bats might be orthologs of HEMO, present in their genomes as pseudogenes. From our analysis, as expected, bat HEMO-like sequences are not under selection.

Long term co-evolution between bats and viruses might confer the ability to survive viral infections, and to safely harbour viruses and large numbers of highly active transposable elements in their genomes (Ray et al., 2007, 2008). It is reasonable to expect that this co-evolution at some point resulted in co-option of viral genes that may have played a role in antiviral immunity (Chuong et al., 2016). It is possible that bats might have a different set-point for triggering the immune response and harbour pathogens controlling their replication in tissues rather than clearing them and setting up immune-pathological responses in infected tissues (Xie et al., 2018). As the only mammals capable of active flight, bats have cells that are under high oxidative stress and limiting processes, which might be additionally damaging, and might be a viable explanation for their immunological differences in susceptibility to pathogens (Ahn et al., 2016; Zhang et al., 2013). We hypothesise that bat resilience may also stem from captured ERV genes, which help bats to overcome viral infections and suppress symptoms. There is a need for more analysis and laboratory tests to mark out which of these sequences might have immunosuppressive or antiviral potential, and if any are substantially expressed in tissues, particularly the placenta. To make such analyses more thorough, more bat genome assemblies from different families with good data quality will be needed.

Declaration of Competing Interest

None.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. AK was funded by the Royal Society.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.virusres.2019.197645.

References

Abascal, F., Zardoya, R., Posada, D., 2005. ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21 (9), 2104–2105.
Agnarsson, I., Zambrana-Torrelio, C.M., Flores-Saldana, N.P., May-Collado, L.J., 2011. A time-calibrated species-level phylogeny of bats (Chiroptera, Mammalia). PLoS Curr. 3, RRN1212.
Ahn, M., Cui, J., Irving, A.T., Wang, L.-F., 2016. Unique loss of the PYHIN gene family in bats amongst mammals: implications for inflammasome sensing. Sci. Rep. 6, 21722.
Alcami, A., Koszinowski, U.H., 2000. Viral mechanisms of immune evasion. Trends Microbiol. 8 (9), 410–418.
Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215 (3), 403–410.
Anisimova, M., Bielawski, J.P., Yang, Z., 2001. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. Mol. Biol. Evol. 18 (8), 1585–1592.
Anisimova, M., Gil, M., Dufayard, J.-F., Dessimoz, C., Gascuel, O., 2011. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. Syst. Biol. 60 (5), 685–699.
Arnaud, F., Caporale, M., Varela, M., Biek, R., Chessa, B., Alberti, A., et al., 2007a. A paradigm for virus–host coevolution: sequential counter-adaptations between endogenous and exogenous retroviruses. PLoS Pathog. 3 (11), e170.
Arnaud, F., Murcia, P.R., Palmarini, M., 2007b. Mechanisms of late restriction induced by an endogenous retrovirus. J. Virol. 81 (20), 11441–11451.
Aswad, A., Katzourakis, A., 2012. Paleovirology and virally derived immunity. Trends Ecol. Evol. (Amst.) 27 (11), 627–636.
Baer, G.M., Bales, G.L., 1967. Experimental rabies infection in the Mexican freetail bat. J. Infect. Dis. 117 (1), 82–90.
Bénit, L., Dessen, P., Heidmann, T., 2001. Identification, phylogeny, and evolution of retroviral elements based on their envelope genes. J. Virol. 75 (23), 11709–11719.
Blaise, S., de Parseval, N., Bénit, L., Heidmann, T., 2003. Genomewide screening for fusogenic human endogenous retrovirus envelopes identifies syncytin 2, a gene conserved on primate evolution. Proc. Natl. Acad. Sci. 100 (22), 13013–13018.
Blaise, S., Mangeney, M., Heidmann, T., 2001. The envelope of Mason–Pfizer monkey virus has immunosuppressive properties. J. Gen. Virol. 82 (7), 1597–1600.
Bustamante Rivera, Y.Y., Brütting, C., Schmidt, C., Volkmer, I., Staege, M.S., 2018. Endogenous retrovirus 3–history, physiology, and pathology. Front. Microbiol. 8, 2691.
Calisher, C.H., Childs, J.E., Field, H.E., Holmes, K.V., Schountz, T., 2006. Bats: important reservoir hosts of emerging viruses. Clin. Microbiol. Rev. 19 (3), 531–545.
Chinwalla, A.T., Cook, L.L., Delehaunty, K.D., Fewell, G.A., Fulton, L.A., Fulton, R.S., et al., 2002. Initial sequencing and comparative analysis of the mouse genome. Nature 420 (6915), 520–562.
Chu, D., Poon, L., Guan, Y., Peiris, J., 2008. Novel astroviruses in insectivorous bats. J. Virol. 82 (18), 9107–9114.
Chua, K.B., Koh, C.L., Hooi, P.S., Wee, K.F., Khong, J.H., Chua, B.H., et al., 2002. Isolation of Nipah virus from Malaysian Island flying-foxes. Microbes Infect. 4 (2), 145–151.
Chuong, E.B., Elde, N.C., Feschotte, C., 2016. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. Science 351 (6277), 1083–1087.
Cianciolo, G.J., Copeland, T.D., Oroszlan, S., Snyderman, R., 1985. Inhibition of lymphocyte proliferation by a synthetic peptide homologous to retroviral envelope proteins. Science 230 (4724), 453–455.
Cornelis, G., Funk, M., Vernochet, C., Leal, F., Tarazona, O.A., Meurice, G., et al., 2017. An endogenous retroviral envelope syncytin and its cognate receptor identified in the viviparous placental Mabuya lizard. Proc. Natl. Acad. Sci., 201714590.
Cornelis, G., Vernochet, C., Carradec, Q., Souquere, S., Mulot, B., Catzeflis, F., et al., 2015. Retroviral envelope gene captures and syncytin exaptation for placentation in marsupials. Proc. Natl. Acad. Sci. 112 (5), E487–E496.
Cornelis, G., Vernochet, C., Malicorne, S., Souquere, S., Tzika, A.C., Goodman, S.M., et al., 2014. Retroviral envelope syncytin capture in an ancestrally diverged mammalian clade for placentation in the primitive Afrotherian tenrecs. Proc. Natl. Acad. Sci. 111 (41), E4332–E4341.
Cui, J., Wang, L.-F., 2015. Genomic mining reveals deep evolutionary relationships between Bornaviruses and bats. Viruses 7 (11), 5792–5800.
Denner, J., 2016. Expression and function of endogenous retroviruses in the placenta. Apmis. 124 (1-2), 31–43.
Dunlap, K.A., Palmarini, M., Varela, M., Burghardt, R.C., Hayashi, K., Farmer, J.L., et al., 2006. Endogenous retroviruses regulate periimplantation placental growth and differentiation. Proc. Natl. Acad. Sci. 103 (39), 14390–14395.
Dupressoir, A., Marceau, G., Vernochet, C., Bénit, L., Kanellopoulos, C., Sapin, V., Heidmann, T., 2005. Syncytin-A and syncytin-B, two fusogenic placenta-specific murine envelope genes of retroviral origin conserved in Muridae. Proc. Natl. Acad.

Sci. 102 (3), 725–730.
Esnault, C., Cornelis, G., Heidmann, O., Heidmann, T., 2013. Differential evolutionary fate of an ancestral primate endogenous retrovirus envelope gene, the EnvV syncytin, captured for a function in placentation. PLoS genetics 9 (3) p.e1003400.
Feschotte, C., Gilbert, C., 2012. Endogenous viruses: insights into viral evolution and impact on host biology. Nat. Rev. Genet. 13 (4), 283–296.
Field, H., Young, P., Yob, J.M., Mills, J., Hall, L., Mackenzie, J., 2001. The natural history of Hendra and Nipah viruses. Microbes Infect. 3 (4), 307–314.
Finlay, B.B., McFadden, G., 2006. Anti-immunology: evasion of the host immune system by bacterial and viral pathogens. Cell 124 (4), 767–782.
Gilbert, C., Feschotte, C., 2010. Genomic fossils calibrate the long-term evolution of hepadnaviruses. PLoS Biol. 8 (9), e1000495.
Gopalakrishna, A., Karim, K.B., 1979. Fetal membranes and placentation in Chiroptera. J. Reprod. Fertil. 56 (1), 417–429.
Haig, D., 2012. Retroviruses and the placenta. Curr. Biol. 22 (15), R609–R613.
Hanna, J.N., Carney, I.K., Smith, G.A., Tannenberg, A.E., Deverill, J.E., Botha, J.A., et al., 2000. Australian bat lyssavirus infection: a second human case, with a long incubation period. Med. J. Aust. 172 (12), 597–599.
Heidmann, O., Vernochet, C., Dupressoir, A., Heidmann, T., 2009. Identification of an endogenous retroviral envelope gene with fusogenic activity and placenta-specific expression in the rabbit: a new" syncytin" in a third order of mammals. Retrovirology 6 (1), 107.
Heidmann, O., Béguin, A., Paternina, J., Berthier, R., Deloger, M., Bawa, O., et al., 2017. HEMO, an ancestral endogenous retroviral envelope protein shed in the blood of pregnant women and expressed in pluripotent stem cells and tumors. Proc. Natl. Acad. Sci. 114 (32), E6642–E6651.
Henzy, J.E., Gifford, R.J., Kenaley, C.P., Johnson, W.E., 2016. An intact retroviral gene conserved in spiny-rayed fishes for over 100 my. Mol. Biol. Evol msw262.
Henzy, J.E., Johnson, W.E., 2013. Pushing the endogenous envelope. Philos. Trans. R. Soc. Lond., B, Biol. Sci. 368 (1626), 20120506.
Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: bayesian inference of phylogenetic trees. Bioinformatics. 17 (8), 754–755.
Januszeski, M.M., Cannon, P.M., Chen, D., Rozenberg, Y., Anderson, W.F., 1997. Functional analysis of the cytoplasmic tail of Moloney murine leukemia virus envelope protein. J. Virol. 71 (5), 3613–3619.
Johnson, R.T., 2003. Emerging viral infections of the nervous system. J. Neurovirol. 9 (2), 140–147.
Katzourakis, A., Gifford, R.J., 2010. Endogenous viral elements in animal genomes. PLoS Genet. 6 (11), e1001191.
Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 28 (12), 1647–1649.
Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., et al., 2001. Initial sequencing and analysis of the human genome. Nature 409 (6822), 860–921.
Larsson, A., 2014. AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics 30 (22), 3276–3278.
Lavialle, C., Cornelis, G., Dupressoir, A., Esnault, C., Heidmann, O., Vernochet, C., et al., 2013. Paleovirology of 'syncytins', retroviral env genes exapted for a role in placentation. Philos. Trans. Biol. Sci. 368 (1626), 20120507.
Leiser, R., Koob, B., 1993. Development and characteristics of placentation in a carnivore, the domestic cat. J. Exp. Zool. 266 (6), 642–656.
Leroy, E.M., Kumulungui, B., Pourrut, X., Rouquet, P., Hassanin, A., Yaba, P., et al., 2005. Fruit bats as reservoirs of Ebola virus. Nature 438 (7068), 575–576.
Lisón, F., Haz, A., González-Revelles, C., Calvo, J.F., 2014. Sexual size dimorphism in greater mouse-eared bat Myotis myotis (Chiroptera: vespertilionidae) from a Mediterranean region. Acta Zool. 95 (2), 137–143.
Mackenzie, J.S., Field, H.E., Guyatt, K.J., 2003. Managing emerging diseases borne by fruit bats (flying foxes), with particular reference to henipaviruses and Australian bat lyssavirus. J. Appl. Microbiol. 94 (Suppl), 59S–69S.
Malik, H.S., 2012. Retroviruses push the envelope for mammalian placentation. Proc. Natl. Acad. Sci. 109 (7), 2184–2185.
Mangeney, M., de Parseval, N., Thomas, G., Heidmann, T., 2001. The full-length envelope of an HERV-H human endogenous retrovirus has immunosuppressive properties. J. Gen. Virol. 82 (10), 2515–2518.
Mangeney, M., Heidmann, T., 1998. Tumor cells expressing a retroviral envelope escape immune rejection in vivo. Proc Natl Acad Sci U S A. 95 (25), 14920–14925.
Mangeney, M., Renard, M., Schlecht-Louf, G., Bouallaga, I., Heidmann, O., Letzelter, C., et al., 2007. Placental syncytins: genetic disjunction between the fusogenic and immunosuppressive activity of retroviral envelope proteins. Proc. Natl. Acad. Sci. 104 (51), 20534–20539.
Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y., et al., 2014. CDD: NCBI's conserved domain database. Nucleic Acids Res. 43 (D1), D222–D226.
Middleton, D., Morrissy, C., Van Der Heide, B., Russell, G., Braun, M., Westbury, H., et al., 2007. Experimental Nipah virus infection in pteropid bats (Pteropus poliocephalus). J. Comp. Pathol. 136 (4), 266–272.
Morozov, V.A., Morozov, A.V., Semaan, M., Denner, J., 2012. Single mutations in the transmembrane envelope protein abrogate the immunosuppressive property of HIV-1. Retrovirology 9, 67.
Myers, P., 1978. Sexual dimorphism in size of vespertilionid bats. Am. Nat. 112 (986), 701–711.
O'Mara, M.T., Bauer, K., Blank, D., Baldwin, J.W., Dechmann, D.K., 2016. Common noctule bats are sexually dimorphic in migratory behaviour and body size but not wing shape. PLoS One 11 (11), e0167027.
Omatsu, T., Watanabe, S., Akashi, H., Yoshikawa, Y., 2007. Biological characters of bats in relation to natural reservoir of emerging viruses. Comparative immunology. microbiology and infectious diseases 30 (5), 357–374.
Patel, M.R., Emerman, M., Malik, H.S., 2011. Paleovirology—ghosts and gifts of viruses past. Curr. Opin. Virol. 1 (4), 304–309.
Ploegh, H.L., 1998. Viral strategies of immune evasion. Science 280 (5361), 248–253.
Plowright, R.K., Eby, P., Hudson, P.J., Smith, I.L., Westcott, D., Bryden, W.L., et al., 2015. Ecological dynamics of emerging bat virus spillover. Proc. Biol. Sci. 282 (1798), 20142124.
Pötgens, A., Schmitz, U., Bose, P., Versmold, A., Kaufmann, P., Frank, H.-G., 2002. Mechanisms of syncytial fusion: a review. Placenta 23, S107–S113.
Quinlan, A.R., 2014. BEDTools: the Swiss-army tool for genome feature analysis. Curr. Protoc. Bioinformatics 11 (2), 34 1-.2.
Ray, D.A., Feschotte, C., Pagan, H.J., Smith, J.D., Pritham, E.J., Arensburger, P., et al., 2008. Multiple waves of recent DNA transposon activity in the bat. Myotis lucifugus. Genome research.
Ray, D.A., Pagan, H.J., Thompson, M.L., Stevens, R.D., 2007. Bats with hATs: evidence for recent DNA transposon activity in genus Myotis. Mol. Biol. Evol. 24 (3), 632–639.
Redelsperger, F., Cornelis, G., Vernochet, C., Tennant, B.C., Catzeflis, F., Mulot, B., et al., 2014a. Capture of syncytin-Mar1, a fusogenic endogenous retroviral envelope gene involved in placentation in the Rodentia squirrel-related clade. J. Virol. 88 (14), 7915–7928.
Redelsperger, F., Cornelis, G., Vernochet, C., Tennant, B.C., Catzeflis, F., Mulot, B., et al., 2014b. Capture of syncytin-Mar1, a fusogenic endogenous retroviral envelope gene involved in placentation in the Rodentia squirrel-related clade. J. Virol. 88 (14), 7915–7928.
Redelsperger, F., Raddi, N., Bacquin, A., Vernochet, C., Mariot, V., Gache, V., et al., 2016. Genetic evidence that captured retroviral envelope syncytins contribute to myoblast fusion and muscle sexual dimorphism in mice. PLoS Genet. 12 (9), e1006289.
Schlecht-Louf, G., Mangeney, M., El-Garch, H., Lacombe, V., Poulet, H., Heidmann, T., 2014. A targeted mutation within the feline leukemia virus (FeLV) envelope protein immunosuppressive domain to improve a canarypox virus-vectored FeLV vaccine. J. Virol. 88 (2), 992–1001.
Schlecht-Louf, G., Renard, M., Mangeney, M., Letzelter, C., Richaud, A., Ducos, B., et al., 2010. Retroviral infection in vivo requires an immune escape virulence factor encrypted in the envelope protein of oncoretroviruses. Proc Natl Acad Sci U S A. 107 (8), 3782–3787.
Shapiro, B., Rambaut, A., Drummond, A.J., 2005. Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. Mol. Biol. Evol. 23 (1), 7–9.
Sinha, A., Johnson, W.E., 2017. Retroviruses of the RDR superinfection interference group: ancient origins and broad host distribution of a promiscuous Env gene. Curr. Opin. Virol. 25, 105–112.
Smith, I., Wang, L.-F., 2013. Bats and their virome: an important source of emerging viruses capable of infecting humans. Curr. Opin. Virol. 3 (1), 84–91.
Sonigo, P., Barker, C., Hunter, E., Wain-Hobson, S., 1986. Nucleotide sequence of Mason-Pfizer monkey virus: an immunosuppressive D-type retrovirus. Cell 45 (3), 375–385.
Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30 (9), 1312–1313.
Stevens, R.D., Johnson, M.E., McCulloch, E.S., 2013. Absolute and relative secondary-sexual dimorphism in wing morphology: a multivariate test of the 'Big Mother'hypothesis. Acta Chiropt. 15 (1), 163–170.
Sulkin, S.E., Allen, R., 1974. Virus infections in bats. Monogr. Virol. 8, 1–103.
Taylor, D.J., Leach, R.W., Bruenn, J., 2010. Filoviruses are ancient and integrated into mammalian genomes. BMC Evol. Biol. 10 (1), 193.
Teeling, E.C., Springer, M.S., Madsen, O., Bates, P., O'brien, S.J., Murphy, W.J., 2005. A molecular phylogeny for bats illuminates biogeography and the fossil record. Science 307 (5709), 580–584.
Wang, L.-F., Walker, P.J., Poon, L.L., 2011. Mass extinctions, biodiversity and mitochondrial function: are bats 'special'as reservoirs for emerging viruses? Curr. Opin. Virol. 1 (6), 649–657.
Xie, J., Li, Y., Shen, X., Goh, G., Zhu, Y., Cui, J., et al., 2018. Dampened STING-Dependent interferon activation in bats. Cell Host Microbe.
Yang, Z., 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Bioinformatics. 13 (5), 555–556.
Zhang, G., Cowled, C., Shi, Z., Huang, Z., Bishop-Lilly, K.A., Fang, X., et al., 2013. Comparative analysis of bat genomes provides insight into the evolution of flight and immunity. Science. 339 (6118), 456–460.
Zhuo, X., Feschotte, C., 2015. Cross-species transmission and differential fate of an endogenous retrovirus in three mammal lineages. PLoS Pathog. 11 (11), e1005279.
Zhuo, X., Rho, M., Feschotte, C., 2013. Genome-wide characterization of endogenous retroviruses in the bat Myotis lucifugus reveals recent and diverse infections. J. Virol. 87 (15), 8493–8501.