

RESEARCH

Open Access



# Identification of a homogenous structural basis for oligomerization by retroviral Rev-like proteins

Chijioke N. Umunnakwe<sup>1,2,5</sup>, Karin S. Dorman<sup>2,3,4</sup>, Drena Dobbs<sup>2,3</sup> and Susan Carpenter<sup>1\*</sup> 

## Abstract

**Background:** Rev-like proteins are post-transcriptional regulatory proteins found in several retrovirus genera, including lentiviruses, betaretroviruses, and deltaretroviruses. These essential proteins mediate the nuclear export of incompletely spliced viral RNA, and act by tethering viral pre-mRNA to the host CRM1 nuclear export machinery. Although all Rev-like proteins are functionally homologous, they share less than 30% sequence identity. In the present study, we computationally assessed the extent of structural homology among retroviral Rev-like proteins within a phylogenetic framework.

**Results:** We undertook a comprehensive analysis of overall protein domain architecture and predicted secondary structural features for representative members of the Rev-like family of proteins. Similar patterns of  $\alpha$ -helical domains were identified for Rev-like proteins within each genus, with the exception of deltaretroviruses, which were devoid of  $\alpha$ -helices. Coiled-coil oligomerization motifs were also identified for most Rev-like proteins, with the notable exceptions of HIV-1, the deltaretroviruses, and some small ruminant lentiviruses. In Rev proteins of primate lentiviruses, the presence of predicted coiled-coil motifs segregated within specific primate lineages: HIV-1 descended from SIVs that lacked predicted coiled-coils in Rev whereas HIV-2 descended from SIVs that contained predicted coiled-coils in Rev. Phylogenetic ancestral reconstruction of coiled-coils for all Rev-like proteins predicted a single origin for the coiled-coil motif, followed by three losses of the predicted signal. The absence of a coiled-coil signal in HIV-1 was associated with replacement of canonical polar residues with non-canonical hydrophobic residues. However, hydrophobic residues were retained in the key 'a' and 'd' positions, and the  $\alpha$ -helical region of HIV-1 Rev oligomerization domain could be modeled as a helical wheel with two predicted interaction interfaces. Moreover, the predicted interfaces mapped to the dimerization and oligomerization interfaces in HIV-1 Rev crystal structures. Helical wheel projections of other retroviral Rev-like proteins, including endogenous sequences, revealed similar interaction interfaces that could mediate oligomerization.

**Conclusions:** Sequence-based computational analyses of Rev-like proteins, together with helical wheel projections of oligomerization domains, reveal a conserved homogeneous structural basis for oligomerization by retroviral Rev-like proteins.

**Keywords:** Retroviruses, Rev, Rev-like proteins, Coiled-coil, Ancestral inference, Helical wheel

\*Correspondence: scarp@iastate.edu

<sup>1</sup> Department of Animal Science, Iowa State University, Ames, IA 50011, USA

Full list of author information is available at the end of the article

## Background

Retroviruses such as HIV-1 require nuclear export of incompletely spliced and unspliced viral mRNA to generate viral proteins and genomic RNA for replication. Lentiviruses, deltaretroviruses and a few betaretroviruses encode a regulatory protein that mediates this essential step. Lentiviruses encode the regulatory protein Rev [1–9]; deltaretroviruses encode a Rev analog, Rex [10, 11]; and three betaretroviruses, mouse mammary tumor virus (MMTV), Jaagsitke sheep retrovirus (JSRV), and the human endogenous retrovirus type K (HERV-K HML2), encode the regulatory proteins Rem [12, 13], Rej [14], and Rec, respectively [15]. This family of proteins, collectively termed Rev-like proteins, is believed to share a similar mechanism of function despite being highly divergent at the sequence level.

HIV-1 Rev is the best characterized of the retroviral Rev-like proteins. Following translation, HIV-1 Rev traffics to the nucleus, binds a ~350 nucleotide target in the viral pre-mRNA, termed the Rev-responsive element (RRE), oligomerizes on the RRE, and associates with host export factors to form an export-competent complex that exits the nucleus [16, 17]. Specific motifs or domains in HIV-1 Rev and other Rev-like proteins mediate distinct steps of the Rev pathway, including nuclear import, RNA binding, oligomerization and nuclear export [18–21].

Early studies established that an arginine-rich motif (ARM) within the  $\alpha$ -helical region of HIV-1 Rev recognizes a high affinity binding site within the HIV-1 RRE [22–26]. This initial binding of HIV-1 Rev to viral RNA is followed by assembly of additional Rev molecules along the RRE and recruitment of the host cellular factor CRM1 for nucleocytoplasmic export of the ribonucleoprotein complex [27–29]. More recent structural studies revealed a re-ordering of HIV-1 Rev dimers upon RRE binding, the result of plasticity in Rev oligomerization domains [30]. This plasticity could allow Rev to adopt a wide array of dimeric and multimeric conformations that facilitate interactions with a genetically variable and conformationally pliable RRE [31, 32].

Equine infectious anemia virus (EIAV) Rev is functionally homologous to primate lentivirus Revs, but is also unusual in several aspects. EIAV Rev is genetically variable, differs in the organization of functional domains [33], and contains a novel bipartite RNA binding domain (RBD) comprising two ARMs (ARM-1 and ARM-2) separated by 79 amino acids in primary sequence [34]. More recent findings also suggest that RNA binding requires coiled-coil interactions that mediate dimerization of EIAV Rev monomers, thereby bringing ARM-1 and ARM-2 into close proximity for RNA binding [35]. This is reminiscent of structural studies of HIV-1 Rev/RRE interactions, which suggest that dimerization of HIV-1

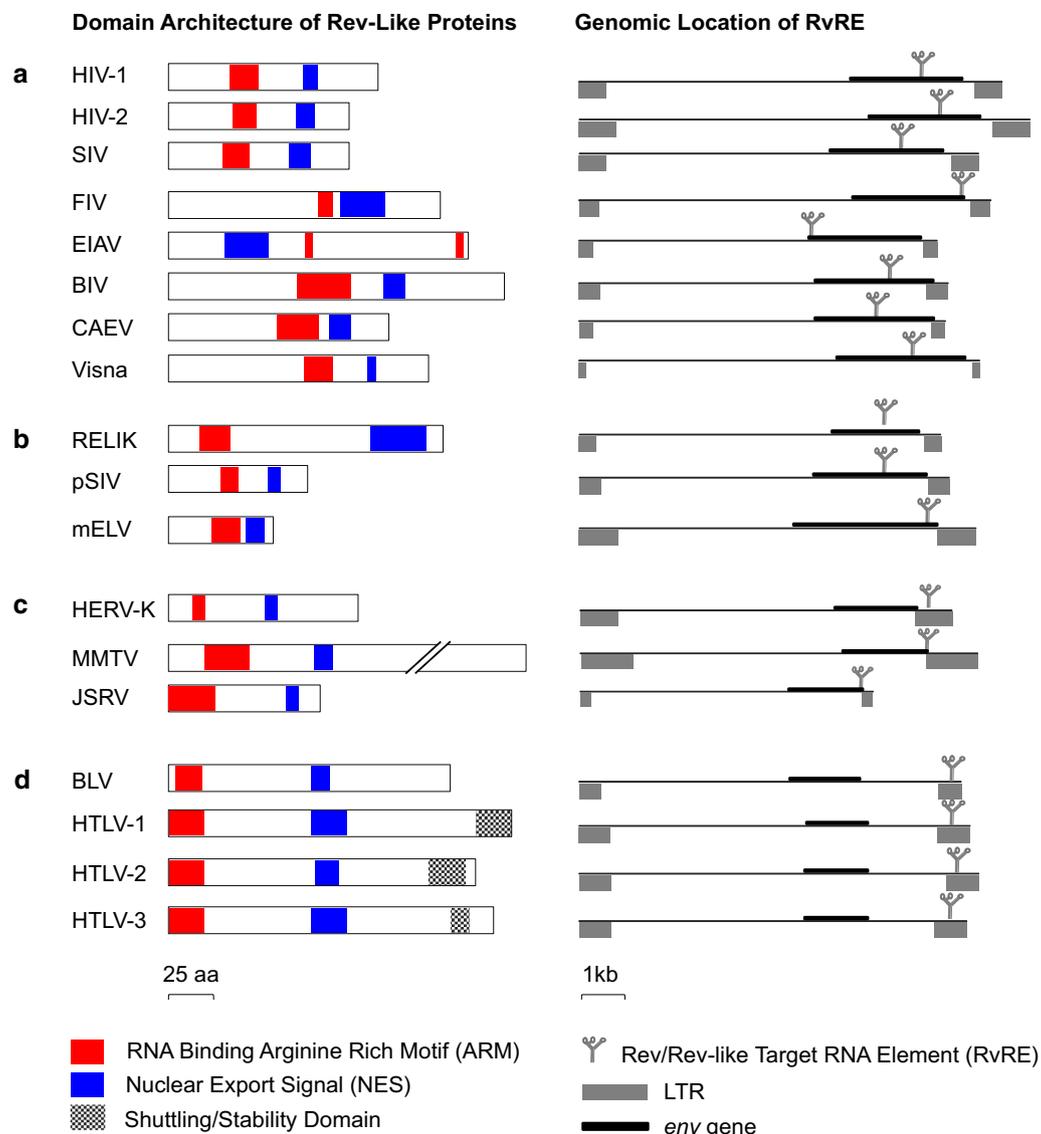
Rev juxtaposes the ARMs of each monomer at high affinity binding sites on the RRE [29, 36, 37]. Thus, while the bipartite RBD of EIAV Rev is novel among lentivirus Revs, it appears that dimerization and juxtaposition of Rev ARMs on the RRE could be a shared feature of lentiviral Rev-RRE interactions.

The possible parallels in the manner by which EIAV and HIV Rev associate with their cognate RNA elements raises the possibility that a coiled-coil motif might also exist within the oligomerization domain of HIV-1 Rev, and perhaps other retroviral Rev-like proteins. In the present study, we investigated the extent of structural homology among retroviral Rev-like proteins by undertaking a comprehensive computational analysis of overall protein domain architecture and specific secondary structural elements for representative members of the Rev-like family of proteins. Results demonstrated that a coiled-coil motif is likely an ancestral trait in retroviral Rev-like proteins and is still functional in HIV-1 Rev, where it is involved in dimerization and oligomerization. Coiled-coil or coiled-coil derived motifs were found associated with oligomerization domains of other retroviral Rev-like proteins, indicating that a common structural motif likely underlies essential protein–protein interactions in the Rev-like family of proteins.

## Results

### Rev-like proteins share a similar functional organization

To determine whether the overall architecture of Rev-like proteins follows a conserved structural pattern, we compared the organization of protein domains of representative Rev-like members of lentiviruses, deltaretroviruses and betaretroviruses. Although Rev-like proteins vary in length, a similar organization of protein domains in which the NES is located C-terminal to the ARM was observed in all members except EIAV Rev (Fig. 1, left). In addition, only the Rev-like proteins of human deltaretroviruses contained a stability/shuttling domain, proximal to the C-terminus (Fig. 1d) [38]. The RNA target elements of all Rev-like proteins, hereafter termed RvRE, are always found in the *env*/LTR region of the genome (Fig. 1, right). Lentiviral RvRE are typically located in the central region of *env*; however, the RvRE of EIAV is located at the 5' end of *env*, while the RvRE of FIV and the endogenous ferret lentivirus, mELV, are located at the 3' end of *env* (Fig. 1a, b). In beta- and deltaretroviruses, the RvRE is located near the start of or within the U3 region of LTR (Fig. 1c, d). Despite some variations, the overall organization of protein domains in retroviral Rev-like proteins and the location of target RNA elements within their cognate viral genomes revealed a similar structural organization. Interestingly, EIAV Rev is an outlier among Rev-like proteins with respect to the overall organization



**Fig. 1** Domain organization of Rev-like proteins (*left*) and location of cognate RNA target elements (RvRE) in the proviral genome (*right*) in: **a** Exogenous lentiviruses; **b** endogenous lentiviruses; **c** betaretroviruses; **d** deltaretroviruses. *Red colored boxes* in protein schematics represent RNA binding arginine rich motifs (ARMs), *blue boxes* represent nuclear export signals (NES), and *checkered boxes* represent shuttling/stability domains unique to HTLV members. Stem loop structures in viral genomes (*right*) represent location of the RvRE target. LTRs are shown as *grey terminal boxes* and the *env* gene is shown as a *thick black rectangle* within the genome

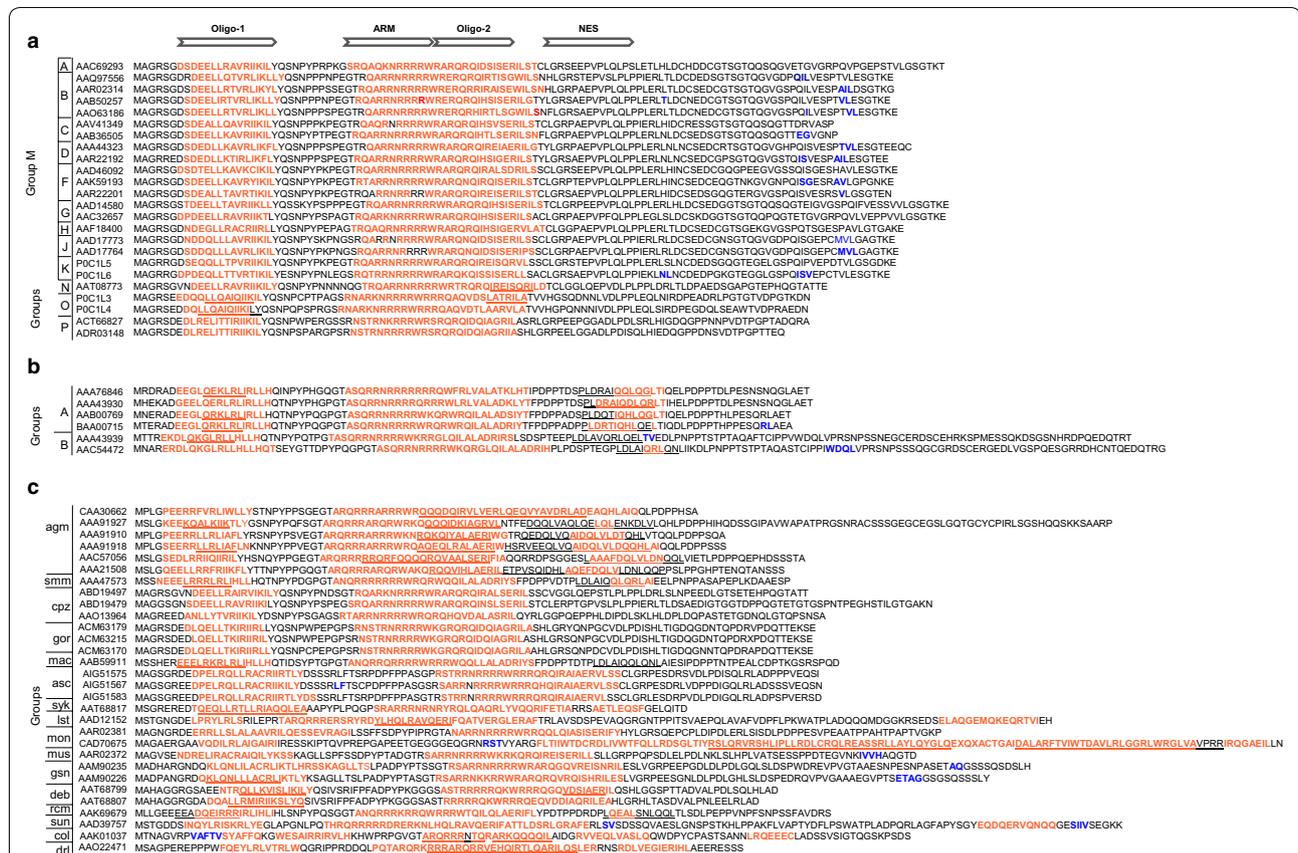
of functional domains and location of the RvRE. It is not clear if/how this may relate to some of the unique aspects of EIAV pathogenesis [33].

#### **$\alpha$ -Helical secondary structures are a common feature of Rev-like proteins**

In HIV-1 Rev, a helix-turn-helix structural motif mediates intra- and intermolecular protein-protein interactions required for high affinity RNA binding and oligomerization [29, 37]. To determine whether other

Rev-like proteins contain similar local structural topologies, we analyzed specific predicted secondary structural elements in representative retroviral Rev-like proteins.

Secondary structure predictions indicated that the two  $\alpha$ -helices located in the N-terminal half of HIV-1 Rev (red residues, Fig. 2a) are conserved in Rev proteins of the other primate lentiviruses (Fig. 2b, c). The first helix encompasses the N-terminal oligomerization domain (Oligo-1) while the second helix overlaps the arginine-rich motif (ARM) and the downstream oligomerization



**Fig. 2** Distribution of  $\alpha$ -helices and coiled-coil signals in representative Rev proteins of primate lentiviruses, **a** HIV-1, **b** HIV-2, and **c** SIV. Rev sequences, including accession numbers are ordered by groups, and HIV-1 group M is further ordered by clade, A–K. *White arrows* above all sequences indicate locations of the following Rev functional domains: N-terminal oligomerization domain (Oligo-1); C-terminal oligomerization domain (Oligo-2); RNA binding arginine rich motif (ARM); and nuclear export signal (NES). Residues in *red* indicate regions of predicted  $\alpha$ -helices; residues in *blue* indicate predicted  $\beta$ -strand regions, and underlined residues represent regions of predicted coiled-coils. Sequences in the figure are a subset of the complete dataset in Additional file 2, which includes genome, Pol, and Env accession numbers for each sequence

domain (Oligo-2). Thus, the helix-turn-helix structure present in the N-terminal half of the HIV-1 Rev crystal structure [29] is a conserved structural feature of all primate lentivirus Rev proteins. Extended unstructured regions were present in the C-terminal half of HIV-1 and HIV-2 Revs and in most of the SIV Revs, although an additional short  $\alpha$ -helical region, overlapping the NES, was predicted for some HIV-2 and SIV Revs (Fig. 2b, c).

At least two  $\alpha$ -helical domains were also predicted for Rev-like proteins of non-primate lentiviruses and betaretroviruses (Additional file 1). Overall, however, these Rev-like proteins contained relatively more  $\alpha$ -helical regions, and less unstructured regions, as compared to primate lentivirus Revs. Surprisingly, the deltaretrovirus Rev-like proteins contained little or no detectable ordered secondary structure (Additional file 1). The presence of intrinsically disordered regions in some retroviral Rev-like proteins may contribute to the structural flexibility recently described for HIV-1 Rev [17].

**Predicted coiled-coil motifs are widely distributed among Rev-like proteins**

Coiled-coils are a subset of  $\alpha$ -helices that mediate protein–protein interactions. In EIAV Rev, computational and biochemical studies indicated that a predicted coiled-coil in the central region of the protein is required for dimerization and RNA binding [35]. The CCHMM\_PROF coiled-coil server [39], was used to assess whether coiled-coils were predicted for other retroviral Rev-like proteins. Predicted coiled-coil motifs were identified in the Rev proteins of all primate lentivirus species analyzed; however, their distribution varied within groups (Fig. 2, Additional file 1). Coiled-coils were predicted in all Rev sequences analyzed from HIV-2 Groups A and B, and from SIV African green monkey (agm), sooty mangabey (smm), and macaque (mac). In contrast, coiled-coils

were not identified in any HIV-1 group M Rev sequences, in SIV chimpanzee (cpz), or red tailed monkey (asc). When present in the primate lentivirus Revs, the coiled-coils were usually localized within or overlapping oligomerization domains (Fig. 2), suggesting they may play a role in protein–protein interactions.

Coiled-coil motifs are also widely distributed in the Rev sequences of non-primate lentiviruses, with the exception of the small ruminant lentivirus (SRLV) group (Additional file 1). Two out of three betaretroviruses Rev-like proteins, MMTV Rem and HERV-K Rec, were predicted to contain coiled-coils (Additional file 1). As expected, predicted coiled-coils were absent in the deltaretrovirus Rex proteins, which were predicted to lack  $\alpha$ -helices (Additional file 1). Taken together, these results indicate that the coiled-coil motif is a common structural feature within specific  $\alpha$ -helical regions of Rev-like proteins.

#### **Predicted coiled-coils segregate along phylogenetic lineages in primate lentiviruses**

The differential distribution of coiled-coils in the Rev proteins of primate lentiviruses may reflect the phylogenetic lineages of HIV-1 and HIV-2. HIV-1 is derived from SIVcpz, which arose from recombination events resulting in a mosaic genome wherein *pol* is derived from red-capped mangabey (SIVrcm) and *env/rev* is derived from SIVgsn/mon/mus [40]. In contrast, HIV-2 is derived from SIVsmm. This evolutionary history suggested that the absence of coiled-coils in HIV-1 Revs reflects the absence of coiled-coils in their SIV ancestors. To test this hypothesis, the presence or absence of coiled-coils was mapped onto an Env-based phylogeny of primate lentiviruses. As shown in Fig. 3, HIV-1 and SIVcpz formed a monophyletic clade with the SIVgsn/mon/mus lineage that was largely bereft of coiled-coils (indicated by red circles). In contrast, HIV-2 and SIVsmm formed a monophyletic clade with the SIVagm/rcm lineage in which the coiled-coil signal was present in every member. From these results, we conclude that the presence/absence of coiled-coils in the Rev proteins of primate lentiviruses is a lineage-specific trait.

#### **Coiled-coil motifs are an ancestral feature of retrovirus Rev-like proteins**

The identification of coiled-coils in Rev-like proteins of divergent retroviruses combined with their phylogenetic distribution in primate lentiviruses, indicated that coiled-coils may be an ancestral feature of retroviral Rev-like proteins. To explore this idea, the ancestral state of the coiled-coils was modeled on the Env-based phylogeny of primate lentiviruses depicted in Fig. 3. Ancestral inference by maximum likelihood yielded a 0.98 probability that the ancestral Revs of all primate lentiviruses

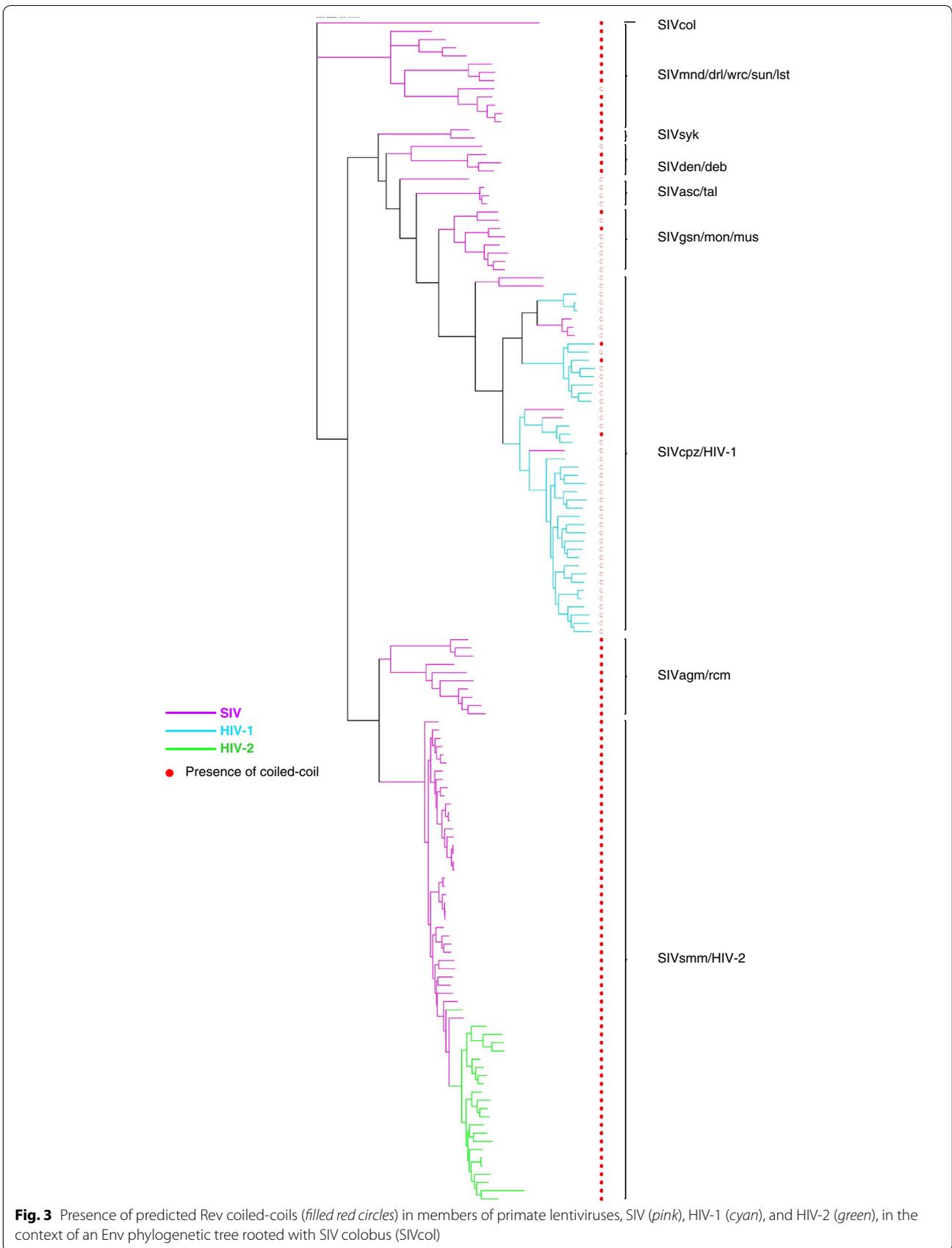
contained a coiled-coil motif (Fig. 4). The analysis also indicated one major loss of the coiled-coil signal in the primate lentiviruses (Fig. 4, red arrow), resulting in the absence of coiled-coils in SIVcpz/HIV-1 group.

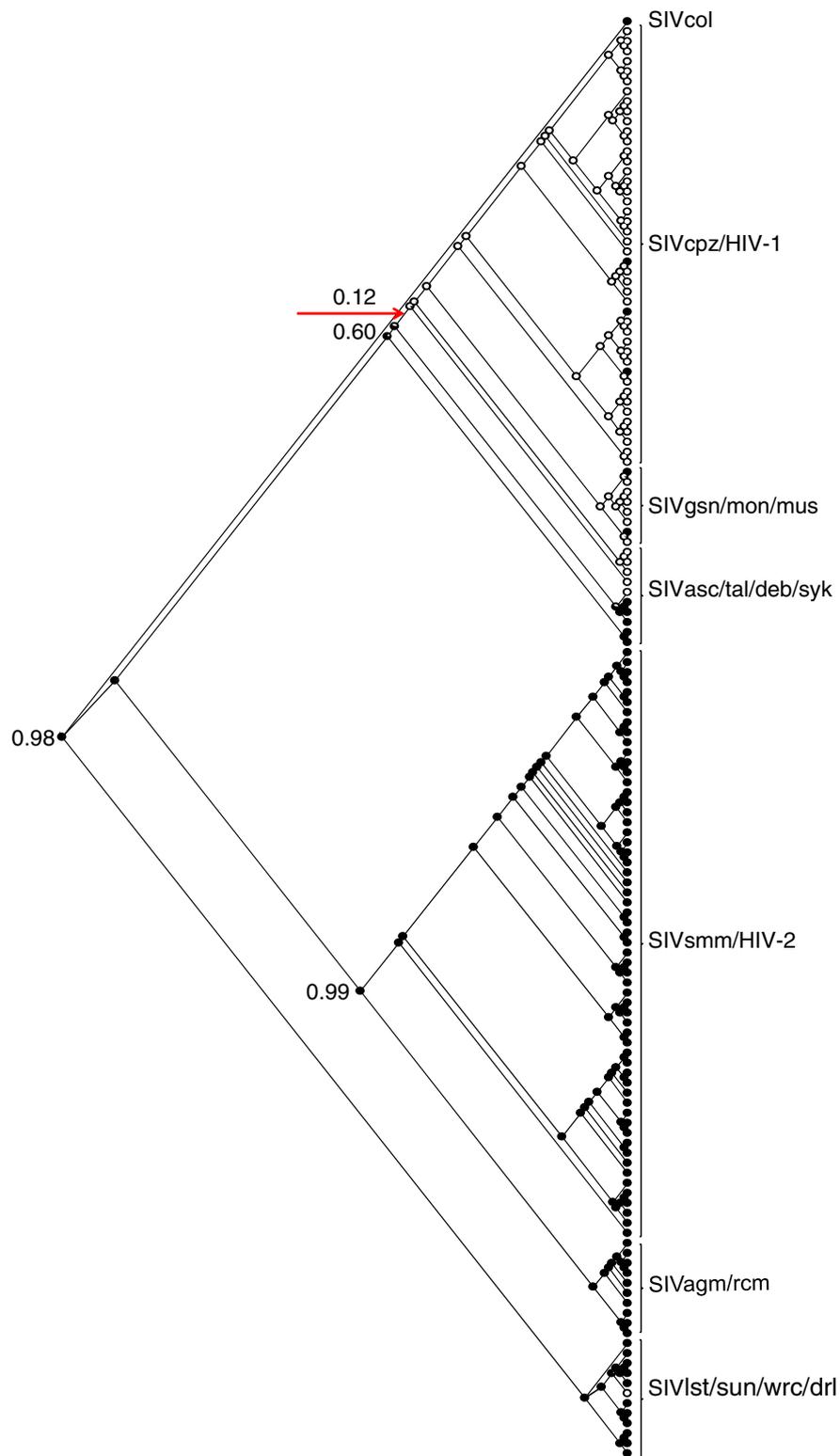
Having obtained strong evidence that the coiled-coil phenotype is an ancestral feature in primate lentivirus Revs, we expanded our analyses to include all retroviral members that contain Rev-like proteins. The ancestral state of coiled-coils was modeled on a Pol-based phylogeny of all retroviruses encoding Rev-like proteins (Fig. 5). Ancestral inference by maximum likelihood yielded a 0.72 probability that the coiled-coil signal was present in the ancestral sequence to all retroviral members that encode Rev-like proteins (Fig. 5, black arrow). The Pol-based tree depicts two groups, one comprising all primate lentiviruses and the other comprising all other Rev-like encoding retroviruses. In addition to the loss of the coiled-coil signal in the SIVcpz/HIV-1 lineage of the primate lentivirus group (Fig. 5, top red arrow), there were two additional losses in the group comprising all other Rev-like encoding retroviruses: one occurring on the branch leading to the small ruminant lentiviruses, and the other occurring on the branch leading to the deltaretroviruses (Fig. 5, bottom two red arrows). These losses are reflected in the absence of coiled-coil signals in the small ruminant lentivirus and deltaretrovirus groups (see Additional files 1, 2).

To ensure that the absence of coiled-coils in the SIVcpz/HIV-1, SRLVs, and deltaretrovirus groups was not due to a bias in the prediction algorithm used, we reanalyzed all sequences for coiled-coil motifs using another coiled-coil server, COILS, which utilizes a different prediction method [41]. The overall correlation between predictions from the two different servers was 0.68, and the two servers agreed on the absence of coiled-coils in the SIVcpz/HIV-1, SRLVs, and deltaretrovirus groups (Additional file 2). Thus, these analyses provided additional support that coiled-coils are an ancestral structural motif of retroviral Rev-like proteins.

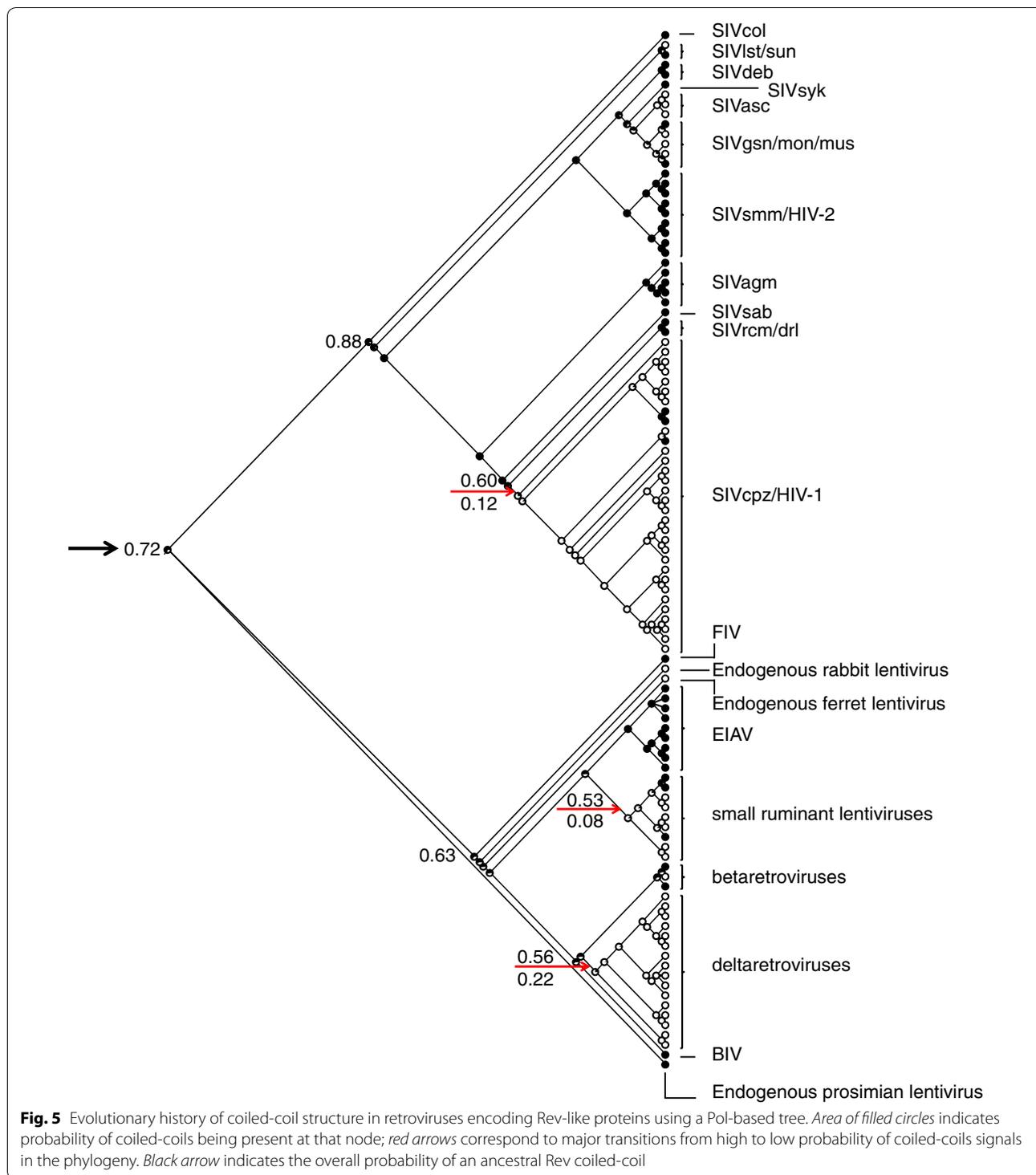
#### **The loss of a coiled-coil signal in HIV-1 Rev is associated with genetic variation in the canonical coiled-coil sequence motif**

Prediction of coiled-coil motifs in a given protein is based on sequence features present in well characterized canonical coiled-coils [42]. Retroviral sequences are highly variable, and the absence of predicted coiled-coil signals in certain lineages such as SIVcpz/HIV-1 could be due to genetic variation that weakens the designation criteria. To explore this possibility, we analyzed genetic variation within the N-terminal oligomerization domains overlapping the predicted coiled-coils of primate lentiviruses and compared the sequence logos of groups



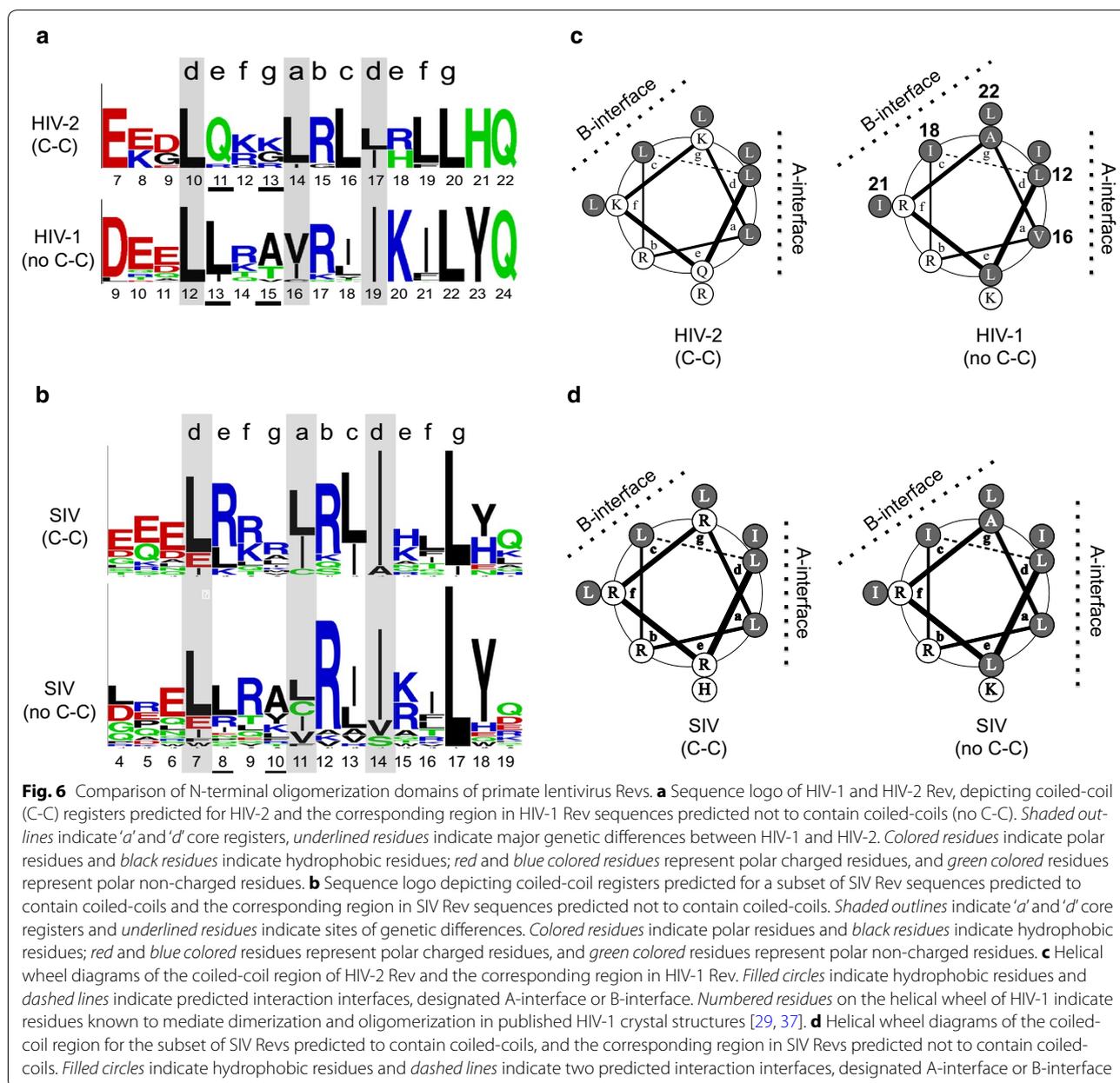


**Fig. 4** Ancestral inference of coiled-coils in Rev proteins of primate lentiviruses. Analysis was performed using the Env-based tree shown in Fig. 3. *Black area of filled circles* indicates the probability of Rev coiled-coils being present at that node; *red arrow points* to a single major transition from higher to lower probability of the coiled-coil signal in the phylogeny (0.60–0.12 probability). The probability of a Rev coiled-coil in the ancestral node of SIVsmm/HIV-2 and SIVagm/rcm (0.99), and the overall probability of an ancestral Rev coiled-coil for all lentiviruses (0.98) are indicated



that have a strong coiled-coil signature with groups that do not (Fig. 6). The HIV-2 Rev oligomerization domain displays hydrophobic residues in core 'a' and 'd' registers of an 'abcdefg' heptad motif, which is characteristic

of canonical coiled-coils (Fig. 6a). Charged and polar residues predominate at the 'e' and 'g' positions, which again is consistent with the canonical coiled-coil formulation [43]. Interestingly, HIV-1 Rev also contained



hydrophobic residues in the core 'a' and 'd' registers, conforming to the basic signature of a coiled-coil (Fig. 6a, shaded). The major sequence differences between HIV-1 and HIV-2 Rev are in the first 'e' and 'g' positions of the 'abcdefg' heptad motif (Fig. 6a, underlined), where HIV-1 Rev contains hydrophobic residues rather than the polar/charged residues found in HIV-2 and other canonical coiled-coils. The presence of hydrophobic residues at these sites may explain why coiled-coils are not predicted for most HIV-1 Rev sequences, as this runs counter to the canonical coiled-coil formulation.

Comparison of SIV Rev sequences that differ with respect to presence/absence of coiled coils showed a mixed population of hydrophobic and charged residues at the first 'e' and 'g' positions (Fig. 6b, underlined). Charged residues predominate in SIV Revs predicted to contain a coiled-coil, whereas hydrophobic residues predominate in SIV Revs predicted to lack a coiled-coil. It is likely, therefore, that genetic variation and selection for additional hydrophobic residues in the N-terminal oligomerization domain weakened the coiled-coil signal in Rev proteins of some primate lentiviruses, including HIV-1 Rev.



with an interaction interface similar to the betaretrovirus members predicted to contain coiled-coils. The Rev proteins of endogenous lentiviruses (Fig. 7c) also contained interaction interfaces that follow a pattern similar to that seen in FIV, BIV, EIAV (Fig. 7a) and the betaretroviruses (Fig. 7b). Each of the SRLV Revs with a predicted coiled-coil motif had an interaction interface containing a mix of hydrophobic and polar residues (Fig. 7d). In contrast, most SRLV Revs predicted to lack a coiled-coil have proline residues, which are helix breakers, in the analogous region (Fig. 7e). Similar to what we observed for Revs of primate lentivirus lineages, weakened or eliminated coiled-coil signals as a result of genetic variation correlated with loss of the coiled-coil signal in some Revs of the SRLV lineage.

Apart from HIV-1 Rev, few structural studies have been performed on Rev-like proteins. Therefore, there is little biophysical data documenting intermolecular interactions mediated by the interfaces predicted for most of the Rev-like proteins shown in Fig. 7. Molecular genetic studies, however, have provided some evidence that the predicted interfaces play a role in oligomerization of the respective Rev-like proteins. We previously reported that residues in the predicted EIAV Rev interface are critical for dimerization and RNA binding [35]. Both HERV-K Rec and SRLV Rev are known to oligomerize, and deletion of residues overlapping the predicted interaction interface in CAEV was found to abrogate dimerization [44]. When considered with available experimental data, our findings indicate that, with the exception of deltaretroviruses, oligomerization in retroviral Rev-like proteins is likely mediated by coiled-coil or coiled-coil-like structural motifs.

## Discussion

Dimerization and oligomerization of HIV-1 Rev along the RRE is an essential prerequisite for RNA export and viral replication [45]. In HIV-1 Rev, distinct dimerization and oligomerization domains mediate Rev assembly on the RNA target [17, 27]. Our previous studies of EIAV Rev identified a coiled-coil motif in the predicted oligomerization domain, which was suggested to mediate dimerization [35]. Here, we identified and characterized coiled-coils in Rev-like proteins from diverse retroviral genera. Coiled-coils were found to be an ancestral trait that largely persists in extant retroviruses, but whose signal is attenuated along certain phylogenetic lineages. For example, within primate Rev proteins of the lentivirus genera, the coiled-coil signal was dependent on HIV/SIV ancestry: Rev proteins descending from the HIV-2/SIVsmm lineage contain strong coiled-coil signals, whereas those from the HIV-1/SIVcpz lineage do not. Despite the lack of a predicted coiled-coil signal, the

N-terminal oligomerization domain of HIV-1 Rev could be configured as a helical wheel with two hydrophobic interfaces, each containing residues shown in crystal structures to mediate dimerization or oligomerization of HIV-1 Rev. Similar helical wheels could be configured for most Rev-like proteins, with the exception of those in deltaretroviruses, which lack predicted secondary structure. The prevalence of coiled-coil signals across highly divergent Rev-like proteins, combined with the evolutionary patterns in their distribution and the observation that a coiled-coil-like motif mediates oligomerization in HIV-1 and EIAV Rev, indicate that the coiled-coil motif constitutes an ancestral and homogeneous mechanism of oligomerization in retroviral Rev-like proteins.

Sequence-based computational prediction of coiled-coil signals is generally reliable because coiled-coils have common and well-characterized features [46]. Sequences with residues that conform to the canonical coiled-coil motif are easily identifiable by coiled-coil prediction servers. However, coiled-coil like sequences that deviate from the canonical formulation by a few residues, especially in core registers of the coiled-coil sequence, can be predicted to have weak or no coiled-coil signals. A complementary, structure-based method to identify potential coiled-coil motifs is to represent sequences as helical wheels and compare their interfaces with those of known or predicted coiled-coils. When such analyses were applied to Rev sequences from the HIV-1/SIVcpz lineage, the structure-based signature of a coiled-coil was present, even though the sequences do not perfectly conform to the canonical coiled-coil signature. The protein–protein interfaces formed in the helical wheel configurations of HIV-1 Rev contain the specific residues shown to be involved in HIV-1 Rev dimerization and oligomerization in available crystal structures [29, 37]. Thus, Rev proteins in the HIV-1/SIVcpz lineage likely adopt a functional coiled-coil structure, even though a coiled-coil sequence signal was not readily identifiable by prediction servers.

Canonical coiled-coil motifs typically contain a combination of hydrophobic and hydrophilic residues that constrain oligomerization interactions in a characteristic manner [43, 46]. The coiled-coil-like motif identified here in HIV-1 Rev differs from canonical coiled-coils, such as those found in HIV-2 Rev or EIAV Rev, by the presence of hydrophobic residues at positions that typically contain polar residues. The additional hydrophobic residues in HIV-1 Rev not only mask the coiled-coil sequence signal, but also give rise to a more hydrophobic interface. The potential selective advantage of a more hydrophobic interface is not known; however, it could increase the variety of available hydrophobic interactions and consequent oligomeric conformations, thereby enhancing the plasticity of HIV-1 Rev. Interestingly, a major theme

emerging from recent HIV-1 Rev-RRE-CRM1 structural studies is the plastic nature of the individual components of the ternary complex, especially Rev [17]. The plasticity of HIV-1 Rev oligomerization and RNA binding domains gives rise to a wide range of Rev binding conformations that can interact with different RRE structural conformations and influence functional activity [17, 31, 47]. Increased plasticity resulting from increased hydrophobicity may explain the loss of the canonical coiled-coil sequence signal, but not the structural motif, in the HIV-1/SIVcpz lineage of Rev proteins.

Presumably, the loss of the coiled-coil motif in deltaretroviruses and in some small ruminant lentivirus is also accompanied by a concomitant selective advantage. Modeling of deltaretrovirus Rex proteins (not shown) yields a globular protein with an extremely high degree of disorder and flexibility. Analogous to what we suggest for HIV-1 Rev, this increased flexibility could enable HTLV-1 Rex to bind diverse RRE structures. This hypothesis is supported by the observation that HTLV-1 Rex can functionally replace HIV-1 Rev, while the reverse is not true [48]. Within SRLV Rev proteins, the absence of the coiled-coil signal is often associated with the presence of a helix-breaking proline residue. This interruption in the helical structure most likely results in a more flexible region and could also contribute to plasticity of the protein. In support of this, the predicted structure of these SRLV Rev proteins yields a globular molecule with flexible loops, reminiscent of predicted protein structures of deltaretrovirus Rex proteins (not shown). Interestingly, there is regular cross-species transmission between goat and sheep lentiviruses [49], and the increased flexibility in some SRLV Rev might facilitate interaction with RRE targets from both ovine and caprine lineages. In most cases, therefore, the loss of coiled-coiled signals in retroviral Rev-like proteins is associated with predicted structural changes that would increase structural plasticity and expand the repertoire of potential Rev-RRE interactions.

## Conclusions

Analyses of sequence-based computational predictions of retroviral Rev-like proteins in a phylogenetic context identified an ancestral coiled-coil motif that is maintained across all retroviral genera, with the exception of deltaretroviruses. Helical wheel projections of HIV-1 Rev and other retroviral Rev-like proteins revealed interaction interfaces that could mediate oligomerization. When considered together, these complementary approaches suggest that the coiled-coil motif constitutes an ancestral and homogeneous mechanism of oligomerization in retroviral Rev-like proteins. Results from this study could inform strategies that seek to target Rev oligomerization and inhibit HIV RNA export and replication.

## Methods

### Domain architecture of Rev-like protein and genomic location of their RNA targets

Residue or nucleotide boundaries of arginine rich motifs, nuclear export signals, and Rev RNA targets of representative Rev/Rev-like proteins were mapped onto the corresponding protein/genomic schematic for each member. All schematics were represented according to scale. Accession codes used as references to construct schematics are as follows: HIV-1: K03455; HIV-2: U27200; SIV: X07805; FIV: M36968; EIAV: AF028232; BIV: M32690; CAEV: EU293537; Visna: L06906; HERV-K (HML-2): AAF88167; MMTV: M15122 and ABB02515; BLV: FJ914764; HTLV-1: U19949; HTLV-2: M10060; HTLV-3: DQ093792. The RELIK Rev sequence and motif boundaries used were described in [50], and sequences and motif boundaries for the endogenous lentiviruses mELV and pSIV were described in [51], and [52], respectively.

### Prediction of secondary structural elements

Secondary structure predictions for all Rev-like proteins were obtained using the JPred 3 server (<http://www.compbio.dundee.ac.uk/jpred/index.html>) [53]. JPred 3 provides a 3-state prediction (alpha-helix, beta-strand, other) with an accuracy of 81.5% [53]. Coiled-coil motif predictions were performed using both the CCHMM\_PROF ([http://gpcr.biocomp.unibo.it/cgi/predictors/cchmmprof/pred\\_cchmmprof.cgi](http://gpcr.biocomp.unibo.it/cgi/predictors/cchmmprof/pred_cchmmprof.cgi)) and COILS ([http://www.ch.embnet.org/software/COILS\\_form.html](http://www.ch.embnet.org/software/COILS_form.html)) web servers [39, 41]. For CCHMM\_PROF, default parameters were used; for COILS, the MTIDK matrix was used, with all window lengths, and with 2.5-fold weighting of positions 'a' and 'd' of coiled-coil repeats. All results of coiled-coil predictions for each of the Rev-like sequences used in this study are listed in Additional file 2.

### Phylogenetic reconstruction

Phylogenetic trees were constructed by Bayesian inference with MrBayes [54] v3.2 (<http://mrbayes.sourceforge.net/download.php>) using the rtRev amino acid substitution matrix, developed specifically for retroviral sequences [55]. The primate lentivirus Env tree (Fig. 3) was derived from an alignment of 144 Env sequences using the 2014 HIV Sequence Compendium [56] as a reference for collecting all HIV/SIV sequences. The Pol tree (Fig. 5) was derived from an alignment of 105 retroviral Pol amino acid sequences, and used to infer the overall ancestral state of coiled-coils in retroviral Rev-like proteins. Sequences used in the tree included sequences of publicly available retroviruses that encoded a Rev-like protein for which Pol and Env amino acid sequences were also available. Accession codes for all sequences included in all phylogenetic analyses are listed in Additional file 2.

MrBayes analysis was run for 5,000,000 generations, sampling trees every 1000 generations and discarding the first 25% of samples as the burn-in fraction, as suggested by the authors [54]. Two Bayesian chains were run to ensure adequate mixing. Convergence was indicated by an average standard deviation of split frequencies (ASDSF) <0.05 between the two chains and a potential scale reduction factor (PSRF) value ~1 for all parameters. The 50% majority consensus tree was selected as the final tree. MrBayes analyses were run on the XSEDE cluster using the CIPRES Science Gateway for inference of large phylogenetic trees [57] (<https://www.phylo.org/>) with the BEAGLE library enabled. The Env tree in Fig. 3 was visualized using the iTOL webserver [58] (<http://itol.embl.de/>).

### Ancestral inference

Ancestral state reconstruction was performed with the MESQUITE software v3 [59] (<https://mesquiteproject.wikispaces.com/>) using both parsimony and maximum likelihood methods on each of the phylogenetic trees constructed. For all ancestral inferences performed, the predicted coiled-coil phenotype was modeled as a discrete trait. Maximum likelihood was performed using the “Markov k-state 1 parameter” model (MK1), a generalization of the Jukes–Cantor model in which there is a single parameter, the rate of change, and both forward (gain) or backward (loss) changes are equally likely.

### Sequence alignment, logos, and helical wheels

Sequence alignments were performed with the MAFFT webserver [60] (<http://mafft.cbrc.jp/alignment/server/>) using default parameters. Sequence logos were generated using the WebLogo server [61] (<http://weblogo.berkeley.edu/logo.cgi>), and helical wheels were obtained by using the DrawCoil 1.0 webserver (<http://www.grigoryanlab.org/drawcoil/>).

### Additional files

**Additional file 1.** Distribution of  $\alpha$ -helices and coiled-coil structure in representative Rev-like proteins in (A) non-primate lentiviruses, (B) betaretroviruses, and (C) deltaretroviruses. Residues in red indicate regions of predicted  $\alpha$ -helices; residues in blue represent predicted  $\beta$ -strand regions, and underlined residues represent regions of predicted coiled-coils.

**Additional file 2.** Complete dataset for sequences used in this study. For each SIV, HIV-1, HIV-2, EIAV, FIV, SRLV, BIV, BLV, PTLV, HERVK, MMTV, Jaagsiekte, and endogenous lentivirus Rev-like sequence, accession code information for the genome from which it was derived as well as Pol, Env, and Rev proteins are provided, where available. Results for the prediction of coiled-coils by CCHMM\_PROF ([http://gpcr.biocomp.unibo.it/cgi/predictors/cchmmprof/pred\\_cchmmprof.cgi](http://gpcr.biocomp.unibo.it/cgi/predictors/cchmmprof/pred_cchmmprof.cgi)) and COILS webservers ([http://www.ch.embnet.org/software/COILS\\_form.html](http://www.ch.embnet.org/software/COILS_form.html)) [39, 41] are indicated for each sequence, and the overall correlation between the two servers for all sequences, as well as the correlation for a smaller subset comprising BLV, PTLV, SIV, and HIV-1 are also included.

### Authors' contributions

CNU and SC conceived and designed the study and CNU implemented the research. KD assisted with phylogenetics and ancestry reconstruction and in editing the manuscript. SC supervised the study and SC, KD and DD provided advice throughout. CNU, SC and DD contributed equally to writing the manuscript. All authors read and approved the final manuscript.

### Author details

<sup>1</sup> Department of Animal Science, Iowa State University, Ames, IA 50011, USA. <sup>2</sup> Program in Bioinformatics and Computational Biology, Iowa State University, Ames, IA, USA. <sup>3</sup> Department of Genetics, Developmental and Cell Biology, Iowa State University, Ames, IA, USA. <sup>4</sup> Department of Statistics, Iowa State University, Ames, IA, USA. <sup>5</sup> Present Address: HIV Dynamics and Replication Program, National Cancer Institute, 1050 Boyles St, Frederick, MD 21702, USA.

### Acknowledgements

The authors thank Amanda Kapfer for assistance in preparation of the manuscript.

### Competing interests

The authors declare that they have no competing interests.

### Availability of data and materials

All data analyzed during this study are included in this published article and supplementary information.

### Funding

This work was supported in part by grants from National Institutes of Health CA128568 (SC) and ISU Presidential Initiative for Innovative Research (DD and SC).

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 11 July 2017 Accepted: 11 August 2017

Published online: 22 August 2017

### References

- Sodroski J, Goh WC, Rosen C, Dayton A, Terwilliger E, Haseltine W. A second post-transcriptional trans-activator gene required for HTLV-III replication. *Nature*. 1986;321:412–7.
- Feinberg MB, Jarrett RF, Aldovini A, Gallo RC, Wong-Staal F. HTLV-III expression and production involve complex regulation at the levels of splicing and translation of viral RNA. *Cell*. 1986;46:807–17.
- Malim MH, Hauber J, Le SY, Maizel JV, Cullen BR. The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. *Nature*. 1989;338:254–7.
- Felber BK, Hadzopoulou-Cladaras M, Cladaras C, Copeland T, Pavlakis GN. Rev protein of human immunodeficiency virus type 1 affects the stability and transport of the viral mRNA. *Proc Natl Acad Sci USA*. 1989;86:1495–9.
- Dillon PJ, Nelbock P, Perkins A, Rosen CA. Structural and functional analysis of the human immunodeficiency virus type 2 Rev protein. *J Virol*. 1991;65:445–9.
- Oberste MS, Greenwood JD, Gonda MA. Analysis of the transcription pattern and mapping of the putative rev and env splice junctions of bovine immunodeficiency-like virus. *J Virol*. 1991;65:3932–7.
- Martarano L, Stephens R, Rice N, Derse D. Equine infectious anemia virus trans-regulatory protein Rev controls viral mRNA stability, accumulation, and alternative splicing. *J Virol*. 1994;68:3102–11.
- Schoborg RV, Saltarelli MJ, Clements JE. A Rev protein is expressed in caprine arthritis encephalitis virus (CAEV)-infected cells and is required for efficient viral replication. *Virology*. 1994;202:1–15.
- Meyer BE, Meinkoth JL, Malim MH. Nuclear transport of human immunodeficiency virus type 1, visna virus, and equine infectious anemia virus Rev proteins: identification of a family of transferable nuclear export signals. *J Virol*. 1996;70:2350–9.

10. Hidaka M, Inoue J, Yoshida M, Seiki M. Post-transcriptional regulator (rex) of HTLV-1 initiates expression of viral structural proteins but suppresses expression of regulatory proteins. *EMBO J*. 1988;7:519–23.
11. Rosenblatt JD, Cann AJ, Slamon DJ, Smalberg IS, Shah NP, Fujii J, et al. HTLV-II transactivation is regulated by the overlapping tax/rex nonstructural genes. *Science*. 1988;240:916–9.
12. Indik S, Gunzburg WH, Salmons B, Rouault F. A novel, mouse mammary tumor virus encoded protein with Rev-like properties. *Virology*. 2005;337:1–6.
13. Mertz JA, Simper MS, Lozano MM, Payne SM, Dudley JP. Mouse mammary tumor virus encodes a self-regulatory RNA export protein and is a complex retrovirus. *J Virol*. 2005;79:14737–47.
14. Hofacre A, Nitta T, Fan H. Jaagsiekte sheep retrovirus encodes a regulatory factor, *Rej*, required for synthesis of Gag protein. *J Virol*. 2009;83:12483–98.
15. Lower R, Tonjes RR, Korbmacher C, Kurth R, Lower J. Identification of a Rev-related protein by analysis of spliced transcripts of the human endogenous retroviruses HTDV/HERV-K. *J Virol*. 1995;69:141–9.
16. Pollard VW, Malim MH. The HIV-1 Rev protein. *Annu Rev Microbiol*. 1998;52:491–532.
17. Fernandes J, Jayaraman B, Frankel A. The HIV-1 Rev response element: an RNA scaffold that directs the cooperative assembly of a homo-oligomeric ribonucleoprotein complex. *RNA Biol*. 2012;9:6–11.
18. Meyer BE, Malim MH. The HIV-1 Rev trans-activator shuttles between the nucleus and the cytoplasm. *Genes Dev*. 1994;8:1538–47.
19. Malim MH, Bohnlein S, Hauber J, Cullen BR. Functional dissection of the HIV-1 Rev trans-activator—derivation of a trans-dominant repressor of Rev function. *Cell*. 1989;58:205–14.
20. Truant R, Cullen BR. The arginine-rich domains present in human immunodeficiency virus type 1 Tat and Rev function as direct importin beta-dependent nuclear localization signals. *Mol Cell Biol*. 1999;19:1210–7.
21. Daly TJ, Cook KS, Gray GS, Maione TE, Rusche JR. Specific binding of HIV-1 recombinant Rev protein to the Rev-responsive element in vitro. *Nature*. 1989;342:816–9.
22. Bartel DP, Zapp ML, Green MR, Szostak JW. HIV-1 Rev regulation involves recognition of non-Watson-Crick base pairs in viral RNA. *Cell*. 1991;67:529–36.
23. Tan R, Chen L, Buettner JA, Hudson D, Frankel AD. RNA recognition by an isolated alpha helix. *Cell*. 1993;73:1031–40.
24. Tan R, Frankel AD. Costabilization of peptide and RNA structure in an HIV Rev peptide-RRE complex. *Biochemistry*. 1994;33:14579–85.
25. Battiste JL, Mao H, Rao NS, Tan R, Muhandiram DR, Kay LE, et al. Alpha helix-RNA major groove recognition in an HIV-1 rev peptide-RRE RNA complex. *Science*. 1996;273:1547–51.
26. Jain C, Belasco JG. A structural model for the HIV-1 Rev-RRE complex deduced from altered-specificity rev variants isolated by a rapid genetic strategy. *Cell*. 1996;87:115–25.
27. Jain C, Belasco JG. Structural model for the cooperative assembly of HIV-1 Rev multimers on the RRE as deduced from analysis of assembly-defective mutants. *Mol Cell*. 2001;7:603–14.
28. Daugherty MD, D'Orso I, Frankel AD. A solution to limited genomic capacity: using adaptable binding surfaces to assemble the functional HIV Rev oligomer on RNA. *Mol Cell*. 2008;31:824–34.
29. Daugherty MD, Liu B, Frankel AD. Structural basis for cooperative RNA binding and export complex assembly by HIV Rev. *Nat Struct Mol Biol*. 2010;17:1337–42.
30. Jayaraman B, Crosby DC, Homer C, Ribeiro I, Mavor D, Frankel AD. RNA-directed remodeling of the HIV-1 protein Rev orchestrates assembly of the Rev-Rev response element complex. *Elife*. 2014;3:e04120.
31. Sherpa C, Rausch JW, Le Grice SF, Hammarskjold ML, Rekosh D. The HIV-1 Rev response element (RRE) adopts alternative conformations that promote different rates of virus replication. *Nucleic Acids Res*. 2015;43:4676–86.
32. Jackson PE, Tebit DM, Rekosh D, Hammarskjold ML. Rev-RRE functional activity differs substantially among primary HIV-1 Isolates. *AIDS Res Hum Retroviruses*. 2016;32:923–34.
33. Carpenter S, Dobbs D. Molecular and biological characterization of equine infectious anemia virus Rev. *Curr HIV Res*. 2010;8:87–93.
34. Lee JH, Murphy SC, Belshan M, Sparks WO, Wannemuehler Y, Liu S, et al. Characterization of functional domains of equine infectious anemia virus Rev suggests a bipartite RNA-binding domain. *J Virol*. 2006;80:3844–52.
35. Umunnakwe CN, Loyd H, Cornick K, Chavez JR, Dobbs D, Carpenter S. Computational modeling suggests dimerization of equine infectious anemia virus Rev is required for RNA binding. *Retrovirology*. 2014;11:115.
36. Fang X, Wang J, O'Carroll IP, Mitchell M, Zuo X, Wang Y, et al. An unusual topological structure of the HIV-1 Rev response element. *Cell*. 2013;155:594–605.
37. DiMattia MA, Watts NR, Stahl SJ, Rader C, Wingfield PT, Stuart DI, et al. Implications of the HIV-1 Rev dimer structure at 3.2 Å resolution for multimeric binding to the Rev response element. *Proc Natl Acad Sci USA*. 2010;107:5810–4.
38. Xie L, Kesic M, Yamamoto B, Li M, Younis I, Lairmore MD, et al. Human T-cell leukemia virus type 2 Rex carboxy terminus is an inhibitory/stability domain that regulates Rex functional activity and viral replication. *J Virol*. 2009;83:5232–43.
39. Bartoli L, Fariselli P, Krogh A, Casadio R. CCHMM\_PROF: a HMM-based coiled-coil predictor with evolutionary information. *Bioinformatics*. 2009;25:2757–63.
40. Sharp PM, Hahn BH. Origins of HIV and the AIDS pandemic. *Cold Spring Harb Perspect Med*. 2011;1:a006841.
41. Lupas A, Van Dyke M, Stock J. Predicting coiled coils from protein sequences. *Science*. 1991;252:1162–4.
42. Lupas AN, Gruber M. The structure of alpha-helical coiled coils. *Adv Protein Chem*. 2005;70:37–78.
43. Mason JM, Arndt KM. Coiled coil domains: stability, specificity, and biological implications. *ChemBioChem*. 2004;5:170–6.
44. Abelson ML, Schoborg RV. Characterization of the caprine arthritis encephalitis virus (CAEV) rev N-terminal elements required for efficient interaction with the RRE. *Virus Res*. 2003;92:23–35.
45. Edgcomb SP, Aschrafi A, Kompfner E, Williamson JR, Gerace L, Hennig M. Protein structure and oligomerization are important for the formation of export-competent HIV-1 Rev-RRE complexes. *Protein Sci*. 2008;17:420–30.
46. Grigoryan G, Keating AE. Structural specificity in coiled-coil interactions. *Curr Opin Struct Biol*. 2008;18:477–83.
47. Sloan EA, Kearney MF, Gray LR, Anastos K, Daar ES, Margolick J, et al. Limited nucleotide changes in the Rev response element (RRE) during HIV-1 infection alter overall Rev-RRE activity and Rev multimerization. *J Virol*. 2013;87:11173–86.
48. Rimsky L, Hauber J, Dukovich M, Malim MH, Langlois A, Cullen BR, et al. Functional replacement of the HIV-1 rev protein by the HTLV-1 rex protein. *Nature*. 1988;335:738–40.
49. Shah C, Boni J, Huder JB, Vogt HR, Muhlherr J, Zanoni R, et al. Phylogenetic analysis and reclassification of caprine and ovine lentiviruses based on 104 new isolates: evidence for regular sheep-to-goat transmission and worldwide propagation through livestock trade. *Virology*. 2004;319:12–26.
50. Katzourakis A, Tristem M, Pybus OG, Gifford RJ. Discovery and analysis of the first endogenous lentivirus. *Proc Natl Acad Sci USA*. 2007;104:6261–5.
51. Han GZ, Worobey M. Endogenous lentiviral elements in the weasel family (Mustelidae). *Mol Biol Evol*. 2012;29:2905–8.
52. Gilbert C, Maxfield DG, Goodman SM, Feschotte C. Parallel germline infiltration of a lentivirus in two Malagasy lemurs. *PLoS Genet*. 2009;5:e1000425.
53. Cole C, Barber JD, Barton GJ. The Jpred 3 secondary structure prediction server. *Nucleic Acids Res*. 2008;36:W197–201.
54. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*. 2012;61:539–42.
55. Dimmic MW, Rest JS, Mindell DP, Goldstein RA. rtREV: an amino acid substitution matrix for inference of retrovirus and reverse transcriptase phylogeny. *J Mol Evol*. 2002;55:65–73.
56. Foley B, Apetrei C, Mizrahi I, Rambaut A, Korber B, Leitner T, et al. HIV Sequence Compendium 2014. Los Alamos: Theoretical Biology and Biophysics Group, Los Alamos National Library; 2014.
57. Miller MA, Pfeiffer W, Schwartz T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: Gateway computing environments workshop (GCE) 2010.
58. Letunic I, Bork P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics*. 2007;23:127–8.
59. Maddison WP, Maddison D. Mesquite: a modular system for evolutionary analysis. 2001.
60. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30:772–80.
61. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. *Genome Res*. 2004;14:1188–90.