Highlights:

Two new CRESSDNA viruses are described.

They were found on an *Avipoxvirus* lesion from a tanager species in Ecuador.

They present a replication-associated protein and several open reading frames.

Characterized Intrinsically disordered regions and nuclear localization signals indicate putative capsid proteins.

Both genomes share features with *Circoviridae* although the similarity with the members of the family is below 65%

*

Research Paper

Identification of two novel CRESS DNA viruses associated with an *Avipoxvirus* lesion of a blue-and-gray Tanager (*Thraupis episcopus*)

Michaël A. J. Moens[1], Javier Pérez-Tris[1], Martí Cortey[2]& Laura Benítez[3*]

[1] Department of Zoology and Physical Anthropology. Faculty of Biology. Complutense University of Madrid. Calle José Antonio Novais 12, 28040, Madrid, Spain. Email: m.moens@bio.ucm.es, Tel. +34 913944949, Fax: +34 913944947

[1] Department of Zoology and Physical Anthropology. Faculty of Biology. Complutense University of Madrid. Calle José Antonio Novais 12, 28040, Madrid, Spain. Email: jperez@bio.ucm.es, Tel. +34 913944949, Fax: +34 913944947

[2]IRTA-CReSA, Campus de la Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain. Email: marti.cortey@irta.cat, Tel. +34 934674040 (Ext 1708), Fax: +34 5814490

[3] Department of Microbiology III, Complutense University of Madrid. Calle José Antonio Novais 12, 28040, Madrid, Spain. Email: lbenitez@bio.ucm.es , Tel. +34 913944963, Fax: +34 913944964

***Corresponding Author:** Laura Benítez, Email: lbenitez@bio.ucm.es

Keywords: CRESS DNA virus; Tanager; Ecuador; Intrinsically Disordered Regions

Abstract

The discovery of circular rep-encoding single stranded (CRESS) DNA viruses has increased spectacularly over the past decade. They represent the smallest animal viruses known worldwide infecting a wide variety of invertebrates and vertebrates in different natural and human-made environments. The extremely low similarity of nucleotide and protein sequences among different CRESS DNA genomes has challenged their classification. Moreover, the existence of putative capsid proteins (Cp) remains difficult to demonstrate which is crucial to understand the structural properties of these viruses. Here we describe two unclassified CRESS DNA viruses isolated from a cutaneous lesion, caused by a strain of *Avipoxvirus*, from a blue and gray tanager (*Thraupis episcopus*) in Southern Ecuador. Both viruses present replication-associated proteins (Rep) and one to two open reading frames (ORF), one of which represents a putative Cp. The two new Rep are long proteins characterized by the existence of the several highly conserved amino acid residues characteristic of rolling circle replication. Within the putative Cp we detected intrinsically disordered regions (IDR), potential protein and DNA binding regions, and nuclear localization signals (NLS), providing further evidence of presumed Cp. Despite being found on the same host lesion, both viruses show low similarity between each other (<60%) and other known CRESS DNA viruses. Furthermore, we analyze the evolutionary relationships within the CRESS DNA diversity. Additional sampling is needed to explore the possible pathogenetic effects, prevalence and diversity (both phylogenetical and structural) of these viruses in wild bird populations.

1. Introduction

Many CRESS DNA (circular rep-encoding single stranded) viruses have been discovered in a wide diversity of natural and human-made environments, such as estuaries (Dayaram et al., 2015), sewage oxidation ponds (Kraberger et al., 2015), oceans (Labonté and Suttle, 2013) and in many animal groups spanning from arthropods to vertebrates, including humans (Halary et al., 2016; Rosario et al., 2015; Steel et al., 2016). These viruses are the smallest known viruses to infect eukaryotic organisms and have the potential to evolve rapidly due to their high recombination and mutation rates (Duffy et al., 2008; Lefeuvre et al., 2009; Martin et al., 2011), which make them good candidates to develop into emerging pathogens (Rosario et al., 2012). Their genome is usually smaller than 6 kb and they use a rolling circle mechanism for genome replication, also relying on host cellular proteins for successful replication (Rosario et al., 2012). Many CRESS DNA viruses present novel genome organizations, although they include similarities with Rep (replication-associated protein) of other viruses. Still the function of other proteins such as capsid proteins (Cp) needs to be explored in order to understand the mechanisms how these viruses replicate in a wide diversity of hosts (Rosario et al., 2015).

In birds several ssDNA viruses have been detected, mainly belonging to the *Circoviridae* family: psittacine beak and feather disease virus (BFDV), chicken anemia virus (CAV) and circovirus of pigeon, goose, gull, canary, raven, duck, swan, finch and starling, all of them considered different species (Breitbart et al., 2017). According to the International Committee on Taxonomy the family *Circoviridae* comprises two genera: *Circovirus*, which include the most of these avian viruses and *Cyclovirus*. The genus *Gyrovirus*, where CAV is included, has been recently moved to the family of *Anelloviridae* (Rosario et al., 2017). More than 70 species have been described in the family, based on the species demarcation threshold of 80% genome-wide nucleotide sequence identity (Breitbart et al., 2017). The genomes of this family typically have a circular ambisense organization consisting of two or more major open reading frames (ORFs), encoding a Rep and a Cp and an origin of replication marked by a predicted stem loop structure exhibiting the conserved nonamer motif (5'-NANTATTAC-3'), located between the

5’-ends of both ORFs (Rosario et al., 2012). Both genera are distinguished by the length of intergenic regions, the position of the origin of replication or the existence of one intron inside the Rep-enconding ORF in *Cyclovirus* (Breitbart et al., 2017). Avian circoviruses have been associated with a variety of illness symptoms such as immuno-supression, delayed growth, feather disorders and developmental abnormalities (Stewart et al., 2006; Todd, 2000, 2004). These viruses are probably widespread and diverse, but they are poorly studied. Therefore, it is urgent to improve our knowledge of their phylogenetic and functional diversity, their ecology (how they interact with their hosts or other pathogens) and their biogeography.

In this paper, we describe a singular yet informative case in which two new genomes of CRESS DNA viruses were isolated from a pox lesion of a common tropical bird, the blue-and-gray tanager (*Thraupis episcopus*). This bird from which the two new genomes were identified was found infected with an *Avipoxvirus* strain which caused a cutaneous lesion on the foot (Moens et al., 2017). Moreover we analyze the structure of the genomes and the characterization of the potential ORFs. Because of the low identity of two novel putative Cps, we investigate the presence of motifs in the amino acid sequences and the existence of intrinsically disordered regions (IDRs), which are regions within a protein that lack an ordered structure (He et al., 2009). These IDRs allow a protein to exist in different states depending on the substrate they interact with (Dunker et al., 2001) and could reveal the possible existence of Cp in CRESS DNA viruses (Rosario et al., 2015). Within these IDRs we aim to detect the presence of protein binding regions which are disordered in isolation but which can undergo disorder-to-order transition upon binding (Mészáros et al., 2009). Finally, we will analyze their evolutionary relationships within the known diversity of CRESS DNA viruses.

2. Materials and Methods

The case reported here was sampled during a mist-netting campaign designed to investigate the prevalence of various avian pathogens in wild bird communities of the Ecuadorian Andes, which involved blood sampling and revision of possible virus infections. The list of screened pathogens included viruses

causing skin lesions like poxvirus and papillomavirus. In total, we inspected 941 birds of 135 species, of which six showed cutaneous lesions on their feet compatible with virus infection. The lesions were weighted, homogenized by mechanical force using a sterile plastic crusher and viral DNA was extracted with a standard phenol-chloroform-isoamyl protocol followed by isopropanol precipitation (Pérez-Tris et al., 2011). We extracted total DNA from blood samples with a standard ammonium acetate protocol (Green et al., 2012). DNA-extracts were controlled for sample quality on an agarose gel stained with Gel Red. We estimated the total nucleic acid concentration with a Nanodrop ND100 system (Nanodrop Technologies, ThermoScientific, Wilmington, DE) and stored the samples at -20 °C. First a multiplex PCR, designed for the combined detection of *Avipoxvirus* and *Papillomavirus*, was performed under conditions described by Pérez-Tris et al. (2011). This was done with a total reaction volume of 25 µl with 50 pmol/ml of each primer, 0.8 mM of each deoxynucleotide triphospate, 4.0 mM of MgCl2, 0.125 µl AmpliTaq DNA polymerase and 2.5µl of Buffer solution 10x (Applied Biosystems, Warrington, UK) under the following PCR conditions: 3 min. of initial denaturation (95 °C), 45 cycles of 95°C for 1 min., 50°C for 1 min. and 72°C for 1 min., and a final extension step at 72°C for 5 min. Samples which tested positive for poxvirus were visualized on an 2 % agarose gel and were further analysed by amplifying part of the P4b core protein gene (Lee and Lee, 1997). Amplified samples were visualized on an agarose gel, one positive sample for *Avipoxvirus* was sequenced by Macrogen (Netherlands) and the sequence was compared with known P4b-isolates on Genbank.

DNA was amplified using randomly-primed rolling cycle amplification (RCA) (Johne et al., 2009) using TempliphiTM 100 Amplification (GE Healthcare) following the manufacturer's instructions (0.5 µl of DNA in a total reaction volume of 10 µl). Out of six birds with lesions, three tested positive for *Avipoxvirus*, one of which also produced RCA products.

The RCA positive bird was captured on the 22 of June 2012 near San Pedro de Vilcabamba (4°15'S, 79°13'W, 1800 m elevation), in a suburban mosaic of scrub and riverine vegetation on the Rio Chamba river banks. After the completion of the RCA reaction the product was digested with the *EcoRI* enzyme.

We ran an agarose gel of 1.5 % stained with ethidium bromide to visualize the digested product. Four DNA fragments with different intensities were extracted from the agarose gel (QIAquick Gel Extraction Kit, QIAGEN, Germany) and cloned. The vector pUC19, cut with *EcoRI* and processed with shrimp alkaline phosphatase (Roche Applied Sciences) to avoid re-ligation, was ligated to the RCA digested fragments in a total volume of 10 µl with the T4 DNA ligase (Roche Applied Sciences). One shot TOPO10 competent *E. coli* (Invitrogen) was transformed with the resulting plasmids. We extracted the plasmid DNA from recombinant clones with a QIAprep Miniprep Spin kit (Qiagen).

Next, successful cloning products were sequenced. After a BLAST search on GenBank we detected nucleotide similarity with Rep proteins of other CRESS DNA viruses. We designed two sets of back-to-back primers located inside of two sequenced fragments corresponding to both Rep proteins and two sets of conventional (non-overlapping) primers (Supplementary table 1). A standard PCR with both types of primers was performed on the RCA products with TaKaRa taq polymerase (TaKaRa Bio Inc, Otsu, Shiga, Japan). Additionally, a PCR with back-to-back sets of primers was done on the original DNA extraction of the lesion (1µl). We ran an extra PCR on the extraction of the blood sample (2µl) of the infected bird to control if the infection can be detected in the blood. Both PCRs were run under the following conditions: 1 min. of initial denaturation (94 °C), 30 cycles of 98°C for 10 seconds, 43°C for 30 seconds and 68°C for 10 min., and a final extension step at 72°C for 10 min. A 2 % agarose gel was run to visualize the PCR products.

The multiple sequences obtained from all protocols were assembled into two distinct genomes using the DNAstar software (Madison, Wisconsin, USA). The genomes were compared to others by means of a BLAST search of the GenBank database. Putative ORFs were assigned by means of SMS ORF Finder (http://www.bioinformatics.org/sms2/orf_find.html). Secondary structures of the proteins were analysed by a web-based version of mfold (Zuker and Jacobson, 1995). Tandem repeats were analysed by Tandem Repeats Finder (Benson, 1999).

In order to study the phylogenetic position of both Reps within the CRESS DNA diversity we performed a Bayesian analysis with BEAST 2.0 (Bouckaert et al., 2014). We decided to infer phylogenetic

relationships of the Rep since the amino acids are easily aligned compared to the other ORFS, which do not show many similarities among each other. Moreover, the Rep has been proven useful in deciphering evolutionary relationships in CRESS DNA viruses. We aligned amino acid sequences of the Rep of 159 CRESS DNA viruses, based on the highest similarities found in the BLAST searching GenBank. We used the most appropriate substitution model for the Rep amino acid sequences according to the Bayesian Information Criterion implemented in MEGA 5.2 (Tamura et al., 2011): LG+G. We specified the parameters for the BEAST-run in BEAUTI 2.0 (Bouckaert et al., 2014) and Monte Carlo Markov Chains (MCMC's) were run for 10^9 generations, sampling every 100.000 trees. Traces were inspected for convergence with Tracer 1.5 (Rambaut and Drummond, 2007). The 10.000 resulting trees were summarized with TreeAnnotator v2.1.2 (Rambaut and Drummond, 2007) and the phylogenies with the posterior probabilities of the nodes were displayed in FigTree v1.4.2 (Maddison and Maddison, 2011) for further analysis. We plotted the host group in which all CRESS DNA viruses were found to analyze if these viruses infect hosts of restricted or wide phylogenetic ancestry. In order to identify possible patterns of similarity of the Rep of these viruses with different CRESS virus types already described in the literature, a pairwise identity matrix of the same Rep and genomes was created with the sequence demarcation tool of the SDT software (Muhire et al., 2014). We used the same selection of 159 CRESS DNA genomes and their correspondent Reps. All sequences were aligned with MUSCLE (Edgar, 2004) within the SDT software.

In order to predict nuclear localization signals and DNA binding residues in the amino acid sequences of both genomes we used PredictNLS (Cokol et al., 2000) and BindN (Wang and Brown, 2006) respectively. We predicted intrinsically unstructured regions within the putative ORF's with the IUPred software (Dosztanyi et al., 2005). IUPred calculates a pairwise energy profile along the amino acid sequence, where values above 0.5 indicate disordered regions within the protein. Results were also compared with the DisProt VL3 disorder predictor (Obradovic et al., 2003; Sickmeier et al., 2007) and the PONDR-fit software which combines several predictors like VL3, VL2 and VLXT (Xue et al., 2010). Protein binding

178 regions within the IDR's were predicted with the ANCHOR software (Dosztányi et al., 2009). ANCHOR
179 pursues to identify segments within disordered regions, which cannot form enough favorable intra-chain
180 interactions to fold on their own and which are likely to gain stabilizing energy by interacting with a
181 globular protein partner (Dosztányi et al., 2009). ANCHOR then generates a probability score profile of
182 the amino acid residues, indicating the likelihood of the residue to be a part of a disordered region along
183 the sequence (Dosztányi et al., 2009). Regions with scores above 0.5 indicate disordered binding regions.
184 The novel CRESS DNA viruses were uploaded to GenBank under the following accession numbers
185 (MF804497 and MF804498).

186 **3. Results and Discussion**

187 A cutaneous lesion sample from a blue and gray tanager (*Thraupis episcopus*), where the existence of an
188 *Avipoxvirus* has been previously confirmed (KU356758) (Moens et al., 2017), was analysed by RCA to
189 detect a possible coinfection with papillomavirus and various fragments were obtained after digestion
190 with *EcoRI* (Fig. 1A). The lack of equimolar amounts of DNA suggested at least two different viruses,
191 indicating the possible existence of more than one original virus. Four of these fragments were cloned and
192 sequenced showing two very divergent Rep-encoding genes. The design of several sets of back-to-back
193 (overlapping) and non-overlapping primers along the Rep-encoding regions and several walking primers
194 allowed the complete sequencing of two different circular viral genomes (Fig. 1B). We successfully
195 amplified one of both viruses with a PCR on the original lesion extract (Fig. 1C). The second virus was
196 not amplified by PCR on the lesion extract (Fig. 1C), probably due to its low concentration. We did not
197 amplify the viruses in the blood sample (Fig. 1C). After all analyses we effectively assembled two
198 CRESS DNA virus genomes. We named them Tanager-associated CRESS DNA virus TaCV1 (3,398 nt)
199 and TaCV2 (2,967 nt) (Fig. 2). TaCV1 showed an ambisense genome organization with two predicted
200 main ORFs of 1,182 and 624 nt long. The longer ORF encodes a Rep of 393 amino acids in sense
201 direction and the second a protein of 207 amino acids on the complementary-sense strand which could
202 correspond to the Cp. A third open reading frame (ORF3) of 420 nts (139 aa), partially overlapping with

Cp-encoding ORF, has been identified but we do not know if it is functional. TaCV2 showed also an ambisense genome organization, however, the two major ORFs are arranged in the same orientation and they encode the Rep (321 aa) and the Cp (174 aa). Similarly, an ORF3 (166 aa) on the complementary strand was detected. TaCV1 had four *EcoRI* restriction sites while TaCV2 contained a single one, which coincides with the results of RCA-digestion (Fig. 1A). The top fragment corresponds to TaCV2 while the remaining fragments belong to TaCV1 (Fig. 1A). The intensity of the latter fragments is lower suggesting a smaller concentration of the initial virus load for TaCV1 (Fig. 1A).

The 5´intergenic regions (IR), where the origin of viral replication (*ori*) is placed, are located between the 5´ends of the Rep and Cp-encoding ORFs in TaCV1 and the 3´ends of the ORF3 and Rep-encoding ORF in TaCV2 (Fig. 2B and 2C). These IR regions are characterized because they are exceptionally long (978 nts in TaCV1 and 873 nts in TaCV2) and contain a high amount of A-T rich regions (approx. 65%), contrarily to described members of the familiy *Circoviridae* (Rosario et al., 2017). The *ori* of TaCV1 and TaCV2 are integrated in a typically stem-loop structure with a putative hairpin (14/13 nt in TaCV1/TaCV2), coincident to the length described in *Circovirus*, plus a well conserved nonanucleotide motif (5-´T/GAGTATTAC-3´), where rolling circle replication is initiated in circular ssDNA replicons (Fig. 2B and 2C). Moreover 3.5 units of tandem direct repeats of a 26 nt sequence (TaCV1) and 2.6 units of a 23 nt (TaCV2), which usually are associated to promoter-enhancer activity, were located in both noncoding regions (Fig. 2 B and 2C). Adjacently to the hairpin structure in TaCV1, two tandem repeats were found of the sequence 5'-GGAGCCA-3', described as putative binding sites for the Rep (Phenix et al., 2001) (Fig. 2 B and 2C).

The two long Rep of these novel CRESS DNA viruses exhibit the three highly conserved sequences located at the N-terminus, known to be involved in rolling circle replication (RCR): motif I (16-FTIFN/24-FTLNN), motif II (54-PHIQG/57-PHLQG) and motif III (100-YITK/96-YCKK) in TaCV1/TaCV2 (Rosario et al., 2017). Moreover, some of the motifs related to the superfamily 3 helicase have been identified within them: Walker A (187-GPSGSGKS/-169-GESGSGKS in TaCV1/TaCV2),

228 Walker B (206-IIDDF in TaCV2) and motif C slightly modified (311-KLSN) in TaCV1 (Rosario et al.,
229 2017). Interestingly, highly conserved human and avian circovirus/cyclovirus motifs (125-WWNGY) and
230 (226-DRYP) have been found in TaCV2-Rep (de Sales Lima et al., 2015; Garigliany et al., 2014).

231 The Reps of both genomes showed low similarity (26% identity, 100% coverage) between each other.
232 However, the TaCV1-Rep showed higher similarity with published Rep of a strain of *Penguinpox virus*
233 (30%, 48% coverage), and the TaCV2-Rep was more similar to *Canarypox virus* (61%, 88% coverage).
234 The similarity with other *Circovirus* or CRESS-DNA Reps is lower than 47%. The pairwise identity
235 matrix shows that the TaCV1-Rep shares 32% pairwise identity (100% coverage) with *Acartia tonsa*
236 copepod circovirus (Genbank Accession nr: AFN42891.1) while the TaCV2-Rep shares 47% identity
237 (100% coverage) with two Reps belonging to bat circoviruses (Genbank Accession nrs: AEL28813,
238 AFH02742) (Fig. 3A). Interestingly, the nucleotide identity of the ORF encoding TaCV2-Rep even
239 reaches 81 % identity (48% coverage of total *rep* gene towards the 5-ends) with the *Canarypox virus rep*
240 gene. This suggests that it is probably a recombinant sequence, where the 5´and 3´ends have distinct
241 evolutionary origins, especially considering the existence of an avipoxvirus identified in the same lesion
242 (Moens et al., 2017). In fact, the existence of recombination of circoviruses Rep-encoding genes from
243 unrelated viruses has been proposed (Gibbs et al., 2006). Recombination and reassortment among ssDNA
244 viruses has not been studied in detail compared to dsDNA viruses and this warrants further research.
245 Different functional domains of CRESS DNA virus genes can have origins in different virus families,
246 which increases their capacity to exploit new niches and to switch hosts (Krupovic et al., 2015; Lefeuvre
247 and Moriones, 2015), but complicates their characterization and classification.

248 The SDT genome analysis shows that TaCV1 shares 64% pairwise identity with a bat circovirus from
249 China (JN377580) and several uncultured marine viruses (Genbank Accession nrs: JX904147, JX904605,
250 JX904185, JX904344) (Dunlap et al., 2013; Labonté and Suttle, 2013) while TaCV2 is 64% identical to
251 bat guano circovirus (Genbank Accession nr: HM228875) (Linlin et al., 2010) (Fig. 3B). The Bayesian
252 analyses of the Rep reveals that the TaCV1-Rep forms a sister clade with a wide diversity of CRESS

DNA Reps, in which the Rep protein of TaCV2 is embedded (Fig. 4). The Rep protein of TaCV2 is most related to Rep proteins of bat circoviruses (Genbank Accession nrs: AFH02742, AEL28813) (Fig. 4). The clade in which both viruses are placed infect hosts of wide phylogenetic origins, including insects and mammals. Moreover, a considerable part of viruses in this clade were found in marine environments (Fig. 4).

The second longest ORFs of TaCV1 and TaCV2 could encode the Cp, although they did not show any significant similarity using BLAST or BLASTX on GenBank. However, analysis *in silico* suggests that both of them could represent putative Cps. Firstly, the two putative Cp of TaCV1 (207 aa) and TaCV2 (174 aa) are basic proteins of 24 and 20 KDa with an isoelectric point (pI) of 9.66 and 9.54 respectively, as opposed to ORF3 of TaCV2 with pI 5. Basic amino acids (K+R) represented 20% of the total amino acids in both TaCV1/2-Cps, but were not concentrated exclusively in the N-terminal region such as in other Cp of circoviruses, which normally contain several arginine-clusters. This highly basic region seems to be involved in packing of the viral genome into the viral capsid (Johne et al., 2004). On the other hand, the existence of a putative bipartite nuclear location signal (NLS) corresponds to the classic consensus sequence $[(K/R)_2X_{10\text{-}12}(K/R)_3]$ which was located in the N-terminal region of Cp in TaCV1 (3-**KR**HTRSYLQEINTF**KKK**-19). This motif was first described in *Xenopus laevis* nucleoplasmine and has been found in varicella zoster virus (Huang et al., 2014). It is known that the Cp participates in the attachment, entry and shuttling of the viral genome across the nuclear pore complex because for these viruses the host cell nucleus is the site of virus replication (Trible and Rowland, 2012). Overlapping with this region a DNA binding domain (2-SKRHTRSY-9) was identified in the same region. In contrast, no canonical NLS was found in Cp of TaCV2, although there was a potential motif (9-**RR**Y**KK**-13), slightly different from a type of monopartite NLS containing 3-5 basic amino acids with the weak consensus KR/KXR/K, similar to simian virus 40 large antigen (Chelsky et al., 1989). Additionally, the same motif was identified as a DNA binding region. Finally, the data derived from the energy profiles and putative binding regions of the potential Cp (Mészáros et al., 2009; Rosario et al., 2015) point in the same way.

Based on the pairwise energy profiles created by IUPred, the potential Cp of TaCV1 and TaCV2 seem to be intrinsically disordered regions (IDR) over their entire length with the exception of a region between residues 10 and 50 (Fig. 5). Both proteins show a very similar pairwise energy profile (Fig. 5). However, they are non-matching profiles with the conserved patterns proposed in other CRESS-DNA viruses (Rosario et al., 2015), although structural proteins of different viruses contain IDRs (Liu and Huang, 2014). It has been shown that smaller viruses exhibit more intrinsic disorder which may be involved in encoding multifunctional proteins (Pushker et al., 2013; Xue et al., 2012). The found residues also show α-helical structure, possibly relevant to the physiological structure of this protein. Within these residues ANCHOR detected putative binding regions when interacting with other globular proteins (Fig. 5). These represent sites that can undergo a disorder-to-order transition upon binding (Mészáros et al., 2009). We did not detect IDRs in the protein encoded by ORF3 of TaCV2 with IUPred but further analyses with PONDR-fit and DISPROT VL3 detected an IDR between residues 100 and 150 (Fig. 5). Moreover, at least four regions can be identified as potential binding regions in TaCV1-Cp and TaCV2-Cp, while none is recognized in the protein encoded by ORF3 in TaCV2 by the ANCHOR-software (Fig. 5).

In summary, we describe two novel CRESS DNA genomes isolated from a lesion caused by an avipoxvirus. They share many features with members of the family *Circoviridae*, and may be considered as new species (Breitbart et al., 2017), because the similarity of the complete genome with other circovirus genomes is under 80%. However the atypical size of the viral genomes and the 5´IR, and the architecture of TaCV2 could challenge their classification. Both viruses have been detected in a sample of a cutaneous pox lesion (wart-like growths) where a strain of *Fowlpox virus* had been also identified. We do not exclude the tissue can be contaminated by fecal material, because many CRESS DNA virus have been identified in fecal samples from birds (Mahzounieh et al., 2014; Stenzel et al., 2015). However circoviruses and cycloviruses have also been reported in other samples as chicken muscle tissue or spleen (Johne et al., 2006; Li et al., 2016). Little is known about the prevalence of CRESS DNA viruses in wild avian communities and this warrants further research. Experiments are needed to understand the potential

impacts of these viruses on avian health. Since both new viruses were detected in a common tanager species, future studies on the prevalence of CRESS DNA viruses in the spectacular diversity of tanagers are warranted.

Acknowledgements

The fieldwork was funded through a travel grant from the Ministry of Economy and Competitiveness (MINECO, EEBB-I-12-05798) from Spain and hosted by Nikolay Aguirre at the Universidad Nacional de Loja in Ecuador. Joy Horton and Curtis Hofmann kindly provided accommodation during field work. The collection permits were provided by the Ministry of Environment of the Loja province under the research project N°009-2012-IC-FAU-DPL-MA. The samples were transported under the following exportation permits: CITES 021/VS and 016-2012-IC-FLO-DPL-MAD. Access to genetic resources in this study was granted by Ministerio del Ambiente (Ecuador), under research permit MAE- DNBCM-2015-0017. Our research was funded by the Spanish Ministry of Economy and Competitiveness (projects CGL2010-15734/BOS and CGL2013-41642-P/BOS to JP).

References

Benson, G., 1999. Tandem Repeats Finder: a program to analyse DNA sequences. Nucleic Acids Res. 27, 573–578.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., 2014. BEAST 2: A software platform for bayesian evolutionary analysis. PLoS Comput. Biol. 10, e1003537.

Breitbart, M., Delwart, E., Rosario, K., Segales, J., Varsani, A., Consortium, I.R., 2017. ICTV virus taxonomy profile: Circoviridae. J. Gen. Virol. 53–54. doi:10.1099/jgv.0.000871

Chelsky, D., Ralph, R., Jonak, G., 1989. Sequence requirements for synthetic peptide-mediated translocation to the nucleus. Mol Cell Biol 9, 2487–2492. doi:10.1128/MCB.9.6.2487

Cokol, M., Nair, R., Rost, B., 2000. Finding nuclear localization signals. EMBO Rep. 1, 411–415. doi:10.1093/embo-reports/kvd092

Dayaram, A., Goldstien, S., Argüello-Astorga, G.R., Zawar-Reza, P., Gomez, C., Harding, J.S., Varsani, A., 2015. Diverse small circular DNA viruses circulating amongst estuarine molluscs. Infect. Genet. Evol. 31, 284–295. doi:10.1016/j.meegid.2015.02.010

de Sales Lima, F., Cibulski, S.P., Fernandes dos Santos, H., Teixera, T.F., Muterla Varela, A.P., Roehe,

P.M., Delwart, E., Franco, A.C., 2015. Genomic characterization of novel circular ssDNA viruses from insectivorous bats in southern Brazil. PLoS One 2, 1–11. doi:10.1371/journal.pone.0118070

Dosztanyi, Z., Csizmok, V., Tompa, P., Simon, I., 2005. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics 21, 3433–3434. doi:10.1093/bioinformatics/bti541

Dosztányi, Z., Mészáros, B., Simon, I., 2009. ANCHOR: Web server for predicting protein binding regions in disordered proteins. Bioinformatics 25, 2745–2746. doi:10.1093/bioinformatics/btp518

Duffy, S., Shackelton, L.A., Holmes, E.C., 2008. Rates of evolutionary change in viruses: patterns and determinants. Nat. Rev. Genet. 9, 267–76. doi:10.1038/nrg2323

Dunker, A.K., Lawson, J.D., Brown, C.J., Williams, R.M., Romero, P., Oh, J.S., Oldfield, C.J., Campen, A.M., Ratliff, C.M., Hipps, K.W., Ausio, J., Nissen, M.S., Reeves, R., Kang, C., Kissinger, C.R., Bailey, R.W., Griswold, M.D., Chiu, W., Garner, E.C., Obradovic, Z., 2001. Intrinsically disordered protein. J. Mol. Graph. Model. 3263, 26–59.

Dunlap, D.S., Ng, T.F.F., Rosario, K., Barbosa, J.G., Greco, A.M., Breitbart, M., Hewson, I., 2013. Molecular and microscopic evidence of viruses in marine copepods. Proc. Natl. Acad. Sci. U. S. A. 110, 1375–80. doi:10.1073/pnas.1216595110

Edgar, R.C., 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797. doi:10.1093/nar/gkh340

Garigliany, M.-M., Hagen, R.M., Frickmann, H., May, J., Schwarz, N.G., Perse, A., Jost, H., Borstler, J., Shahhosseini, N., Desmecht, D., Mbunkah, H.A., Mbunkah, A., Kingsley, M.T., de Mendonca Campos, R., Salete de Paula, V., Randriamampionona, N., Poppert, S., Tannich, E., Rakotozandrindrainy, R., Cadar, D., Schmidt-Chanasit, J., 2014. Cyclovirus CyCV-VN species distribution is not limited to Vietnam and extends to Africa. Sci. Rep. 4, 1–7. doi:10.1038/srep07552

Gibbs, M.J., Smeianov, V. V, Steele, J.L., Upcroft, P., Efimov, B.A., 2006. Two families of Rep-like genes that probably originated by interspecies recombination are represented in viral, plasmid, bacterial, and parasitic protozoan genomes. Mol. Biol. Evol. 23, 1097–1100. doi:10.1093/molbev/msj122

Green, M., Hughes, H., Sambrook, J., MacCallum, P., 2012. Molecular Cloning: A Laboratory Manual (Fourth Edition). CSH Press.

Halary, S., Duraisamy, R., Fancello, L., Monteil-bouchard, S., Jardot, P., Biagini, P., Gouriet, F., Raoult, D., Desnues, C., 2016. Novel single-stranded DNA circular viruses in pericardial fluid of patient with recurrent pericarditis. Emerg. Infect. Dis. 22, 1839–1841.

He, B., Wang, K., Liu, Y., Xue, B., Uversky, V.N., Dunker, A.K., 2009. Predicting intrinsic disorder in proteins : an overview 929–949. doi:10.1038/cr.2009.87

Huang, Y., Zhang, J., Halawa, M.A., Yao, S., 2014. Nuclear localization signals of varicella zoster virus ORF4. Virus Genes 48, 243–251. doi:10.1007/s11262-013-1006-z

Johne, R., Fernandez-de-Luco, D., Hofle, U., Muller, H., 2006. Genome of a novel circovirus of starlings, amplified by multiply primed rolling-circle amplification. J. Gen. Virol. 87, 1189–1195. doi:10.1099/vir.0.81561-0

Johne, R., Mu, H., Rector, A., Ranst, M. Van, Stevens, H., 2009. Rolling-circle amplification of viral

DNA genomes using phi29 polymerase 205–211. doi:10.1016/j.tim.2009.02.004

Johne, R., Raue, R., Grund, C., Kaleta, E.F., Müller, H., 2004. Recombinant expression of a truncated capsid protein of beak and feather disease virus and its application in serological tests. Avian Pathol. 33, 328–336. doi:10.1080/0307945042000220589

Kraberger, S., Argüello-astorga, G.R., Greenfield, L.G., Galilee, C., Law, D., Martin, D.P., Varsani, A., 2015. Characterisation of a diverse range of circular replication-associated protein encoding DNA viruses recovered from a sewage treatment oxidation pond. Infect. Genet. Evol. 31, 73–86. doi:10.1016/j.meegid.2015.01.001

Krupovic, M., Zhi, N., Li, J., Hu, G., Koonin, E. V., Wong, S., Shevchenko, S., Zhao, K., Young, N.S., 2015. Multiple layers of chimerism in a single-stranded DNA Virus discovered by deep sequencing. Genome Biol. Evol. 7, 993–1001. doi:10.1093/gbe/evv034

Labonté, J.M., Suttle, C. a, 2013. Previously unknown and highly divergent ssDNA viruses populate the oceans. ISME J. 7, 2169–2177. doi:10.1038/ismej.2013.110

Lee, L.H., Lee, K.H., 1997. Application of the polymerase chain reaction for the diagnosis of fowl poxvirus infection 63, 113–119.

Lefeuvre, P., Lett, J.-M., Varsani, a, Martin, D.P., 2009. Widely conserved recombination patterns among single-stranded DNA viruses. J. Virol. 83, 2697–2707. doi:10.1128/JVI.02152-08

Lefeuvre, P., Moriones, E., 2015. Recombination as a motor of host switches and virus emergence : geminiviruses as case studies. Curr. Opin. Virol. 10, 14–19. doi:10.1016/j.coviro.2014.12.005

Li, L., Shan, T., Soji, O.B., Alam, M., Kunz, T.H., Zaidi, S.Z., Delwart, E., 2016. Possible cross-species transmission of circoviruses and cycloviruses among farm animals. J. Gen. Virol. 92, 768–772. doi:10.1099/vir.0.028704-0

Linlin, L., Victoria, J.G., Wang, C., Jones, M., Fellers, G.M., Kunz, T.H., Delwart, E., 2010. Bat guano virome: predominance of dietary viruses from insects and plants plus novel mammalian viruses. J. Virol. 84, 6955–6965. doi:10.1128/JVI.00501-10

Liu, Z., Huang, Y., 2014. Advantages of proteins being disordered. Protein Sci. 23, 539–550. doi:10.1002/pro.2443

Maddison, W.P., Maddison, D., 2011. Mesquite: a modular system for evolutionary analysis. Version 2.75.

Mahzounieh, M., Khoei, H.H., Shamsabadi, M.G., Dastjerdi, A., 2014. Detection and phylogenetic characterization of Columbid circoviruses in Chaharmahal va Bakhtiari province, Iran. Avian Pathol. 37–41. doi:10.1080/03079457.2014.966648

Martin, D.P., Biagini, P., Lefeuvre, P., Golden, M., Roumagnac, P., Varsani, A., 2011. Recombination in eukaryotic single stranded DNA viruses. Viruses 3, 1699–1738. doi:10.3390/v3091699

Mészáros, B., István, S., Dosztányi, Z., 2009. Prediction of protein binding regions in disordered proteins. PLoS Comput. Biol. 5, e1000376. doi:10.1371/Citation

Moens, M.A.J., Pérez-tris, J., Milá, B., Benítez, L., 2017. The biological background of a recurrently emerging infectious disease: prevalence, diversity and host specificity of Avipoxvirus in wild Neotropical birds. J. Avian Biol. 48, 1041–1046. doi:10.1111/jav.01240

412 Muhire, B.M., Varsani, A., Martin, D.P., 2014. SDT : A virus classification tool based on pairwise
413 sequence alignment and identity calculation 9. doi:10.1371/journal.pone.0108277

414 Obradovic, Z., Peng, K., Vucetic, S., Radivojac, P., Brown, C.J., Dunker, A.K., 2003. Predicting Intrinsic
415 Disorder From Amino Acid Sequence 572, 566–572.

416 Pérez-Tris, J., Williams, R.A.J., Abel-Fernández, E., Barreiro, J., Conesa, J.J., Figuerola, J., Martinez-
417 Martínez, M., Ramírez, Á., Benítez, L., 2011. A multiplex PCR for detection of poxvirus and
418 Papillomavirus in cutaneous warts from live birds and museum skins. Avian Dis. 55, 545–553.

419 Phenix, K. V, Weston, J.H., Ypelaar, I., Lavazza, A., Smyth, J.A., Todd, D., Wilcox, G.E., Raidal, S.R.,
420 2001. Nucleotide sequence analysis of a novel circovirus of canaries and its relationship to other
421 members of the genus Circovirus of the family Circoviridae. J. Gen. Virol. 82, 2805–2809.

422 Pushker, R., Mooney, C., Davey, N.E., Jacqué, J.-M., Shields, D.C., 2013. Marked variability in the
423 extent of protein disorder within and between viral families. PLoS One 8, e60724.
424 doi:10.1371/journal.pone.0060724

425 Rambaut, A., Drummond, A.J., 2007. Tracer: MCMC Trace Analysis Tool.

426 Rosario, K., Breitbart, M., Harrach, B., Segales, J., Delwart, E., Biagini, P., Varsani, A., 2017. Revisiting
427 the taxonomy of the family Circoviridae: establishment of the genus Cyclovirus and removal of the
428 genus Gyrovirus. Arch. Virol. doi:10.1007/s00705-017-3247-y

429 Rosario, K., Duffy, S., Breitbart, M., 2012. A field guide to eukaryotic circular single-stranded DNA
430 viruses: insights gained from metagenomics. Arch. Virol. 157, 1851–1871. doi:10.1007/s00705-012-
431 1391-y

432 Rosario, K., Schenck, R.O., Harbeitner, R.C., Lawler, S.N., Breitbart, M., 2015. Novel circular single-
433 stranded DNA viruses identified in marine invertebrates reveal high sequence diversity and
434 consistent predicted intrinsic disorder patterns within putative structural proteins. Front. Microbiol.
435 6, 1–13. doi:10.3389/fmicb.2015.00696

436 Sickmeier, M., Hamilton, J.A., LeGall, T., Vacic, V., Cortese, M.S., Tantos, A., Szabo, B., Tompa, P.,
437 Chen, J., Uversky, V.N., Obradovic, Z., Dunker, A.K., 2007. DisProt: the database of disordered
438 proteins. Nucleic Acids Res. 35, 786–793. doi:10.1093/nar/gkl893

439 Steel, O., Kraberger, S., Sikorski, A., Young, L.M., Ryan, J., Stevens, A.J., Ladley, J.J., Coray, D.S.,
440 Stainton, D., Dayaram, A., Julian, L., Bysterveldt, K. Van, Varsani, A., Bysterveldt, V., 2016.
441 Circular replication-associated protein encoding DNA viruses identified in the faecal matter of
442 various animals in New Zealand. Infect. Genet. Evol. 43, 151–164.
443 doi:10.1016/j.meegid.2016.05.008

444 Stenzel, T., Farkas, K., Varsani, A., 2015. Genome sequence of a diverse goose circovirus recovered from
445 Greylag goose. Genome Announc. 3, 7–8. doi:10.1128/genomeA.00767-15.

446 Stewart, M.E., Perry, R., Raidal, S.R., 2006. Identification of a novel circovirus in Australian ravens
447 (Corvus coronoides) with feather disease. Avian Pathol. 35, 86–92.
448 doi:10.1080/03079450600597345

449 Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular
450 evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum
451 parsimony methods. Mol. Biol. 28, 2731–2739. doi:10.1093/molbev/msr121

Todd, D., 2004. Avian circovirus diseases: lessons for the study of PMWS. Vet. Microbiol. 98, 169–174. doi:10.1016/j.vetmic.2003.10.010

Todd, D., 2000. Circoviruses: immunosuppressive threats to avian species: a review. Avian Pathol. 29, 373–94. doi:10.1080/030794500750047126

Trible, B.R., Rowland, R.R.R., 2012. Genetic variation of porcine circovirus type 2 (PCV2) and its relevance to vaccination, pathogenesis and diagnosis. Virus Res. 164, 68–77. doi:10.1016/j.virusres.2011.11.018

Wang, L., Brown, S.J., 2006. BindN: A web-based tool for efficient prediction of DNA and RNA binding sites in amino acid sequences. Nucleic Acids Res. 34, 243–248. doi:10.1093/nar/gkl298

Xue, B., Dunbrack, R.L., Williams, R.W., Dunker, K.A., Uversky, V.N., 2010. PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. Biochim. Biophys. Acta 1804, 996–1010. doi:10.1016/j.micinf.2011.07.011.

Xue, B., Dunker, A.K., Uversky, V.N., 2012. Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. J. Biomol. Struct. Dyn. 30, 137–149. doi:10.1080/07391102.2012.675145

Zuker, M., Jacobson, A.B., 1995. "Well-determined" regions in RNA secondary structure prediction: analysis of small subunit ribosomal RNA. Nucleic Acids Res. 23, 2791–2798.

Figure Captions

Figure 1: Agarose gels with amplified products: (A) *EcoRI*-RCA pattern obtained from four independent amplifications (1-4). The lack of equimolar amounts of DNA suggested at least two different viruses; (B) PCR amplification on RCA product using back-to-back primers CIR1000F1S/R1S (line 1), CIR3000F1S/R1S (line 3) and conventional primers CIR1000F2/R2) (line 2), CIR3000F2/R2 (line 4); (C) Direct PCR amplification on the lesion DNA and blood sample using back-to-back primers CIR1000F1S/R1S (lines 1 to 4) and CIR3000F1S/R1S (lines 5 to 8). Lines 1, 2, 5 and 6 (amplification of 1µl of DNA extraction of lesion), line 3 and 7 (amplification of 2 µl of blood sample), lines 4 and 8 are negative controls. M: GeneRuler 1 kb DNA ladder Plus (ThermoFisher Scientific).

Figure 2: Predicted genome organization of the two novel CRESS DNA viruses described in this study. (A) Circular structure of TaCV1 and TaCV2, lengths of the ORFs and location of the stem-loop structures on the origin of replication. (B) Detail of intergenic regions (IR) and conserved origin of replication stem-loop structure of TaCV1, ORF orientation and tandem repeats of 26 nt (black boxes). Non conserved bases of the repeats are marked by an asterisk. Shaded bases represent the conserved nonamer. Underlined sequences correspond to putative binding sites for Rep (C) Detailed stem-loop structure of TaCV2, ORF orientation and tandem repeats of 23 nt (black boxes). Shaded bases represent the conserved nonamer.

Figure 3: Colour-coded pairwise identity matrices of Rep (A) and complete genomes (B), showing the percentage of Rep amino acid and genome nucleotide identity of TaCV1 and TaCV2 with 159 CRESS DNA viruses.

Figure 4: Evolutionary relationships of Rep of the two genomes characterized in this study (TaCV1 and TaCV2; marked with arrows) inferred with Bayesian analyses. The colors at the end of the branches represent the host group or habitat from which the virus was isolated. Numbers along the branches represent branch support (posterior probabilities).

Figure 5: Representation of intrinsically disordered regions profiles of all non-Rep-encoding ORFs of both CRESS DNA viruses described in this study based on the IUPred software (A) and the ANCHOR software (B). Putative binding regions within amino acid sequences are shown. Values above the 0.5 threshold indicate disordered regions and binding regions.

Figure 1

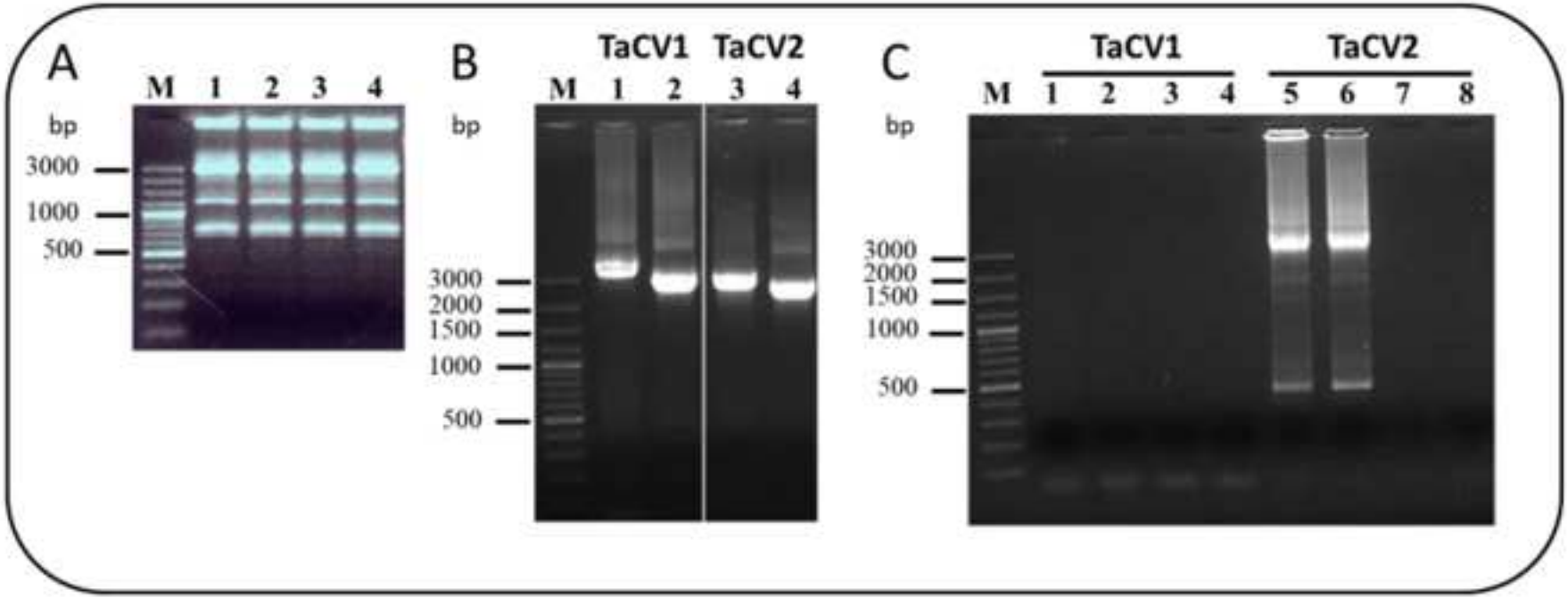

A
M 1 2 3 4
bp
3000
1000
500
B
TaCV1 TaCV2
M 1 2 3 4
bp
3000
2000
1500
1000
500
C
TaCV1
TaCV2
M 1 2 3 4 5 6 7 8
bp
3000
2000
1500
1000
500

Figure 2

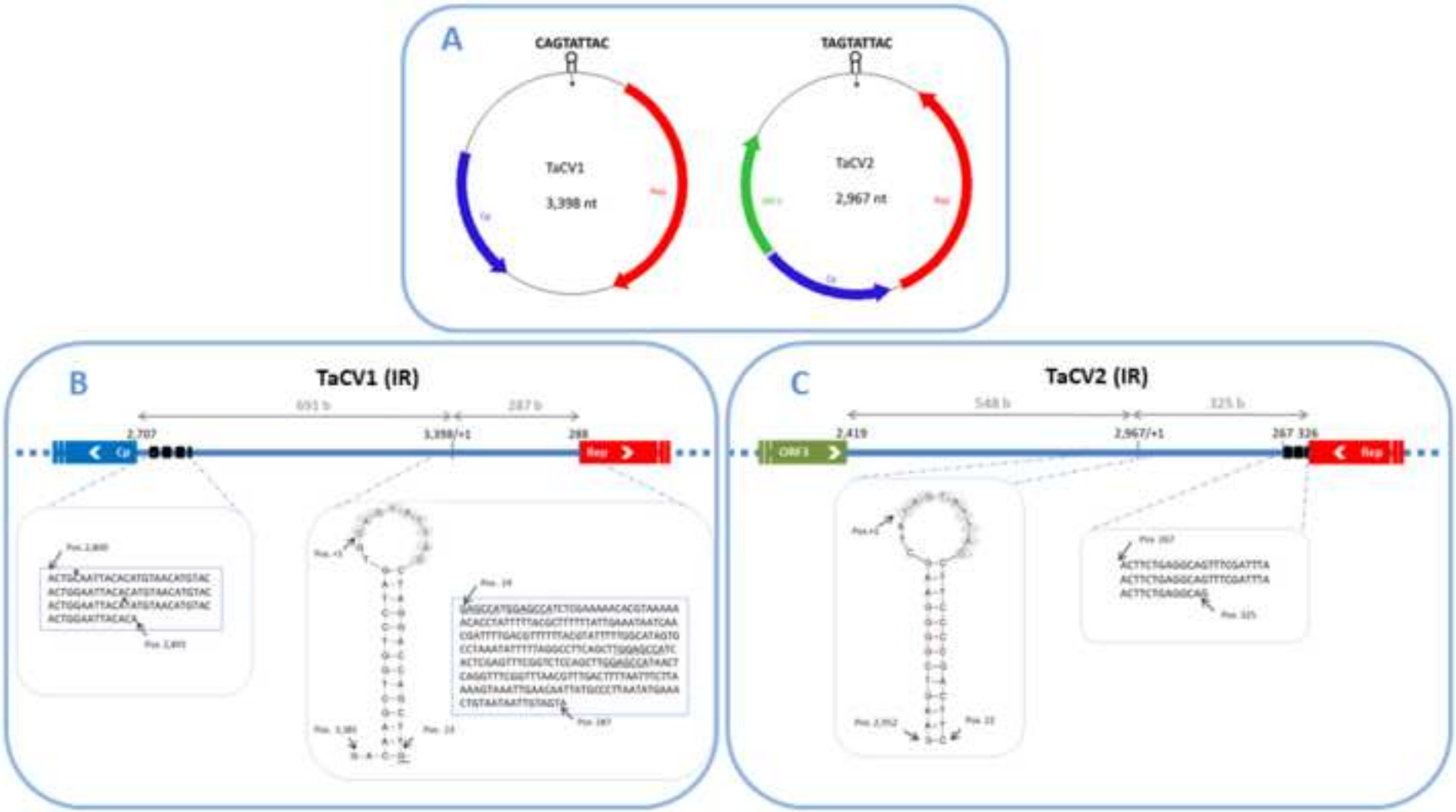

A
CAGTATTAC
TaCV1
3,398 nt
Cp
Rep
TAGTATTAC
TaCV2
2,967 nt
ORF1
Cp
Rep
B
TaCV1 (IR)
691 b
287 b
2,707
3,398/+1
288
Cp
Rep
C
TaCV2 (IR)
548 b
325 b
2,419
2,967/+1
267 326
ORF1
Rep

Figure 3

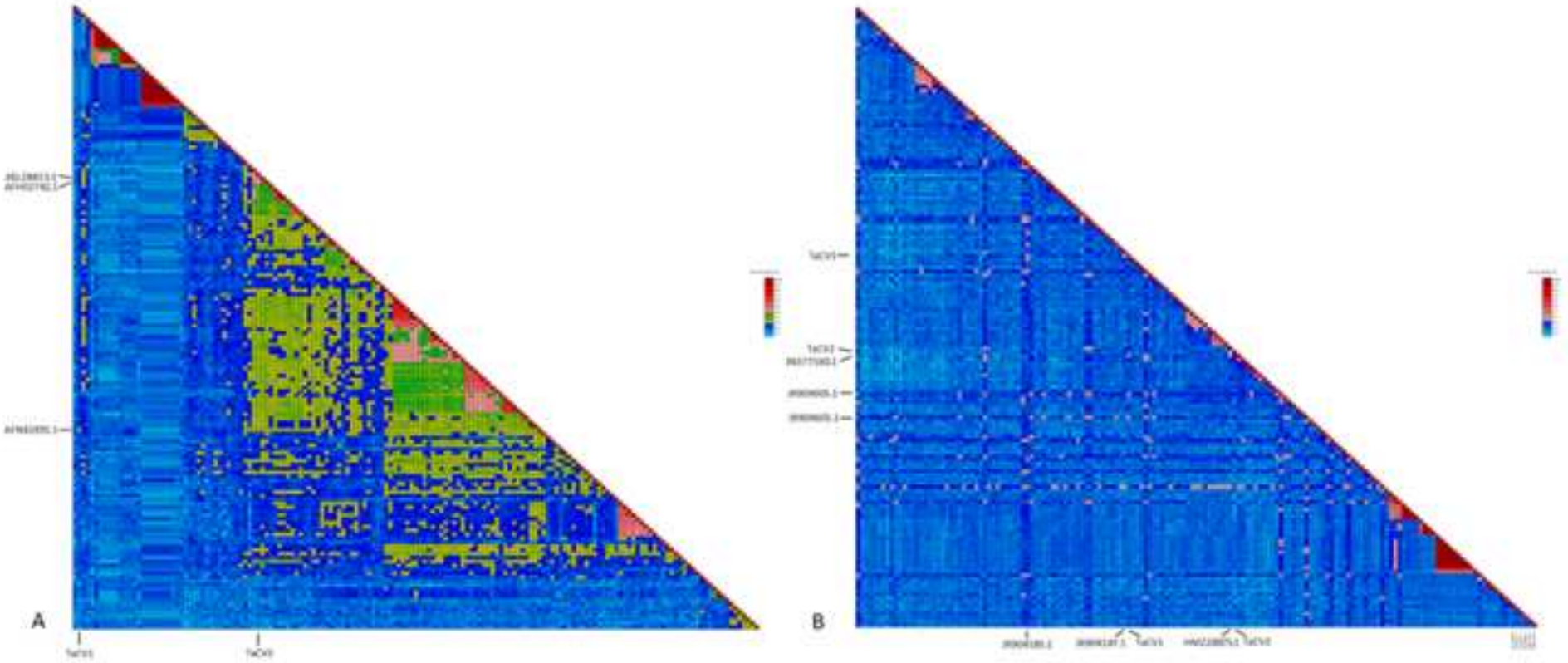


Figure 4

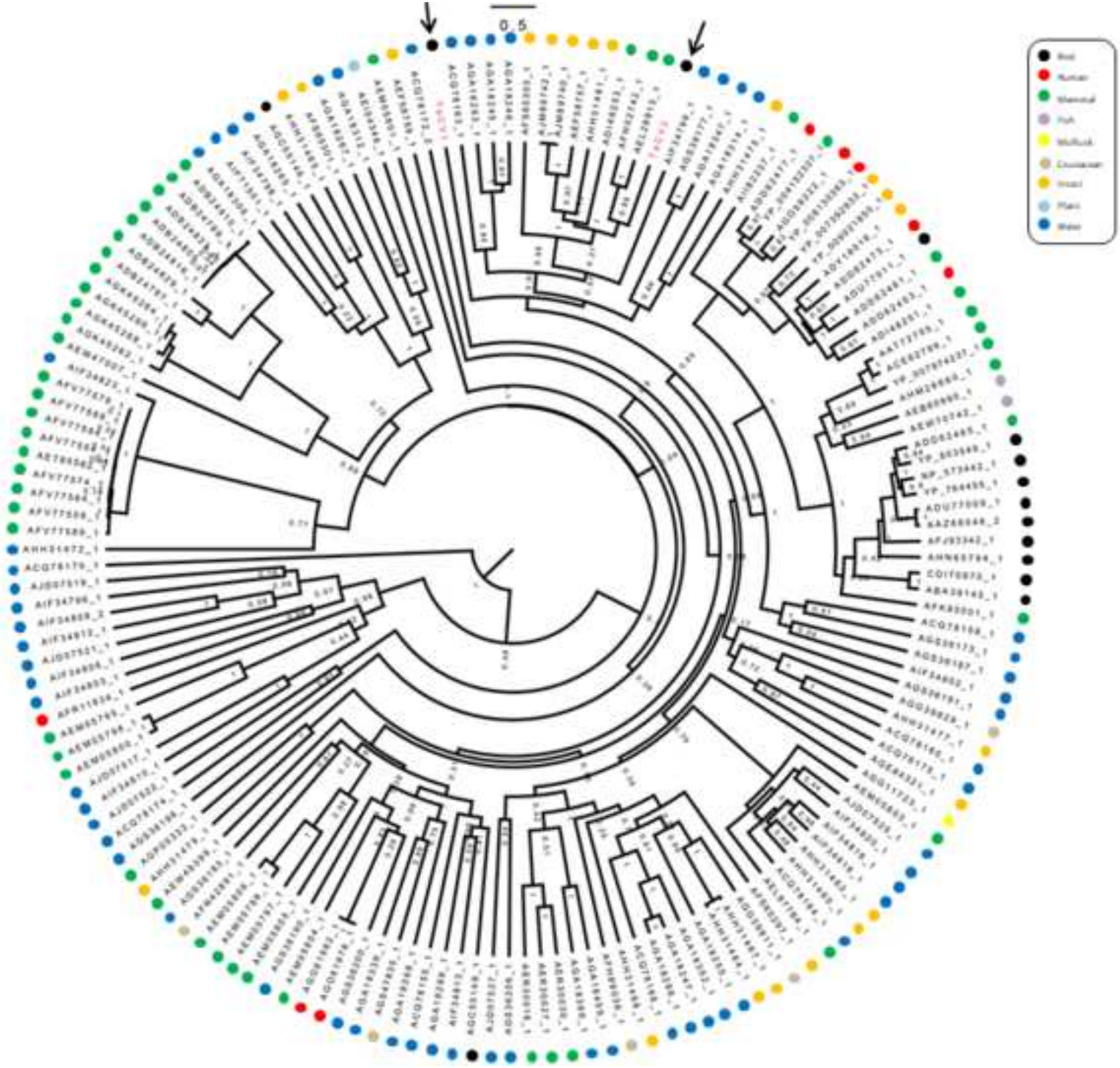
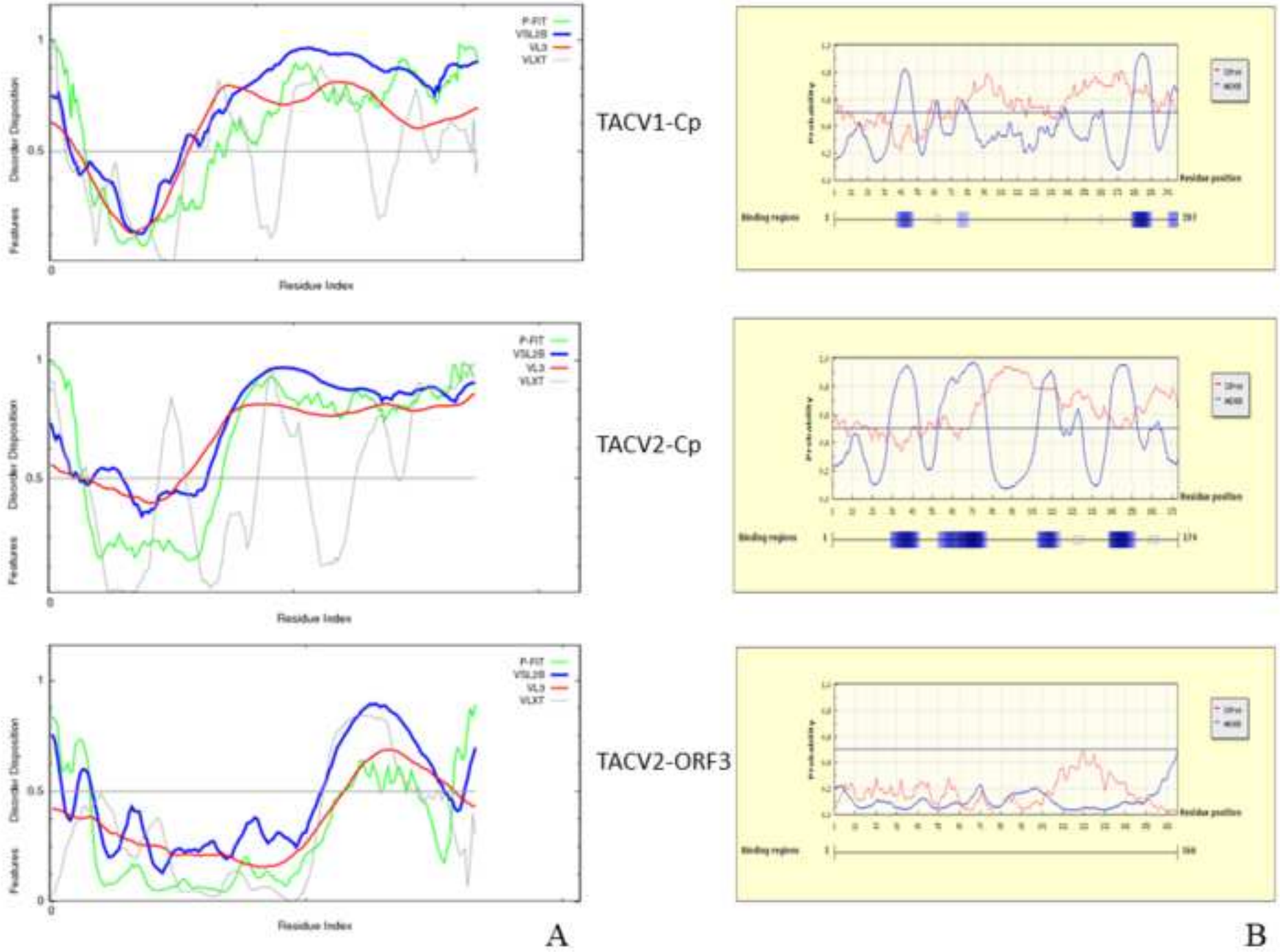


Figure 5



Appendices

Table A.1: List of primer sets used for PCR and walking sequencing

Virus	Primer	Type	Sequence	Used for
TaCV1	CIR1000F1S	PCR Back-to-back	5'-GAAAGAAACGTGATGACTCC-3'	RCA/Not amplified sample
TaCV1	CIR1000R1S	PCR Back-to-back	5'-TCGAATTTGGAGTCATCACG-3'	RCA/Not amplified sample
TaCV1	CIR1000F2	PCR Non-overlapping	5'-CAAGCGGTAGTGGTAAAAG-3'	RCA/Not amplified sample
TaCV1	CIR1000R2	PCR Non-overlapping	5'-CAAGCTGGAGACCGAAACT-3'	RCA/Not amplified sample
TaCV1	Mi500F1	Sequencing	5'-AGCGAGCATTCCCATTTTTCA-3'	Cloned sequences/RCA
TaCV1	Mi10R1	Sequencing	5'-GTTTTCTGGTCCAAGCGGTAGTG-3'	Cloned sequences/RCA
TaCV1	Mi500F2	Sequencing	5'-TAAAACGGAGTGTAAGTGGT-3'	Cloned sequences/RCA
TaCV1	Mi10R2	Sequencing	5'-AAGGCGAGTAATAACAGTAAGTCT-3'	Cloned sequences/RCA
TaCV1	Mi10F3	Sequencing	5'-CATTACCACTTACACTCCGTTTTA-3'	Cloned sequences
TaCV1	Mi500R3	Sequencing	5'-TGTTATTACTCGCCTTCTCAG-3'	Cloned sequences
TaCV1	Mi1000/walk/FW1	Sequencing	5'-AACTCCTTCGTGATATAAATGTAAT-3'	RCA
TaCV1	Mi1000/walk/RV1	Sequencing	5'-ACGAGGAAGGGTTTTGTAATAATA-3'	RCA
TaCV1	Mi1000/walk/FW2	Sequencing	5'-TTCTCTTTTAACTCCTTCGTGATA-3'	RCA
TaCV1	Mi1000/walk/RV2	Sequencing	5'-CGATATTTAACGTTGGTGAGG-3'	RCA
TaCV1	Mi1000/walk/FW3	Sequencing	5'-ATCATTTGGTTATATCATTTTA-3'	RCA
TaCV1	Mi1000/walk/RV3	Sequencing	5'-TCGATAGAAACAAGATACCC-3'	RCA
TaCV1	M13RP	Sequencing	5'-CAGGAAACAGCTATGACC-3'	Cloned sequences
TaCV1	M13FP	Sequencing	5'-TGTAAAACGACGGCCAGT-3'	Cloned sequences
TaCV2	CIR3000F1S	PCR Back-to-back	5'-TAATAGAAACAAAAATAAAGAACC-3'	RCA/Not amplified sample
TaCV2	CIR3000R1S	PCR Back-to-back	5'-TTGTTTCTATTATGTTGTTTATTG-3'	RCA/Not amplified sample
TaCV2	CIR3000F2	PCR Non-overlapping	5'-GCTTATGAAATCTTGCACTACTTG-3'	RCA/Not amplified sample
TaCV2	CIR3000R2	PCR Non-overlapping	5'-TATGTGAATGTAGAAAGTGAGTTGG-3'	RCA/Not amplified sample
TaCV2	Mi3000/walk/FW1	Sequencing	5'-TCAAAAATAGTGGAAAAATAGG-3'	RCA
TaCV2	Mi3000/walk/RV1	Sequencing	5'-CCTCTTCTTTCTTCCCAATGTA-3'	RCA
TaCV2	Mi3000/walk/FW2	Sequencing	5'-TCTAAAGCTTCATCACTCAATCTAT-3'	RCA
TaCV2	Mi3000/walk/RV2	Sequencing	5'-TAGTCCATGGCATATTCCTTTTTA-3'	RCA
TaCV2	Mi3000/walk/FW3	Sequencing	5'-TTGGAAAGATGACACTAAATGGT-3'	RCA
TaCV2	Mi3000/walk/RV3	Sequencing	5'-CTTTCTGTTATAGTAGGGATTTTC-3'	RCA