

Nuclear trafficking of *Anelloviridae* capsid protein ORF1 reflects modular evolution of subcellular targeting signals

Gayle F. Petersen^{1,†}, Silvia Pavan^{2,†}, Daryl Ariawan³, Ole Tietz⁴, Sepehr Nematollahzadeh², Subir Sarker⁴, Jade K. Forwood^{1,5}, Gualtiero Alvisi^{1,2,*}

¹Gulbali Institute, Charles Sturt University, 250 Boorooma Street, Wagga Wagga, NSW 2678, Australia

²Department of Molecular Medicine, University of Padova, Via Gabelli 63, Padova (PD) 35121, Italy

³Dementia Research Centre, Macquarie Medical School, Faculty of Medicine, Health and Human Sciences, 75 Talavera Rd, Macquarie University, Sydney, NSW 2109, Australia

⁴Biomedical Sciences & Molecular Biology, College of Medicine and Dentistry, James Cook University, 1 James Cook drive, Townsville, QLD 4811, Australia

⁵School of Agricultural, Environmental and Veterinary Sciences, Charles Sturt University, 250 Boorooma Street, Wagga Wagga, NSW 2678, Australia

*Corresponding author. Department of Molecular Medicine, University of Padova, Padova 35121, Italy. E-mail: gualtiero.alvisi@unipd.it

[†]Gayle F. Petersen and Silvia Pavan equal contribution

Abstract

Anelloviridae members are ubiquitous viruses with a small, negative sense, single-stranded DNA genome which is replicated by host cell DNA polymerases. Anelloviruses are postulated to interact with the host cell nuclear transport machinery, however, the lack of reliable cell culture models strongly limits our knowledge regarding *Anelloviridae*-host interactions. In particular, capsid nuclear import is a largely uncharacterized process. We addressed this by investigating the relationship between host cell nuclear transport receptors (NTRs) and ORF1, the putative capsid protein from torque teno douroucouli virus (TTDoV). We identified the subcellular targeting signals and NTRs responsible for its nucleolar and nuclear localization, and characterized their relative contribution to ORF1 subcellular localization. In the absence of other viral proteins, ORF1 accumulated in the nucleoli. Bioinformatics analysis revealed a putative classical nuclear localization signal (cNLS) within the highly conserved N-terminal arginine rich motif (ARM) ('NLSn', 27-RRWRRRPRRRRPPYRRPYRRYGRRRKVRRR-57), and an additional C-terminal cNLS ('NLSc', 632-LPPPEKRARWGF-643), which has been specifically acquired by *Anelloviridae* capsids with larger projection domains. Such NLSs play distinct roles in ORF1 subcellular localization by interacting with specific NTRs. NLSn, a non-classical NLS, features broad importin (IMP) binding affinity yet plays a minor role in nuclear import, being responsible for nucleolar targeting likely through interaction with nucleolar components. NLSc, a *bona fide* cNLS, specifically interacts with IMP α and is the main driver of active nuclear transport in an IMP α /β1-dependent fashion. These findings suggest an evolutionary correlation between the acquisition of progressively larger projection domains and the presence of additional cNLSs in *Anelloviridae* capsids, aimed at maximizing IMP α /β1-mediated nuclear import.

Keywords: *Anelloviridae*; karyopherins; shuttling; nucleolus; capsid; NLS; NoLS; projection domain

Introduction

Anelloviridae are small, non-enveloped viruses with circular, single-stranded negative sense DNA genomes, first discovered in 1997 when torque teno virus (TTV) was identified in the blood of a post-transfusion patient with elevated liver enzymes using molecular techniques (Nishizawa et al. 1997). Subsequent advances in metagenomics have led to the identification of a plethora of additional TTVs infecting humans and animals, with high genomic heterogeneity (Nishizawa et al. 1997, Biagini 2009). These viruses are now classified under the *Anelloviridae* family, which comprises 31 established genera with genomes between 1.6 and 3.9 kb in length (Varsani et al. 2021). Larger genomes, such as those from the genera *Alphatorquevirus* and *Zetatorquevirus* ranging from 3.6 to 3.9 kb, usually encode for larger capsid proteins compared to smaller genomes, such as those from the genus *Gyrovirus* ranging from 1.8 to 2.4 kb (Butkovic et al. 2023). All these

isolates feature a shared genetic organization, with a relatively conserved untranslated region containing a GC-rich zone, and a coding region containing the major open reading frames (ORFs), ORF1, which is proposed to encode for the capsid protein (Butkovic et al. 2023), and an overlapping ORF2, which is believed to be a regulatory protein that interferes with host antiviral defences by suppressing the NF- κ B pathway (Zheng et al. 2007). Depending on the isolate, several additional ORFs have also been described (Deb et al. 2021).

TTVs are believed to be the most abundant eukaryotic viruses in the human virome (Virgin et al. 2009, Deb et al. 2021), chronically infecting most of the human population, and have been detected in up to 90% of tested individuals in the absence of clinical symptoms (Vasilyev et al. 2009). Though no disease has been unequivocally linked to TTV infection, viremia has been shown to reflect the degree of immunosuppression and is

being clinically investigated as a marker of solid organ transplant rejection (van Rijn et al. 2023). Asymptomatic anellovirus infections are also common in other mammals (Fahsbender et al. 2017), as well as in chickens and several other avian species, suggesting long term virus-host co-evolution. Given their ability to establish lifelong infection, it is proposed that TTVs have evolved sophisticated mechanisms to manipulate host cell functions and immune defences (Vietzen et al. 2024). Recently, the immune evasion properties of *Anelloviridae* members have been linked to the evolutionary acquisition of a highly variable projection domain between β -strands H and I of the jelly-roll fold within the capsid protein, which appear considerably variable in length and larger than their *Circoviridae* counterparts (Sarker et al. 2016, Butkovic et al. 2023). Consistent with this hypothesis, cryogenic electron microscopy (cryo-EM) analysis of virus-like particles (VLPs) of ORF1 from *Betatorquevirus* isolate LY1 revealed that the variable projection domain protrudes outside the capsid structure at the 5-fold symmetry axis, potentially exposed for recognition by neutralizing antibodies (Liou et al. 2024).

With a circular ssDNA genome of approximately 3000 bases, TTV DNA replication has long been thought to occur through a rolling circle mechanism in the nucleus of infected cells (Okamoto et al. 2000a, 2000b). Therefore, TTV proteins are expected to interact with host nuclear transport receptors (NTRs), specifically importin (IMP) superfamily members, which recognize nuclear localization signals (NLSs) on cargoes to mediate their nuclear import (Gorlich et al. 1995, Timney et al. 2016). NLSs can be categorized into several classes, based on their ability to interact with specific NTRs. The best characterized are the 'classical' NLSs (cNLSs), which are recognized by IMP α paralogs and subsequently transported into the nucleus by the IMP α / β 1 heterodimer (Marfori et al. 2011). Other NLS classes are capable of binding to different NTRs and are generally referred to as 'non-classical' NLSs (Bourgeois et al. 2020). While molecules smaller than ~ 70 kDa can passively diffuse through the nuclear pore complex (Timney et al. 2016), larger ones and those that need to quickly accumulate in the cell nucleus require energy dependent transport (Wing et al. 2022). Further, several proteins from both DNA and RNA viruses need to be actively translocated into the nucleus to foster viral replication and manipulate cell function (Alvisi et al. 2013, Bonamassa et al. 2015). TTV capsid proteins feature a highly basic arginine rich motif (ARM) at the N-terminus, which is conserved in capsids from closely related *Circoviridae* members and other highly divergent icosahedral viruses (Requiao et al. 2020). Stretches of basic amino acids within ARMs can bind to NTRs and have been proposed to act as NLSs (Patterson et al. 2013), however they are generally located inside the viral capsid (Venter et al. 2009, Sarker et al. 2016, Liou et al. 2024), packaging viral genomes of both RNA (Venter et al. 2009) and DNA (Sarker et al. 2016) viruses by electrostatic interaction (Requiao et al. 2020). Analysis of TTV-host interactions have been so far limited by a lack of suitable cellular systems allowing viral replication (Kaczorowska and van der Hoek 2020). Heterogenous subcellular localization has been reported for ORF1 from different TTV species, but the NTRs responsible for nuclear targeting have not yet been identified. ORF1 from TTV genotype 6 HEL32 is mainly restricted to the cytoplasm when expressed as HA- or GFP-tagged fusions in the absence of any other viral protein (Qiu et al. 2005), while in the case of genogroup 1 P/1C1 and suid TTV isolates TTSuV1 type 1C and TTSuV2 subtype 2A, ORF1 was primarily detected in the nucleoli (Mueller et al. 2008, Martinez-Guino et al. 2011).

We set out to investigate the interaction between ORF1 and the host cell nuclear transport machinery using a prototype TTV, the *Zetatorquevirus torque teno douroucouli virus* (TTDoV), which was first identified in the non-human primate *Aotus trivirgatus* (Okamoto et al. 2000b). Our study identified specific sequences that directly interact with NTRs and control the subcellular localization of TTDoV ORF1. ORF1 contains two putative NLSs. The N-terminal NLS (NLSn), which largely overlaps with the ARM, is a non-classical NLS that interacts with a wide range of IMP α and IMP β NTRs. This signal is crucial for nucleolar accumulation but poorly contributes to nuclear transport. By contrast, an additional NLS at the C-terminus (NLSc) selectively binds IMP α 's and is the main driver of IMP α / β 1-mediated nuclear localization, representing a *bona fide* cNLS. Intriguingly, while NLSn is widely conserved across *Anelloviridae* and *Circoviridae* members, NLSc is exclusively found within larger capsids from anelloviruses, correlating with large projection domains and establishing an evolutionary link between acquisition of large projection domains and presence of additional IMP α / β 1-dependent cNLSs at the C-terminus of capsid proteins.

Materials and methods

Bioinformatics

The genomic sequence of TTDoV (isolate At-TTV3) was retrieved from GenBank (accession number: 11862897). The sequence of viral encoded ORF1 was retrieved from UniProt with the code Q9DUB7. The sequences of *Circoviridae* (Varsani et al. 2024) and *Anelloviridae* (Butkovic et al. 2023) ORF1 proteins were retrieved from UniProt. Protein sequences were analysed with cNLS Mapper (Kosugi et al. 2009) to identify putative cNLSs. Structural models were predicted using the AlphaFold3 Server (Abramson et al. 2024) and visualized using the PyMOL Molecular Graphics System (version 3.1.3.1; Schrödinger, LLC). Protein model interactions were analysed using the PDBePISA (<https://www.ebi.ac.uk/pdbe/pisa/>) and PDBsum (<https://www.ebi.ac.uk/thornton-srv/databases/pdbsum/>) web tools. Phylogenetic analysis was performed using Clustal Omega (Madeira et al. 2024) and standard settings.

Plasmids

Plasmids encoding NTRs used in binding assays include human importin alpha 1 (hIMP α 1 Δ IBB), mouse importin alpha 2 (mIMP α 2 Δ IBB), human importin alpha 3 (hIMP α 3 Δ IBB), human importin alpha 5 (hIMP α 5 Δ IBB), and human importin alpha 7 (hIMP α 7 Δ IBB), all truncated to remove the autoinhibitory importin beta binding (IBB) domain, and human importin beta 1 (hIMP β 1), human importin beta 2 (hIMP β 2), and human importin beta 3 (hIMP β 3). Genes were cloned into pET-30a(+) or pMCSG21 vectors. All plasmids contain an N-terminal 6x histidine tag and a tobacco etch virus (TEV) protease cleavage site, except mIMP α 2 Δ IBB (no TEV site). Plasmids pcDNA3.1/NT-GFP and pcDNA3.1/NT-GFP-SV40 LTA, mediating the expression of cycle 3 GFP or cycle 3 GFP fused to SV40 large tumour antigen (LTA) NLS (126-PKKKRKV-132), were described previously (Alvisi et al. 2023). Mammalian expression plasmids encoding TTDoV ORF1 NLSs fused to the C-terminus of cycle 3 GFP were generated by annealing appropriate oligonucleotide pairs in vector pcDNA3.1/NT-GFP-TOPO[®] (Thermo Fisher Scientific, Monza, Italy). Plasmid mCherry-Bimax2, encoding a fusion protein between the spontaneously fluorescent protein mCherry and the competitive inhibitor of the IMP α / β 1 nuclear import pathway Bimax2 (Tsujii et al. 2015), was kindly gifted from Yoshihiro Yoneda and Masahiro Oka (Osaka,

Japan), while plasmids pDsRed-C1-fibrillarin and pDsRed-C1-nucleolin (Gomez Corredor and Archambault 2009) were kindly provided by Denis Archambault (University of Québec, Canada). Plasmid pEGFP-C1-ORF1, with TTDov ORF1 placed downstream of the eGFP ORF and flanked by Gateway attB recombination sites, was synthesized (BioFab Research, Rome, Italy) and substitution derivatives were generated using the Quikchange mutagenesis kit (Agilent Technologies, Cernusco sul Naviglio (MI), Italy) with appropriate oligonucleotide pairs, as previously described (Alvisi et al. 2009). A list of all plasmids used in this study is available in [Supplementary Table S1](#).

Peptides

N-terminal fluorescein isothiocyanate (FITC)-tagged synthetic peptides ORF1 NLSn, ORF1 NLS_c, and the positive control SV40 LTA NLS were synthesized as described previously (Alvisi et al. 2023, Cross et al. 2024a), using standard Fluorenyl methoxycarbonyl (Fmoc)-solid-phase peptide synthesis on low swell 100–200 mesh Wang resin (0.05 mmol reaction scale, 0.5 mmol g⁻¹ loading) on a CEM Liberty Blue™ Peptide Synthesizer (CEM, USA). *Initial amino acid loading*: Wang resin (100–200 mesh; 0.65 mmol g⁻¹, 77 mg, 0.05 mmol) was weighed into a 10 mL polypropylene syringe equipped with a porous polypropylene frit, which was used as the reaction vessel. The resin was washed with dichloromethane (3 × 5 mL) before being allowed to swell in dichloromethane (5 mL) for at least 0.5 h prior to loading of the first amino acid. A solution of Fmoc-AA-OH (4 equiv.) was dissolved in a mixture of dry dichloromethane (2 mL), N,N-dimethylformamide (2 mL), hydroxybenzotriazole (HOBt) (4 equiv.), and N,N'-diisopropylcarbodiimide (DIC) (4 equiv.), taken up into the syringe with resin, and stirred overnight using an orbital shaker. The resin was then capped with acetic anhydride (0.1 mL) and N,N-diisopropylethylamine (DIPEA) (0.1 mL) in dichloromethane (3 mL) for 30 min. The resin was then washed with dichloromethane (3 × 4 mL) and N,N-dimethylformamide (DMF) (3 × 4 mL). *Automated peptide synthesizer*: Resin was pre-swelled in 50/50 DMF and dichloromethane (DCM) for 1 hr. Amino acids were dissolved in DMF at a concentration of 0.2 M before being transferred to the synthesizer. Peptides were synthesized using sequential amid coupling from C- to N-terminus for 5 min at 90 °C, using five equivalents of amino acid with 10 equivalents of activator (0.5 M DIC (N,N'-Diisopropylcarbodiimide) in DMF) and 5 equivalents of activator base (0.5 M Oxyma (Ethyl cyanohydroxyiminoacetate), 0.05 M DIPEA (N,N-Diisopropylethylamine) in DMF), followed by Fmoc deprotection in 20% piperidine in DMF for 3 min at 75 °C and 3x resin wash in DMF. Following final Fmoc deprotection, resin was removed from the synthesizer, transferred to a syringe fitted with a propylene filter, and labelled with FITC (*see below*). Double couplings were performed for arginine residues to ensure complete coupling. *FITC labelling*: Fmoc-6-aminohexanoic acid (Ahx) was coupled to the N-terminus of peptides and deprotected using standard amino acid coupling conditions. Peptide bound resin was removed from the synthesizer and transferred to a syringe fitted with a propylene filter. FITC (2 equiv.), HOBt (3 equiv.), benzotriazol-1-yloxytripyrrolidinophosphonium hexafluorophosphate (PyBOP) (3 equiv.), and DIPEA (6 equiv.) in DMF (4 mL) was taken up in the syringe and agitated overnight in an orbital shaker. Resin was then washed with DMF (x3), DCM (x3), and methanol (x3) and proceeded to cleavage. *Cleavage*: Peptide was cleaved from the resin using a cleavage cocktail of 92.5% TFA (trifluoroacetic acid), 2.5% TIPS (triisopropylsilane), 2.5% thioanisole, and 2.5% H₂O for at least 3 hrs at room temperature, precipitated in ice cold diethyl

ether, dissolved in H₂O, and freeze dried. Peptide ORF1 NLS_c was purified using a Shimadzu LC-20 AD high-performance liquid chromatography (HPLC, Shimadzu, Japan). Mass spectra were obtained on a Shimadzu LCMS-8050 LCMS system (Shimadzu, Japan) in positive electron spray [ESI+] mode, fitted with a Polaris 3 C18-A 50 x 4.6 mm column (Agilent Technologies, USA). Peptide ORF1 NLSn could not be purified due to the high arginine content. FITC-tagged synthetic peptides ORF1 NLSn_A, ORF1 NLSn_B, and the negative control Langya virus matrix protein were synthesized by GenScript (Singapore). A list of all peptides used in this study is available in [Supplementary Table S2](#).

Expression and purification of recombinant proteins

Plasmids encoding NTRs were transformed into BL21(DE3)pLysS E. coli cells and proteins were expressed for 24–30 hrs at 25 °C using the auto-induction method (Studier 2005). Bacterial cells were pelleted via centrifugation at ~7 500 x g at 4 °C and resuspended in His buffer (50 mM phosphate buffer, 300 mM sodium chloride, 20 mM imidazole, pH 8). Cells were lysed by three freeze-thaw cycles, followed by incubation with 20 mg/mL lysozyme (Thermo Fisher Scientific, Waltham, MA, USA) and 50 mg/mL DNase (Sigma-Aldrich, St. Louis, MO, USA). Soluble extract was isolated by centrifugation at 30 000 x g at 4 °C, clarified by 0.45 µm filtration, and injected onto a pre-equilibrated 5 mL HisTrap HP column (Cytiva, Marlborough, MA, USA). The column was washed with 20 column volumes of His buffer, and protein was eluted using a linear gradient of 20–500 mM imidazole. Peak fractions were pooled and the 6x histidine tag was cleaved by incubation with TEV protease at 4 °C overnight (except mIMPα2ΔIBB; no TEV site). Proteins were further purified by size exclusion chromatography (SEC) using a pre-equilibrated HiLoad 26/600 Superdex 75 pg or HiLoad 26/600 Superdex 200 pg column (Cytiva, Marlborough, MA, USA) and SEC buffer (50 mM Tris base, 125 mM sodium chloride, pH 8). Peak fractions were pooled and run through a 5 mL HisTrap HP column, pre-equilibrated with SEC buffer, to remove uncleaved target protein and TEV protease. Flowthrough fractions were pooled and concentrated using an Amicon 10 kDa MWCO ultra centrifugal filter (Merck Millipore, Burlington, MA, USA). Purified NTR protein was flash frozen in liquid nitrogen, aliquoted, and stored at –80 °C for future use. The negative control protein human SOX2 high mobility group (HMG)-box domain (SOX2; UniProt: P48431; residues 39–127) was expressed and purified as previously described (Ghafoori et al. 2024).

Electrophoretic mobility shift assays (EMSAs)

To qualitatively assess the interaction between NLSs and NTRs, EMSAs were performed. Twenty µM NTR protein was combined with 10 µM FITC-tagged NLS peptide, in the presence of 7.5% glycerol, and electrophoretically separated on a 1.5% agarose gel in TB buffer (45 mM Tris base, 45 mM boric acid) at 75 V for 1.5–2 hrs. Gels were imaged for FITC peptide detection before staining with Coomassie Blue for protein detection. Protein only and peptide only controls were also run.

Fluorescence polarization (FP) assays

To quantitatively assess the interaction between NLSs and NTRs, FP assays were performed based on previously described methods (Cross et al. 2024b). Twenty µM NTR protein was titrated in a two-fold dilution series across 23 wells of a black Fluotrac microplate (Greiner Bio-One, Kremsmünster, Austria) and combined with 10 nM FITC-tagged NLS peptide. Wells were made up to a total volume of 200 µL with SEC buffer. Fluorescence polarization was

measured using a CLARIOstar Plus plate reader (BMG Labtech, Ortenberg, Germany). A peptide only control was included and used for gain adjustment. Assays were performed in triplicate. Data were analysed in GraphPad Prism (version 10.2.2; GraphPad Software, Boston, MA, USA) using non-linear regression with one site-specific binding to determine the dissociation constant (Kd) and maximum binding (Bmax) (Bmax constrained to > 50).

Cell culture

HEK293A cells (#R70507; Thermo Fisher Scientific, Monza, Italy) were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% (v/v) fetal bovine serum, 50 U/mL penicillin, 50 U/mL streptomycin, and 2 mM L-glutamine (Thermo Fisher Scientific, Monza, Italy), and passaged when confluent (Alvisi et al. 2018).

Confocal laser scanning microscopy (CLSM)

HEK293A cells were seeded onto glass coverslips in a 24-well plate (4×10^4 cells/well) and the next day transfected with appropriate amounts of expression constructs (range 100–250 ng) using Lipofectamine 2000 (Thermo Fisher Scientific, Monza, Italy), as previously described (Trevisan et al. 2018). At 24 hrs post-transfection (p.t.), cells were incubated for 30 min with DRAQ5 (1:5000 in DMEM, no phenol red), to stain cell nuclei. Cells were subsequently washed with PHEM 1x (60 mM PIPES, 25 mM HEPES, 10 mM EGTA, 4 mM MgSO₄), and fixed with 4% (v/v) paraformaldehyde for 10 min. Following three washes with PHEM 1x and one wash with milliQ water, coverslips were mounted on glass slides with Fluoromount G (Southern Biotech, Birmingham, AL, USA). Subcellular localization of fusion proteins was analysed using a Nikon A1 confocal laser scanning microscope (Nikon, Tokyo, Japan) equipped with a 60x oil immersion objective, as described previously (Smith et al. 2018). GFP (excitation peak: 488 nm; emission peak: 510 nm) and cycle 3 GFP (excitation peaks: 395 and 475 nm; emission peak: 507 nm) fusion proteins were excited with an argon ion laser beam at 488 nm (Melles Griot; IMA101040ALS). Fusion proteins with mCherry (excitation peak: 585 nm; emission peak: 610 nm) and DsRed (excitation peak: 560 nm; emission peak: 583 nm) were excited at 561 nm with a Shappire 561 Optically-Pumped Semiconductor Laser (Coherent; SHAPPIRE 561-20 CDHR). DNA-bound DRAQ5 (excitation peaks: 598 and 643 nm; emission peak: 696 nm) was excited at 642 nm with a Laser Diode (Oxxius; LBX-642-140-CSB-PPA).

Image analysis

DRAQ5 and DsRed-fibrillarin/nucleolin were used to define nuclear and nucleolar masks, respectively, whereas a small area close to DRAQ5 was used to define a cytosolic mask, as described previously (Nematollahzadeh et al. 2024). The fluorescence attributed to autofluorescence/background (Fb) was subtracted from the measurements to calculate the Fn/c and Fno/n ratios according to the formulas $Fn/c = (Fn - Fb)/(Fc - Fb)$ and $Fno/n = (Fno - Fb)/(Fn - Fb)$. Cells with oversaturated signals were excluded from analysis. In some cases, to allow easier detection of nucleoli, cells were co-transfected with DsRed-fibrillarin or DsRed-nucleolin expression plasmids and rgb profile plots were calculated with Fiji (Schindelin et al. 2012). Statistical analysis was performed using GraphPad Prism (version 9; GraphPad Software, Boston, MA, USA) applying Student's t test, one-way ANOVA, or two-way ANOVA as appropriate.

Inhibition of ran-dependent nuclear transport

Ran-dependent nuclear transport was inhibited by depletion of cellular RanGTP resulting from a lack of free GTP (Schwoebel et al. 2002), by incubating cells for 30 min at 37 °C in DMEM containing no glucose, 5% FBS, and supplemented with 10 mmol/L sodium azide, 6 mmol/L 2-deoxy-D-glucose (#D8357; Sigma, Merck Millipore, Milan, Italy) and DRAQ5 (#62251; Thermo Fisher Scientific, Monza, Italy; 1:5000), as described previously (Alvisi et al. 2008, Nematollahzadeh et al. 2024), before being stained, fixed, and analysed by CLSM as detailed above.

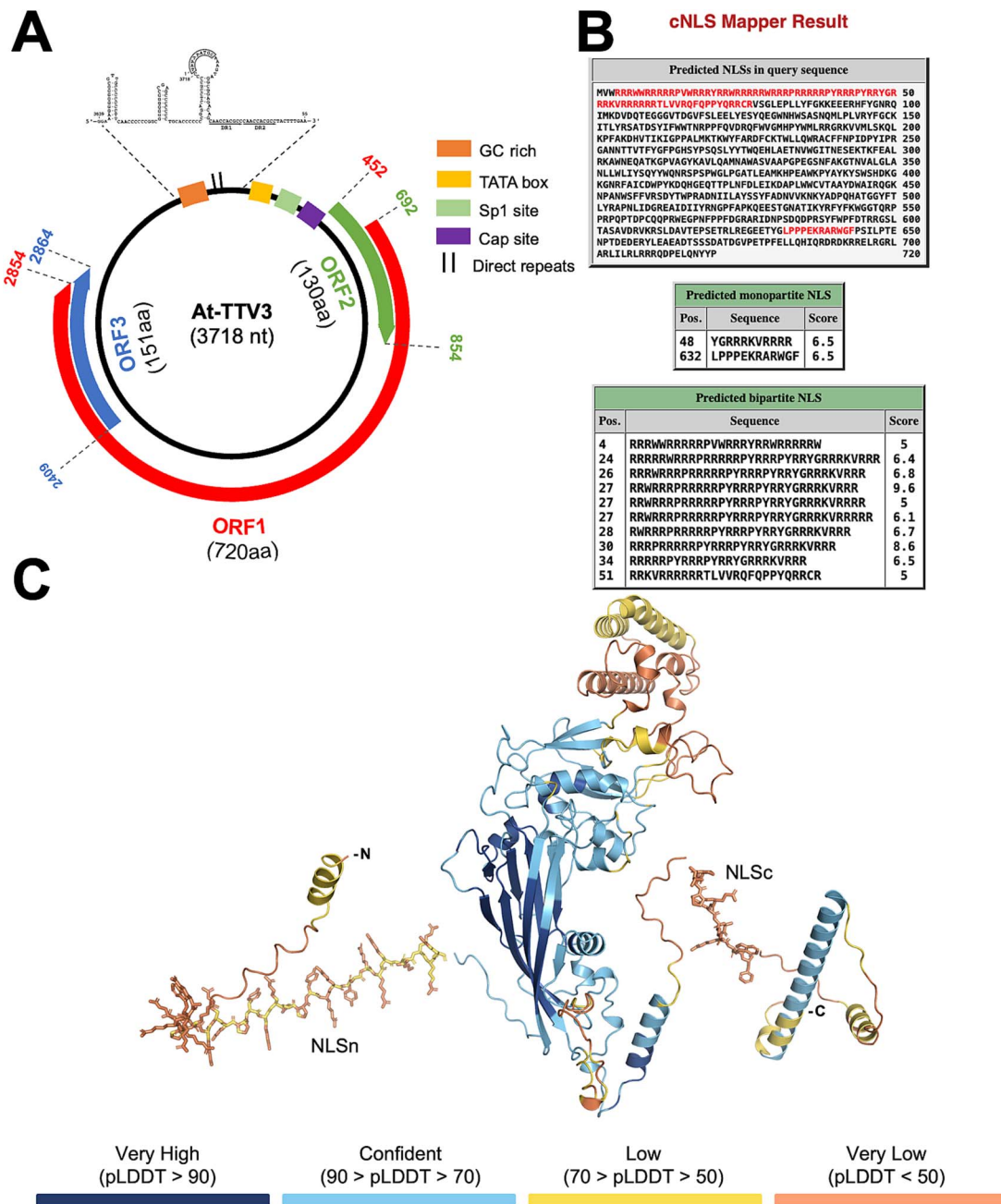
Results

TTDoV ORF1 contains putative cNLSs located at the N-terminus and the C-terminus

Although TTVs are believed to replicate in the cell nucleus, the ability of TTV encoded proteins to interact with host NTRs has never been investigated. Further, previous studies have reported conflicting results concerning the subcellular localization of ORF1 from different TTV isolates (Qiu et al. 2005, Mueller et al. 2008, Martinez-Guino et al. 2011). To shed some light on this topic, we analysed the primary amino acid sequence of TTDoV (isolate At-TTV3; Fig. 1A) ORF1 with the cNLS Mapper software, searching for putative cNLSs. Our analysis revealed that ORF1 possessed several putative cNLSs (Fig. 1B). In particular, eleven partially overlapping cNLSs were predicted in the N-terminal ARM spanning residues 4 to 75 (the strongest being 27-RRWRRRPRRRRPYRRRPYRRYGRRKVRRR-57; NLSn), and one cNLS was predicted at the C-terminus (632-LPPEKRARWGF-643; NLSc). To assess whether the identified cNLSs could be accessible to host cell NTRs, AlphaFold3 was used to predict the structure of full length TTDoV ORF1 protein and putative cNLSs were mapped onto the predicted model. The model of ORF1 (Fig. 1C; Supplementary Table S3) featured largely unstructured N- and C-termini, which is where both putative cNLSs were located. Between these regions (residues 71-546) was a single jelly-roll fold comprised of eight β -strands (B-I) forming two anti-parallel β -sheets, the larger BIDG and the smaller CHEF, with a projection domain extending out between β -strands H and I (Supplementary Fig. S1), as described for ORF1 proteins from Betatorquevirus isolate LY1 and beak and feather disease virus (BFDV), and as predicted for several additional *Anelloviridae* ORF1 proteins (Sarker et al. 2016, Butkovic et al. 2023, Liou et al. 2024). Both putative cNLSs were modelled in unstructured regions of the protein with low confidence scores (pLDDT < 70); this strongly correlates with regions that are intrinsically disordered (Kovalevskiy et al. 2024), thus NLSn and NLSc are proposed to be accessible for binding to NTRs in ORF1 monomers.

Basic residues within TTDoV ORF1 protein mediate high affinity interactions with several NTRs

We set out to functionally validate the newly identified TTDoV ORF1 putative cNLSs by first testing their ability to interact with selected NTRs, including IMP α 1/2/3/5/7 and IMP β 1/2/3. Fluorescein isothiocyanate (FITC)-tagged peptides corresponding to the predicted cNLS motifs were synthesized and subjected to EMSAs. The NLSn peptide consistently precipitated, suggesting poor solubility under assay conditions, thus we were unable to evaluate its binding to NTRs (Fig. 2A). To investigate further, NLSn was divided into two sections based on cNLS Mapper results,



namely the arginine rich region (27-RRWRRRRPRRRRRPYRRRPYRR-47; ORF1 NLSn_A) and the predicted monopartite cNLS (48-YGRRRKVRRR-57; ORF1 NLSn_B). Both NLSn_A and NLSn_B co-migrated with several IMP α and IMP β NTRs, indicating their potential as NLSs, although some precipitation of the arginine rich NLSn_A peptide was still evident (Fig. 2B). Conversely, NLS_c specifically co-migrated with only IMP α NTRs (Fig. 2A). FP assays were subsequently performed to measure the binding affinity of ORF1 NLS peptides for NTRs and confirm EMSA results. Again,

no interaction could be detected between NLSn and NTRs due to poor peptide solubility (Supplementary Fig. S2A,B). NLSn_A bound all NTRs with high affinity (~ 1 -47 nM), exhibiting a preference for IMP β 's (all < 2 nM) (Fig. 2C,D), while ORF1 NLSn_B bound all NTRs with a moderate to low binding affinity (~ 244 -1604 nM), exhibiting a preference for IMP α 1 (290 nM) and IMP α 3 (244 nM) (Fig. 2E,F). ORF1 NLSn_C bound to all tested IMP α isoforms, with moderate to high binding affinity (~ 4 -233 nM) and a preference for IMP α 1 (9 nM) and IMP α 3 (4 nM), but not

IMP β 's (Fig. 2G,H). These results suggest that TTDov ORF1 nuclear import could be mediated by two distinct NLSs with different NTR binding properties: a non-classical N-terminal NLS (NLSn), which interacts with both IMP α and IMP β NTRs, and a C-terminal cNLS (NLSc), specifically binding IMP α paralogs. This raises the possibility that TTDov ORF1 has evolved to simultaneously exploit multiple nuclear import pathways, as recently described for several cellular (Kimura et al. 2017, Mackmull et al. 2017) and viral (Nematollahzadeh et al. 2024) proteins.

Basic residues within TTDov ORF1 protein confer RanGTP- and IMP α / β 1-dependent nuclear targeting properties to heterologous proteins

The observed high affinity of ORF1 NLS peptides for NTRs suggests their involvement in protein nuclear translocation. To investigate this possibility, we measured their ability to confer nuclear targeting properties to GFP. To this end, we quantified the levels of nuclear accumulation of GFP fused to ORF1 NLSs when transiently expressed in mammalian cells, using GFP alone and a GFP-SV40 LTA NLS fusion protein as negative and positive controls, respectively (Fig. 3A). While GFP evenly distributed throughout the cell, GFP-SV40 NLS strongly accumulated in the cell nucleus (Fn/c of 1.1 and 9.7, respectively; Fig. 3B,C). Interestingly, GFP-ORF1 NLSn could be detected both in the nucleus and, to a lesser extent, in the cytoplasm (Fn/c of 3.2; Fig. 3), but strongly accumulated in the nucleoli, as evidenced by extensive co-localization with DsRed-nucleolin in co-transfection experiments (Supplementary Fig. S3A) and an Fno/n of 3.0 (Supplementary Fig. S3B,C). On the other hand, GFP-ORF1 NLSc was primarily localized in the nucleoplasm, with nucleolar exclusion and an Fn/c of 3.0. Depletion of intracellular RanGTP by incubating cells in media containing sodium azide and 2-deoxy-D-glucose (Fig. 3C,D, blue circles) or inhibition of the IMP α / β 1-dependent nuclear import pathway by co-transfection with plasmid mCherry-Bimax2 (Fig. 3C,D, pink circles) resulted in redistribution of all proteins between the nucleus and cytoplasm with a significant loss of nuclear fluorescence, indicating that both NLSs confer active, IMP α / β 1-dependent nuclear transport properties to GFP. However, the nucleolar accumulation of GFP-ORF1 NLSn was unaffected by either GTP depletion or co-expression with mCherry-Bimax2, thus nucleolar localization mediated by the long stretch of arginine residues is most likely due to interactions with cellular nucleic acids within the nucleoli, rather than an active process.

TTDov ORF1 nuclear import is primarily dependent on recognition of NLSc by IMP α / β 1 while accumulation in the nucleoli relies on the interaction of NLSn with nucleolar components

Our data suggest that ORF1 contains a nucleolar localization signal (NoLS)/non classical NLS at residues 27-57 (NLSn) and a cNLS at residues 632-643 (NLSc). To more precisely characterize the process of ORF1 intracellular transport and the specific contribution of such sequences to this process, we investigated the subcellular localization of full length ORF1 and derivatives thereof as fused to GFP (Fig. 4A). Transient expression of GFP-ORF1 in HEK293A cells resulted in the protein being detected at variable levels in the nucleus and cytosol, with an average Fn/c of 4.0 (Fig. 4B,C) and nuclear accumulation in more than 60% of transfected cells (Fig. 4D). Further, strong nucleolar accumulation was observed in all cells, highlighted by co-localization after transfection with a plasmid encoding DsRed-fibrillarin (Fig. 4E) and an average Fno/n

of 1.9 (Fig. 4F), with nucleolar accumulation in 100% of analysed cells (Fig. 4G). Importantly, nuclear targeting was strongly impaired by overexpression of mCherry-Bimax2, demonstrating that ORF1 nuclear import is primarily mediated by the IMP α / β 1 heterodimer (Fig. 4B,D). Deletion of the first 78 amino acids (ORF1 Δ 78), containing the ARM and NLSn, resulted in exclusion from the nucleoli (Fno/n of 0.6; Fig. 4E,G), but did not reduce nuclear accumulation of ORF1. Rather, the observed redistribution from nucleoli to the nucleoplasm caused a significant increase of the average Fn/c value to 12.4 (Fig. 4C). AlphaFold3 models predicted that ORF1 bound IMP α NTRs in the major NLS binding site through NLSc (Fig. 4H; Supplementary Table S3), with K637 interacting with key IMP α residues G150, T155, and D192 in the critical P2 binding pocket (Fig. 4I). Accordingly, introduction of the K637A substitution resulted in strong impairment of nuclear targeting (average Fn/c of 1.4; Fig. 4B,D), but not of nucleolar accumulation (Fno/n of 1.2; Fig. 4E,G). Strikingly, introduction of the K637A substitution together with deletion of residues 1-78 resulted in a predominantly cytosolic protein, which failed to accumulate in the nucleus (average Fn/c of 0.8; Fig. 4B,D) or the nucleolus (Fno/n of 0.7; Fig. 4E,G). Clearly, TTDov has evolved distinct sequences playing complementary roles in determining ORF1 subcellular localization, with NLSc mediating IMP α / β 1-dependent nuclear targeting and NLSn being responsible for nucleolar accumulation, most likely by mediating electrostatic interactions with nucleolar components such as rRNA.

Evolution of additional cNLSs in capsid proteins from Anelloviridae

The presence of a functional cNLS at the C-terminus of TTDov ORF1, in addition to the non-classical NLS located within the ARM, is noteworthy. In members of the *Circoviridae* family, the capsid protein ARM is known to mediate interactions with host NTRs and is considered the principal determinant of nuclear import (Patterson et al. 2013, Sarker et al. 2016). Capsid proteins of *Anelloviridae* are thought to have evolved from those of circoviruses through the stepwise acquisition of increasingly complex projection domains within the jelly-roll fold (Butkovic et al. 2023, Liou et al. 2024). Since TTDov ORF1 is one of the largest *Anelloviridae* capsid proteins reported so far, and it is predicted to possess a very large projection domain (Supplementary Fig. S1) similar to that of alphatorqueviruses (Butkovic et al. 2023), we hypothesized that the acquisition of large projection domains could correlate with the presence of additional cNLSs downstream of the ARM. To verify this hypothesis, we retrieved the amino acid sequences of capsids from both *Anelloviridae* and *Circoviridae* and scanned them for the presence of putative cNLSs (Fig. 5, Supplementary Tables S4,S5). Intriguingly, at least one cNLS was identified in 96.7% of capsids from *Anelloviridae* family members, but only in 50.3% of those from *Circoviridae*. Furthermore, cNLSs outside the ARM were identified in capsids encoded by several *Anelloviridae* family members, but not in gyroviruses which encode shorter capsids (Supplementary Fig. S4, Supplementary Tables S4-S6). Capsids encoded by *Anelloviridae* are significantly longer and more variable in length than those encoded by *Circoviridae* (592 ± 108 versus 295 ± 27 amino acids, respectively; Fig. 5A), consistent with the acquisition of projection domains of increasing size (Butkovic et al. 2023). Furthermore, the number (Fig. 5B) and predicted activity (Fig. 5C) of cNLSs is higher in anelloviruses than in circoviruses. Moreover, in circoviruses the predicted cNLS is most frequently located within the ARM, while more than 70% of anelloviruses possess an additional cNLS downstream (Fig. 5D). Linear regression showed significant

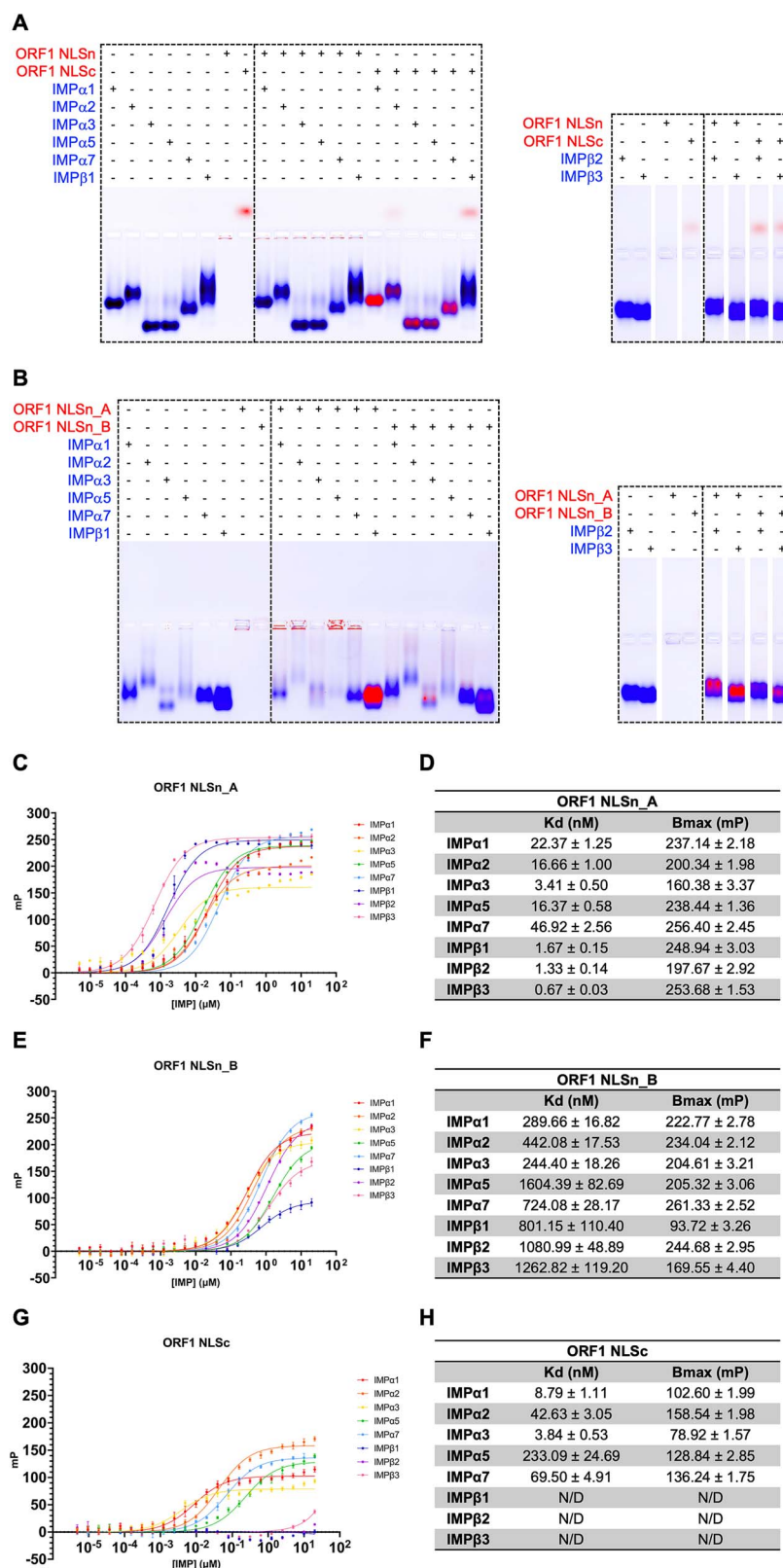


Figure 2. TTDov ORF1 can bind NTRs through both its N-terminal and C-terminal NLS. (A,B) electrophoretic mobility shift assays (EMSAs) assessing binding of NTRs (20 μ M) and FITC-tagged NLS peptides (10 μ M). Proteins are shown in blue; peptides are shown in red; co-migration indicates binding; same IMP β 2/3 protein only controls shown in both a and B. See [Supplementary Fig. S2C-G](#) for individual FITC and Coomassie images and EMSA controls. (C-H) fluorescence polarization assays measuring binding affinity between NTRs (20 μ M starting concentration) and FITC-tagged NLS peptides (10 nM), including calculated dissociation constant (Kd) and maximum binding (Bmax) values. Data shown as n = 3; error bars represent mean $\pm$ standard error of the mean; N/D = not determined; mP = millipolarization units.

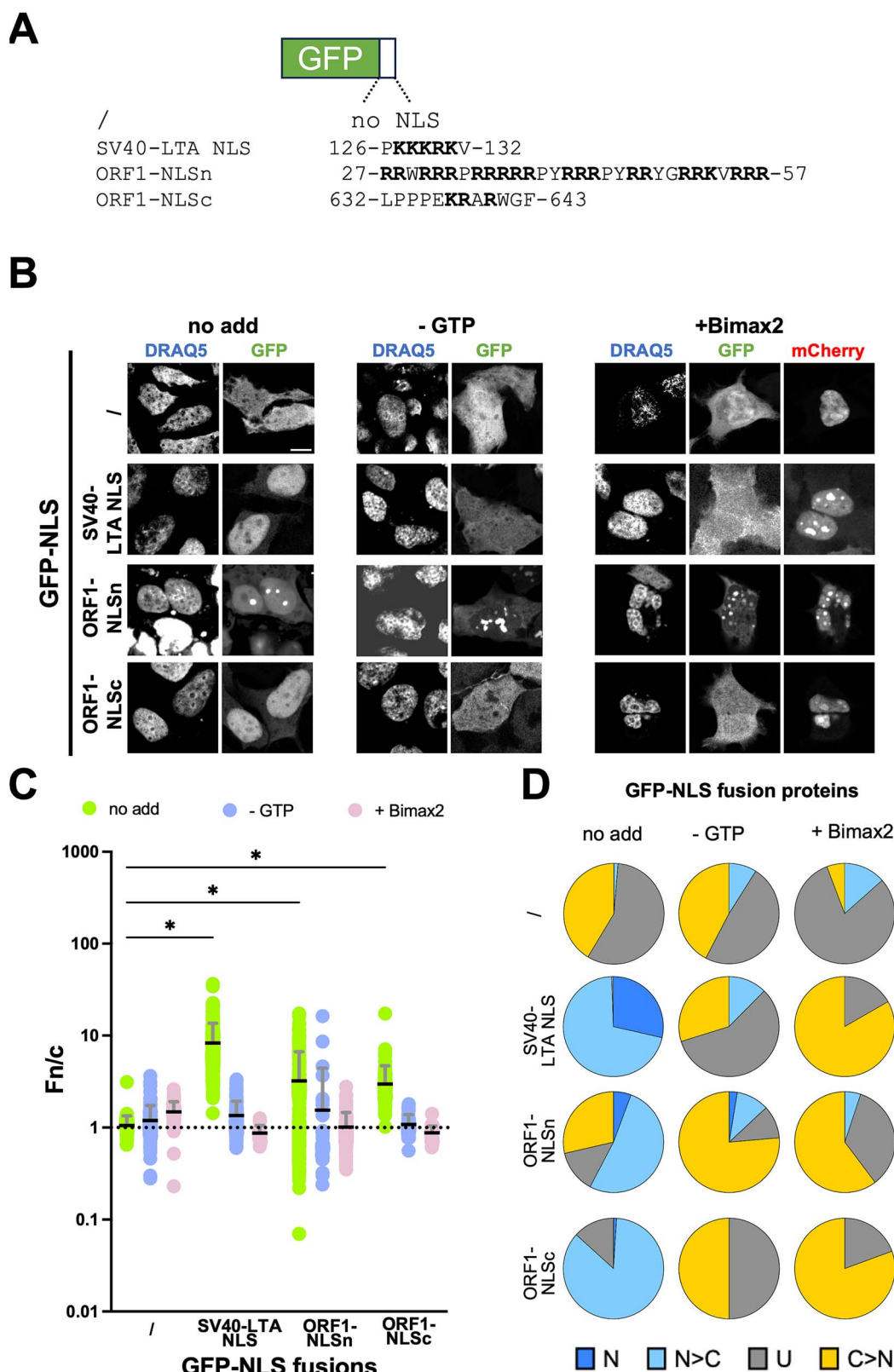


Figure 3. Characterization of nuclear and nucleolar targeting properties of TTDov ORF1 NLSs in a cellular context. (A) HEK293A cells were transfected to express the indicated GFP fusion proteins (basic residues shown in bold), alone or in the presence of mCherry-Bimax2. 24 hrs p.t., cells were either left untreated or incubated for 30 min with an energy depletion media (– GTP) before being stained and fixed for CLSM imaging and analysis. (B) Representative images of the 633 nm (DRAQ5), 488 nm (GFP), and 561 nm (mCherry) laser channels are shown, relative to the indicated GFP fusion proteins. Scale bar = 10 μ m. (C) Images such as those shown in (B) were analysed for quantification of the levels of nuclear accumulation (Fn/c) at the single cell level. Data are shown as individual measurements (circles), along with mean (black horizontal bars) and standard deviation of the mean (grey vertical bars), including the results of the Welch and Brown-Forsythe one-way ANOVA for significance between the indicated proteins (*: $p \leq 0.05$); pooled data from at least two independent experiments. (D) The percentage of cells relative to each indicated fusion protein displaying the indicated subcellular localization is shown. N: Nuclear, $Fn/c \geq 10$; N > C: More nuclear than cytosolic, $2 \leq Fn/c < 10$; U: Ubiquitous, $1 \leq Fn/c < 2$; C > N: More cytosolic than nuclear, $Fn/c < 1$.

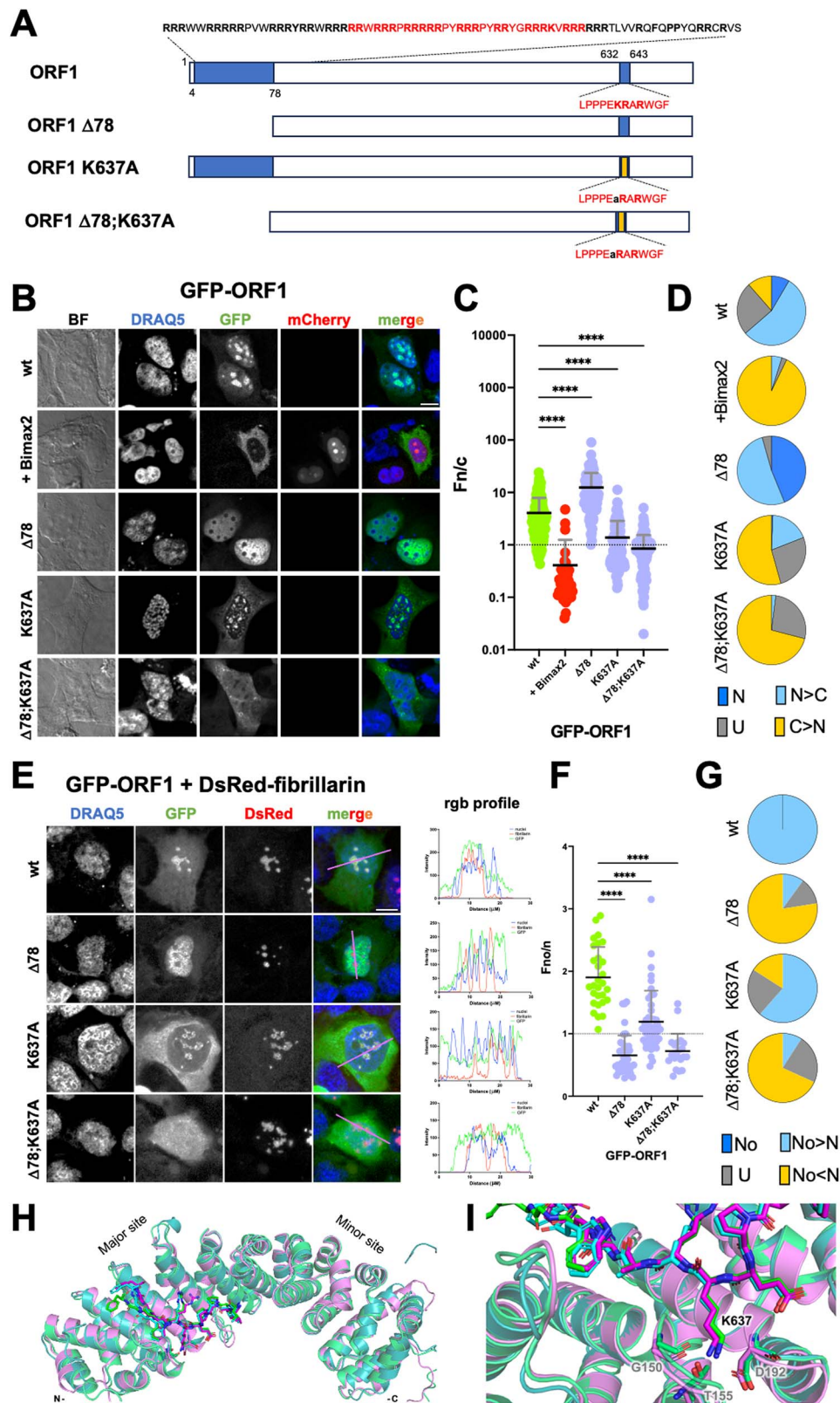


Figure 4. TTDov ORF1 is translocated into the nucleus by IMP α / β 1 which recognizes NLS, and accumulates in the nucleoli through a NoLS at the N-terminus. (A) Schematic representation of GFP-ORF1 fusion proteins analysed, along with the respective position and sequence of the targeting signals identified here. ORF1 amino acid sequence is shown as a white box; identified NLS regions are shown as blue boxes; NLS sequences are in red; mutated NLS is shown as a yellow box. (B) HEK293A cells were transfected with lipofectamine 2000 to express the indicated GFP-ORF1 fusion proteins, alone or in the presence of mCherry-Bimax2. Twenty-four hrs p.t., cells were incubated with DRAQ5 to stain cell nuclei, fixed, and processed for CLSM imaging and analysis. Representative images of the bright field (BF), 633 nm (DRAQ5), 488 nm (GFP), and 561 nm (mCherry) laser channels are shown, along with a merged image (merge). Scale bar = 10 μ m. (C) Images such as those shown in (B) were quantitatively analysed to calculate the levels of nuclear accumulation (Fn/c) at the single cell level. Data are shown as individual measurements (circles), along with mean (black horizontal bars) and standard deviation (grey shaded area).

positive correlation between capsid length and cNLS activity among anelloviruses (Fig. 5E) but not circoviruses (Fig. 5F). Similar results were obtained by analysing the correlation between capsid length and cNLS number (Fig. 5G) and cNLS score (Fig. 5H) among the most abundant *Anelloviridae* genera. In addition, the percentage of capsids bearing an additional cNLS increased progressively with capsid size (Fig. 5I). Such cNLSs were never identified in gyrovirus capsids, which are significantly shorter than those from other genera (432.9 ± 46.7 amino acids), while they were systematically found in all alphatorqueviruses, which encode for the largest capsids (732.4 ± 41.0 amino acids). Pairwise alignment of *Anelloviridae* ORF1 proteins revealed that TTDov NLS_c is frequently conserved among family members and located between 17 and 109 amino acids from the C-terminus (Supplementary Fig. S5, Supplementary Table S6). Therefore, the acquisition of an additional cNLS downstream of the ARM represents a hallmark of *Anelloviridae* evolution and correlates with the presence of a large projection domain.

Discussion

The landscape of *Anelloviridae*-host interaction is largely uncharacterized, mainly due to the lack of suitable cellular systems to study the virus life cycle. In this context, little is known regarding the functional interaction between host cell NTRs and capsid proteins. Molecular studies of the virus-host interface may shed light on their potential pathogenicity. Keeping this in mind, we used TTDov as a model to study the interaction between TTVs and the host cell nuclear transport machinery. Our study identified distinct signals responsible for ORF1 nuclear import and nucleolar targeting, as well as the NTRs responsible for this process. This is the first study to investigate the physical and functional consequences of the interaction between TTV proteins and the host cell nuclear transport machinery, paving the way to a better understanding of TTV-host cell interaction, pathogenicity, and evolution.

Role of TTDov ORF1 ARM in nucleolar accumulation

The capsid proteins from several DNA and RNA viruses, including TTV1 P/1C1, are known to localize in the nucleolus to modulate host transcription, process RNA, and promote viral replication (Hiscox et al. 2001, Mueller et al. 2008, Taylor et al. 2017, Wang et al. 2017, Nair and Zlotnick 2021, Selinger et al. 2022, Zhou et al. 2022). When expressed in the absence of other viral proteins, TTDov ORF1 strongly accumulated in the nucleolus (Fig. 4B, wt). Such nucleolar accumulation depends on an N-terminal NLS_n

located within the ARM. Although the ARM has been implicated in genome packaging and nuclear targeting for closely related *Circoviridae* capsids (Patterson et al. 2013, Sarker et al. 2016), our data indicate that in the context of TTDov, the ARM primarily supports genome packaging and nucleic acid binding rather than nuclear import. Firstly, TTDov ORF1 NLS_n is sufficient to confer strong nucleolar localization to GFP (Fig. 3, ORF1-NLS_n; Supplementary Fig. S3) and is absolutely required for nucleolar—but not nuclear—targeting of the full-length protein, as evidenced by the strong nucleoplasmic localization observed upon deletion of the first 78 amino acids (Fig. 4B-D, Δ78). Secondly, while TTDov ORF1 nuclear import strongly relies on IMP α /β1, nucleolar accumulation is independent of this pathway (Fig. 4).

Role of TTDov ORF1 NLS_c in IMP α /β1-dependent nuclear import

Although TTDov ORF1 NLS_n can bind multiple IMP α and IMP β NTRs with high affinity (Fig. 2B,C,F), ORF1 nuclear import primarily relies on the downstream NLS_c, which preferentially interacts with IMP α 's (Fig. 2A,G,H) and is predicted to bind the major site of IMP α (Fig. 4H). The K637A substitution, targeting a predicted key binding determinant within NLS_c (Fig. 4I), significantly decreased nuclear levels of ORF1 (Fig. 4B-D, K637A). Simultaneous deletion of NLS_n further reduced nuclear targeting (Fig. 4B-D, Δ78;K637A), indicating that while NLS_c is the main driver of ORF1 nuclear import, NLS_n also contributes to the process. Interestingly, although Bimax2-mediated inhibition of the IMP α /β1-dependent nuclear import pathway almost completely blocked nuclear localization of ORF1, some protein still accumulated in the nucleolus (Fig. 4B-D, +Bimax2). This observation is consistent with the ability of ORF1 residues 27-RRWRRRPRRRRPPYRRPYRR-47 to mediate high-affinity interactions with IMP β 1, IMP β 2, and IMP β 3 (Fig. 2C-D). Taken together, our results suggest that nucleolar localization is dependent upon the ability of the arginine-rich NLS_n to interact with nucleic acids, which are highly abundant in the nucleolus, or other nucleolar components, as demonstrated for capsid proteins from other DNA and RNA viruses (Selinger et al. 2022, Zhou et al. 2022, Wang et al. 2023), following nuclear entry mediated by the concerted action of NLS_c and NLS_n.

Acquisition of additional cNLSs as a hallmark of *Anelloviridae* capsid evolution

This redundancy of targeting signals and NTR specificity is not unusual and has previously been reported for other viral proteins, including pVII from both human and animal adenoviruses (Nematollahzadeh et al. 2024), however it does appear to distinguish *Anelloviridae* from *Circoviridae* capsid proteins. Indeed,

standard deviation of the mean (grey vertical bars), including the results of the Welch and Brown-Forsythe one-way ANOVA for significance between the indicated proteins (****: $p \leq 0.001$); pooled data from at least two independent experiments. (D) the percentage of cells relative to each indicated fusion protein displaying the indicated subcellular localization is shown. N: Nuclear, $F_n/c > 10$; N > C: More nuclear than cytosolic, $2 < F_n/c < 10$; U: Ubiquitous, $1 < F_n/c < 2$; C > N: More cytosolic than nuclear, $F_n/c < 1$. (E) The indicated GFP fusion proteins were transiently co-expressed in HEK293A cells with DsRed-fibrillarin by means of lipofectamine 2000 transfection. Twenty-four hrs p.t., cells were incubated with DRAQ5 to stain cell nuclei, fixed, and processed for CLSM imaging and analysis. Representative images of the 633 nm (DRAQ5), 488 nm (GFP), and 561 nm (DsRed) laser channels are shown, along with a merged image (merge) and a rgb profile plot across the indicated area (rgb profile). Scale bar = 10 mm. (F) Images such as those shown in (E) were quantitatively analysed to calculate the levels of nucleolar accumulation (F_{no}/n) at the single cell level. Data are shown as individual measurements (circles), along with mean (black horizontal bars) and standard deviation of the mean (grey vertical bars), including the results of the Welch and Brown-Forsythe one-way ANOVA for significance between the indicated proteins (****: $p \leq 0.001$); pooled data from at least two independent experiments. (G) the percentage of cells relative to each indicated fusion protein displaying the indicated subcellular localization is shown. No: Nucleolar, $F_{no}/n > 10$; No > N: More nucleolar than nuclear, $2 < F_{no}/n < 10$; U: Ubiquitous, $1 < F_{no}/n < 2$; No < N: More nuclear than nucleolar, $F_{no}/n < 1$. (H) Superposition of AlphaFold3 models of TTDov ORF1 NLS_c bound in the major NLS binding site of IMP α 1 (teal), IMP α 3 (green), and IMP α 7 (purple). (I) K637 is shown to interact with key residues in the P2 binding pocket of IMP α , specifically G150, T155, and D192 (IMP α 1 numbering). The top ranked predictions are shown. IMP α shown in cartoon; ORF1 NLS_c and IMP α P2 residues shown in stick representation.

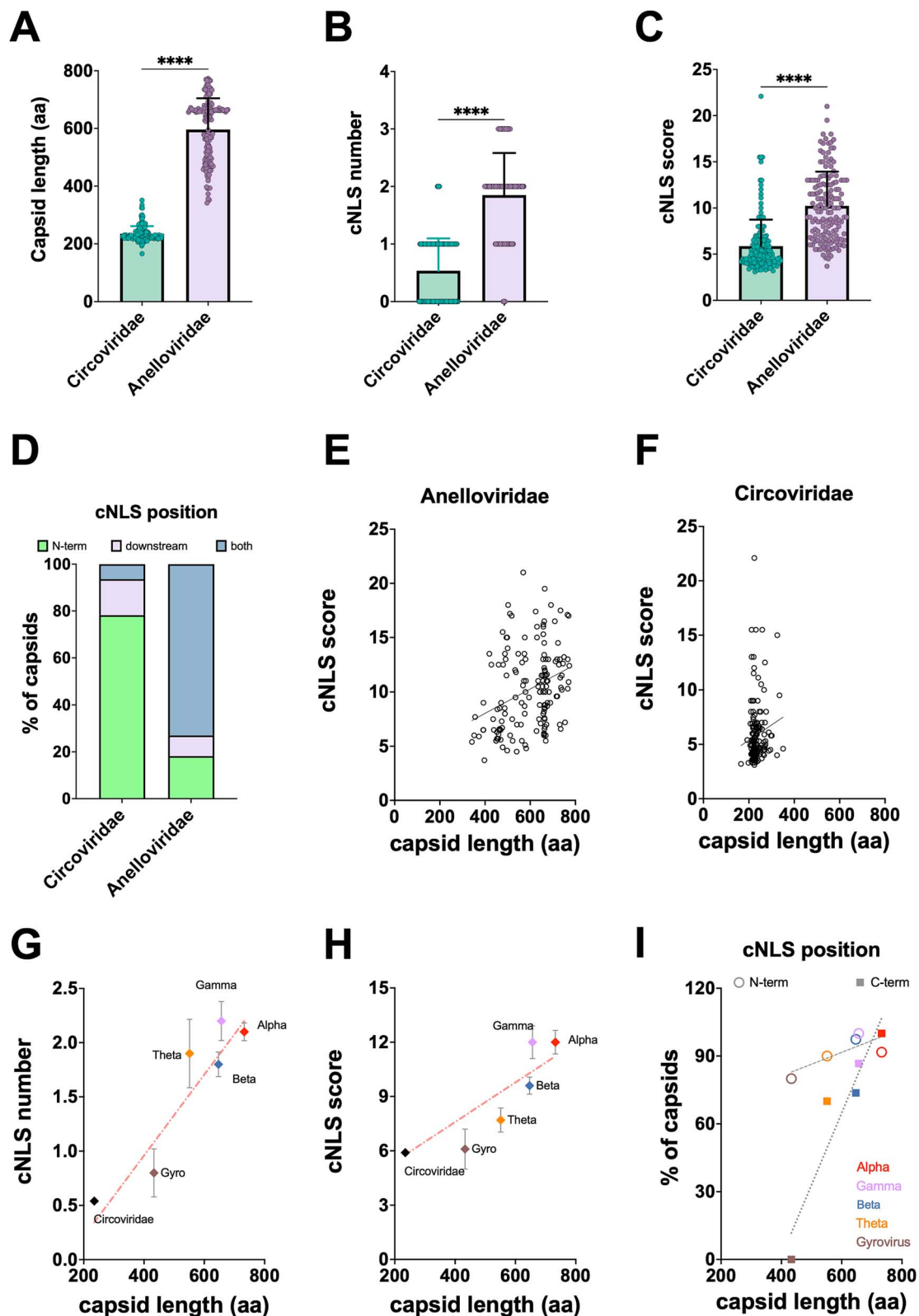


Figure 5. Evolution of C-terminal cNLSs downstream of the ARM correlates with capsid size and represents a potential evolutionary hallmark of Anelloviridae. Capsid protein sequences from members of the Circoviridae and Anelloviridae families were retrieved from UniProt and analysed using cNLS mapper to identify putative classical nuclear localization signals (cNLSs) with a predicted score > 5. (A-C) average capsid protein length (A), number of predicted cNLSs (B), and predicted cNLS activity score (C) for Circoviridae and Anelloviridae. Data are presented as individual values (circles), along with mean (columns) $\pm$ standard deviation (vertical bars). Statistical significance was assessed using the Mann Whitney test (****: $p \leq 0.0001$). (D) Proportion of capsid proteins bearing a predicted cNLS exclusively within the ARM (N-term), exclusively downstream of the ARM (downstream), or in both positions (both). (E-F) linear regression between capsid length and cNLS activity score in Anelloviridae (E) and Circoviridae (F). (G-H) linear regression between capsid length and cNLS number (G) and cNLS activity score (H) across Circoviridae and the indicated anellovirus genera. (I) Linear regression between capsid length and the percentage of capsid proteins containing a cNLS within the ARM (N-term) or downstream (C-term) across Circoviridae and the indicated anellovirus genera.

Anelloviridae capsids are thought to have evolved from that of Circoviridae through the progressive acquisition of increasingly large projection domains between β -strands H-I of the jelly-roll fold (Sarker et al. 2016, Butkovic et al. 2023, Liou et al. 2024). Like alphatorqueviruses (Butkovic et al. 2023), TTDov ORF1 possesses a remarkably large projection domain (Supplementary Fig. S1). Our analyses showed that longer projection domains correlate with the acquisition of additional cNLSs located downstream of the ARM (Fig. 5I). In this respect, gyroviruses appear very similar to Circoviridae in that they lack projection domains and contain only an N-terminal cNLS within the ARM. In contrast, all alphatorqueviruses possess large projection domains and encode an additional downstream cNLS (Fig. 5I; Supplementary Figs. S4,S5). Intriguingly, the capsid of the Circoviridae BFDV lacks a C-terminal cNLS. This protein can form distinct macromolecular assemblies depending on the presence of ssDNA, forming large 60-mer VLPs in the presence of DNA and smaller 10-mer VLPs in its absence. The ARM is positioned on the interior of the capsid interacting with DNA in the large VLPs, but remains exposed and accessible for NTR binding in the smaller ones (Patterson et al. 2013, Sarker et al. 2016, Chen et al. 2020). It is therefore plausible that the acquisition of larger projection domains limits the formation of these smaller capsid assemblies where the ARM is exposed, necessitating the evolution of an additional cNLS downstream. This hypothesis is supported by the cryo-EM structure of the LY1 torque teno mini virus capsid protein that forms only 60-mer icosahedral capsids, with the ARM buried within the interior (Liou et al. 2024). Superposition of the predicted TTDov ORF1 model onto this structure revealed that ORF1 NLSn and NLSc are likely located within the interior and exterior of the viral capsid, respectively (Supplementary Fig. S6).

A model for TTDov capsid nuclear transport

Although our cellular assays were performed using transiently expressed GFP-tagged ORF1 and not intact viral capsids during viral infection, our data suggest that the two NLSs identified here might play different roles during viral entry and assembly. During viral replication, ORF1 synthesized in the cytoplasm could be translocated into the nucleus by the concerted action of NLSn and NLSc, predominantly through IMP α / β 1 and supported by IMP β 1, IMP β 2, and IMP β 3. Subsequently, NLSn located within the ARM could bind viral DNA during packaging and assembly. As a consequence, NLSn would not be available for interaction with NTRs during the first steps of new rounds of viral infection, thus nuclear import of the viral genome is likely mediated by IMP α / β 1 through recognition of NLSc. In this context, the acquisition of an additional downstream cNLS might represent a particular evolutionary path pursued by Anelloviridae members to maximize nuclear import of viral capsid proteins.

Overall, by elucidating the functional interplay between TTDov ORF1 and host NTRs, our study provides critical mechanistic insight into TTV biology and a conceptual framework for understanding the evolution of nuclear trafficking strategies in small DNA viruses.

Acknowledgements

We thank Yoshihiro Yoneda, Masahiro Oka, and Denis Archambault for sharing expression plasmids. We thank Emily Wagon for sharing EMSA control reagents. We thank Jeff Nanson for advice with structural modelling and constructive criticism of the manuscript.

Author contributions

Conceptualization, G.A.; methodology, D.A. and O.T.; validation, G.F.P. and G.A.; formal analysis, G.F.P., S.P., S.N. and G.A.; investigation, G.F.P., S.P. and S.N.; resources, D.A. and O.T.; data curation, G.F.P. and G.A.; writing—original draft preparation, G.F.P. and G.A.; writing—review and editing, G.F.P., S.P., S.N., S.S. and G.A.; visualization, G.F.P., S.P. and G.A.; supervision, J.K.F. and G.A.; project administration, G.A.; funding acquisition, J.K.F. and G.A. All authors have read and agreed to the published version of the manuscript.

Supplementary data

Supplementary data is available at VEVOLU Journal online.

Conflict of interest: The authors declare no conflicts of interest.

Funding

This research was partially funded by the Italian Ministry for Universities and Research (MUR) Progetto PRIN 2022 cod. 2022F2YJNK—Acr. INTERROGA, CUP: C53D23003110006 awarded to G.A. Additionally, G.F.P. and J.K.F. acknowledge funding through the Training Hub promoting Regional Industry and Innovation in Virology and Epidemiology (THRIIVE) program from the Australian Government. S.S. is the recipient of an Australian Research Council Discovery Early Career Researcher Award (grant number DE200100367) funded by the Australian Government. S.N. was supported by a scholarship funded by EU within the NextGeneration EU-MUR PNRR Extended Partnership initiative on Emerging Infectious Diseases (Project no. PE00000007, INF-ACT).

Data availability

The original data presented in the study are openly available in Research data UNIPD at <https://researchdata.cab.unipd.it/1643>

Institutional review board statement

Not applicable.

Informed consent statement

Not applicable.

References

- Abramson J, Adler J, Dunger J et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*. 2024;**630**: 493–500. <https://doi.org/10.1038/s41586-024-07487-w>
- Alvisi G, Avanzi S, Musiani D et al. Nuclear import of HSV-1 DNA polymerase processivity factor UL42 is mediated by a C-terminally located bipartite nuclear localization signal. *Biochemistry*. 2008;**47**: 13764–77. <https://doi.org/10.1021/bi800869y>
- Alvisi G, Roth DM, Camozzi D et al. The flexible loop of the human cytomegalovirus DNA polymerase processivity factor ppUL44 is required for efficient DNA binding and replication in cells. *J Virol* 2009;**83**:9567–76. <https://doi.org/10.1128/JVI.00669-09>
- Alvisi G, Jans DA, Camozzi D et al. Regulated transport into the nucleus of herpesviridae DNA replication core proteins. *Viruses*. 2013;**5**:2210–34. <https://doi.org/10.3390/v5092210>

- Alvisi G, Paolini L, Contarini A et al. Intersectin goes nuclear: secret life of an endocytic protein. *The Biochemical journal* 2018;**475**: 1455–72. <https://doi.org/10.1042/BCJ20170897>
- Alvisi G, Manaresi E, Cross EM et al. Importin alpha/beta-dependent nuclear transport of human parvovirus B19 non-structural protein 1 is essential for viral replication. *Antivir Res* 2023;**213**:105588. <https://doi.org/10.1016/j.antiviral.2023.105588>
- Biagini P. Classification of TTV and related viruses (anelloviruses). *Curr Top Microbiol Immunol* 2009;**331**:21–33.
- Bonamassa B, Ciccarese F, Antonio VD et al. Hepatitis C virus and host cell nuclear transport machinery: a clandestine affair. *Front Microbiol* 2015;**6**:619. <https://doi.org/10.3389/fmicb.2015.00619>
- Bourgeois B, Hutten S, Gottschalk B et al. Nonclassical nuclear localization signals mediate nuclear import of CIRBP. *Proc Natl Acad Sci USA* 2020;**117**:8503–14. <https://doi.org/10.1073/pnas.1918944117>
- Butkovic A, Kraberger S, Smele Z et al. Evolution of anelloviruses from a circovirus-like ancestor through gradual augmentation of the jelly-roll capsid protein. *Virus Evol* 2023;**9**:vead035. <https://doi.org/10.1093/ve/vead035>
- Chen JK, Hsiao C, Lo AR et al. Characterization of the nuclear localization sequence of beak and feather disease virus capsid proteins and their assembly into virus-like particles. *Virus Res* 2020;**289**:198144. <https://doi.org/10.1016/j.virusres.2020.198144>
- Cross, Akbari N, Ghassabian H et al. A functional and structural comparative analysis of large tumor antigens reveals evolution of different importin alpha-dependent nuclear localization signals. *Protein Sci* 2024a;**33**:e4876. <https://doi.org/10.1002/pro.4876>
- Cross EM, Marin O, Ariawan D et al. Structural determinants of phosphorylation-dependent nuclear transport of HCMV DNA polymerase processivity factor UL44. *FEBS Lett* 2024b;**598**: 199–209. <https://doi.org/10.1002/1873-3468.14797>
- Deb B, Uddin A, Chakraborty S. Composition, codon usage pattern, protein properties, and influencing factors in the genomes of members of the family Anelloviridae. *Arch Virol* 2021;**166**:461–74. <https://doi.org/10.1007/s00705-020-04890-2>
- Fahsbender E, Burns JM, Kim S et al. Diverse and highly recombinant anelloviruses associated with Weddell seals in Antarctica. *Virus Evol* 2017;**3**:vex017. <https://doi.org/10.1093/ve/vex017>
- Ghafoori SM, Sethi A, Petersen GF et al. RNA binding properties of SOX family members. *Cells*. 2024;**13**. <https://doi.org/10.3390/cells13141202>
- Gomez Corredor A, Archambault D. The bovine immunodeficiency virus rev protein: identification of a novel lentiviral bipartite nuclear localization signal harboring an atypical spacer sequence. *J Virol* 2009;**83**:12842–53. <https://doi.org/10.1128/JVI.01613-09>
- Gorlich D, Kostka S, Kraft R et al. Two different subunits of importin cooperate to recognize nuclear localization signals and bind them to the nuclear envelope. *Curr Biol* 1995;**5**:383–92. [https://doi.org/10.1016/S0960-9822\(95\)00079-0](https://doi.org/10.1016/S0960-9822(95)00079-0)
- Hiscox JA, Wurm T, Wilson L et al. The coronavirus infectious bronchitis virus nucleoprotein localizes to the nucleolus. *J Virol* 2001;**75**:506–12. <https://doi.org/10.1128/JVI.75.1.506-512.2001>
- Kaczorowska J, van der Hoek L. Human anelloviruses: diverse, omnipresent and commensal members of the virome. *FEMS Microbiol Rev* 2020;**44**:305–13. <https://doi.org/10.1093/femsre/fuaa007>
- Kimura M, Morinaka Y, Imai K et al. Extensive cargo identification reveals distinct biological roles of the 12 importin pathways. *eLife*. 2017;**6**. <https://doi.org/10.7554/eLife.21184>
- Kosugi S, Hasebe M, Matsumura N et al. Six classes of nuclear localization signals specific to different binding grooves of importin alpha. *J Biol Chem* 2009;**284**:478–85. <https://doi.org/10.1074/jbc.M807017200>
- Kovalevskiy O, Mateos-Garcia J, Tunyasuvunakool K. AlphaFold two years on: validation and impact. *Proc Natl Acad Sci USA* 2024;**121**:e2315002121. <https://doi.org/10.1073/pnas.2315002121>
- Liou SH, Boggavarapu R, Cohen NR et al. Structure of anellovirus-like particles reveal a mechanism for immune evasion. *Nat Commun* 2024;**15**:7219. <https://doi.org/10.1038/s41467-024-51064-8>
- Mackmull MT, Klaus B, Heinze I et al. Landscape of nuclear transport receptor cargo specificity. *Mol Syst Biol* 2017;**13**:962. <https://doi.org/10.15252/msb.20177608>
- Madeira F, Madhusoodanan N, Lee J et al. The EMBL-EBI job dispatcher sequence analysis tools framework in 2024. *Nucleic Acids Res* 2024;**52**:W521–5. <https://doi.org/10.1093/nar/gkae241>
- Marfori M, Mynott A, Ellis JJ et al. Molecular basis for specificity of nuclear import and prediction of nuclear localization. *Biochim Biophys Acta* 2011;**1813**:1562–77. <https://doi.org/10.1016/j.bbamcr.2010.10.013>
- Martinez-Guino L, Ballester M, Segales J et al. Expression profile and subcellular localization of torque Teno sus virus proteins. *J Gen Virol* 2011;**92**:2446–57. <https://doi.org/10.1099/vir.0.033134-0>
- Mueller B, Maerz A, Doberstein K et al. Gene expression of the human torque Teno virus isolate P/1C1. *Virology*. 2008;**381**:36–45. <https://doi.org/10.1016/j.virol.2008.08.017>
- Nair S, Zlotnick A. HBV Core protein is in flux between cytoplasmic, nuclear, and nucleolar compartments. *mBio*. 2021;**12**. <https://doi.org/10.1128/mBio.03514-20>
- Nematollahzadeh S, Athukorala A, Donnelly CM et al. Mechanistic insights into an ancient adenovirus precursor protein VII show multiple nuclear import receptor pathways. *Traffic*. 2024;**25**:e12953. <https://doi.org/10.1111/tra.12953>
- Nishizawa T, Okamoto H, Konishi K et al. A novel DNA virus (TTV) associated with elevated transaminase levels in posttransfusion hepatitis of unknown etiology. *Biochem Biophys Res Commun* 1997;**241**:92–7. <https://doi.org/10.1006/bbrc.1997.7765>
- Okamoto H, Takahashi M, Nishizawa T et al. Replicative forms of TT virus DNA in bone marrow cells. *Biochem Biophys Res Commun* 2000a;**270**:657–62. <https://doi.org/10.1006/bbrc.2000.2481>
- Okamoto H, Nishizawa T, Tawara A et al. Species-specific TT viruses in humans and nonhuman primates and their phylogenetic relatedness. *Virology*. 2000b;**277**:368–78. <https://doi.org/10.1006/viro.2000.0588>
- Patterson EI, Dombrowski AK, Swarbrick CM et al. Structural determination of importin alpha in complex with beak and feather disease virus capsid nuclear localization signal. *Biochem Biophys Res Commun* 2013;**438**:680–5. <https://doi.org/10.1016/j.bbrc.2013.07.122>
- Qiu J, Kakkola L, Cheng F et al. Human circovirus TT virus genotype 6 expresses six proteins following transfection of a full-length clone. *J Virol* 2005;**79**:6505–10. <https://doi.org/10.1128/JVI.79.10.6505-6510.2005>
- Requiao RD, Carneiro RL, Moreira MH et al. Viruses with different genome types adopt a similar strategy to pack nucleic acids based on positively charged protein domains. *Sci Rep* 2020;**10**:5470. <https://doi.org/10.1038/s41598-020-62328-w>
- van Rijn AL, Roos R, Dekker FW et al. Torque Teno virus load as marker of rejection and infection in solid organ transplantation - a systematic review and meta-analysis. *Rev Med Virol* 2023;**33**:e2393. <https://doi.org/10.1002/rmv.2393>
- Sarker S, Terron MC, Khandokar Y et al. Structural insights into the assembly and regulation of distinct viral capsid complexes. *Nat Commun* 2016;**7**:13014. <https://doi.org/10.1038/ncomms13014>

- Schindelin J, Arganda-Carreras I, Frise E et al. Fiji: an open-source platform for biological-image analysis. *Nat Methods* 2012;**9**: 676–82. <https://doi.org/10.1038/nmeth.2019>
- Schwoebel ED, Ho TH, Moore MS. The mechanism of inhibition of ran-dependent nuclear transport by cellular ATP depletion. *J Cell Biol* 2002;**157**:963–74. <https://doi.org/10.1083/jcb.2001.11077>
- Selinger M, Novotny R, Sys J et al. Tick-borne encephalitis virus capsid protein induces translational shutoff as revealed by its structural-biological analysis. *J Biol Chem* 2022;**298**:102585. <https://doi.org/10.1016/j.jbc.2022.102585>
- Smith KM, Di Antonio V, Bellucci L et al. Contribution of the residue at position 4 within classical nuclear localization signals to modulating interaction with importins and nuclear targeting. *Biochim Biophys Acta* 2018;**1865**:1114–29. <https://doi.org/10.1016/j.bbamcr.2018.05.006>
- Studier FW. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* 2005;**41**:207–34. <https://doi.org/10.1016/j.pep.2005.01.016>
- Taylor A, Liu X, Zaid A et al. Mutation of the N-terminal region of chikungunya virus capsid protein: implications for vaccine design. *mBio* 2017;**8**. <https://doi.org/10.1128/mBio.01970-16>
- Timney BL, Raveh B, Mironska R et al. Simple rules for passive diffusion through the nuclear pore complex. *J Cell Biol* 2016;**215**: 57–76. <https://doi.org/10.1083/jcb.201601004>
- Trevisan M, Di Antonio V, Radeghieri A et al. Molecular requirements for self-interaction of the respiratory syncytial virus matrix protein in living mammalian cells. *Viruses*. 2018;**10**. <https://doi.org/10.3390/v10030109>
- Tsujii A, Miyamoto Y, Moriyama T et al. Retinoblastoma-binding protein 4-regulated classical nuclear transport is involved in cellular senescence. *J Biol Chem* 2015;**290**:29375–88. <https://doi.org/10.1074/jbc.M115.681908>
- Varsani A, Opriessnig T, Celer V et al. Taxonomic update for mammalian anelloviruses (family Anelloviridae). *Arch Virol* 2021;**166**: 2943–53. <https://doi.org/10.1007/s00705-021-05192-x>
- Varsani A, Kraberger S, Opriessnig T et al. Anelloviridae taxonomy update 2023. *Arch Virol* 2023;**168**:277. <https://doi.org/10.1007/s00705-023-05903-6>
- Varsani A, Harrach B, Roumagnac P et al. 2024 taxonomy update for the family Circoviridae. *Arch Virol* 2024;**169**:176. <https://doi.org/10.1007/s00705-024-06107-2>
- Vasilyev EV, Trofimov DY, Tonevitsky AG et al. Torque Teno virus (TTV) distribution in healthy Russian population. *Virology* 2009;**6**:134. <https://doi.org/10.1186/1743-422X-6-134>
- Venter PA, Marshall D, Schneemann A. Dual roles for an arginine-rich motif in specific genome recognition and localization of viral coat protein to RNA replication sites in flock house virus-infected cells. *J Virol* 2009;**83**:2872–82. <https://doi.org/10.1128/JVI.01780-08>
- Vietzen H, Simonitsch C, Friedel B et al. Torque Teno viruses exhaust and imprint the human immune system via the HLA-E/NKG2A axis. *Front Immunol* 2024;**15**:1447980. <https://doi.org/10.3389/fimmu.2024.1447980>
- Virgin HW, Wherry EJ, Ahmed R. Redefining chronic viral infection. *Cell*. 2009;**138**:30–50. <https://doi.org/10.1016/j.cell.2009.06.036>
- Wang L, Tan H, Wu M et al. Dynamic virus-dependent subnuclear localization of the capsid protein from a Geminivirus. *Front Plant Sci* 2017;**8**:2165. <https://doi.org/10.3389/fpls.2017.02165>
- Wang D, Hou L, Ji Y et al. Ubiquitination-dependent degradation of nucleolin mediated by porcine circovirus type 3 capsid protein. *J Virol* 2023;**97**:e0089423. <https://doi.org/10.1128/jvi.00894-23>
- Wing CE, Fung HYJ, Chook YM. Karyopherin-mediated nucleocytoplasmic transport. *Nat Rev Mol Cell Biol* 2022;**23**:307–28. <https://doi.org/10.1038/s41580-021-00446-7>
- Zheng H, Ye L, Fang X et al. Torque Teno virus (SANBAN isolate) ORF2 protein suppresses NF- κ B pathways via interaction with I κ B kinases. *J Virol* 2007;**81**:11917–24. <https://doi.org/10.1128/JVI.01101-07>
- Zhou J, Wang Y, Qiu Y et al. Contribution of DEAD-box RNA helicase 21 to the nucleolar localization of porcine circovirus type 4 capsid protein. *Front Microbiol* 2022;**13**:802740. <https://doi.org/10.3389/fmicb.2022.802740>