Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Functional screening identifies candidate residues associated with nucleoprotein-mediated suppression of Borna disease virus 1 replication

Kazuyoshi Nakaji
Kyoto University
Houshin Hiramatsu
Kyoto University
Madoka Sakai
Kyoto University
Junna Kawasaki
Chiba University
Keizo Tomonaga
Kyoto University
Akiko Makino
makino@infront.kyoto-u.ac.jp

Kyoto University

Research Article

Keywords: Borna disease virus 1, Nucleoprotein, Functional screening, Mutagenesis, Mutation profiling, Endogenous bornavirus-like nucleoprotein, Viral suppression

Posted Date: May 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9582930/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Borna disease virus 1 (BoDV-1) nucleoprotein (N) is an essential component of the viral ribonucleoprotein complex, yet its overexpression suppresses viral infection. While this effect has been described in newly infected cells, its relevance in persistently infected cells remains unclear. Here, we show that N overexpression reduces infection levels in persistently infected cells with slower kinetics than those reported in newly infected cells. To identify sequence determinants associated with this effect, we performed a functional screening using a randomly mutagenized N library and separated variants into suppressive and non-suppressive populations. Comparative mutation profiling revealed enrichment of nonsynonymous substitutions at a defined subset of amino acid positions in the non-suppressive population. These residues were clustered within specific regions and showed minimal overlap with synonymous variation. Notably, all 23 candidate residues were conserved in a suppressive endogenous bornavirus-like nucleoprotein (itEBLN), whereas several were altered in a non-suppressive human EBLN-1. Together, these findings provide a residue-level map associated with nucleoprotein-mediated suppression of BoDV-1 and suggest that this phenotype is linked to constraints on amino acid variation that are partially reflected in evolutionary conservation.

Introduction

Borna disease virus 1 (BoDV-1) is a non-segmented negative-strand RNA virus that uniquely replicates in the nucleus of infected cells. Viral RNA synthesis is mediated by the viral ribonucleoprotein (vRNP) complex, which consists of the genomic RNA encapsidated by the nucleoprotein (N) and associated with the phosphoprotein (P) and the RNA-dependent RNA polymerase (L) [1–4]. The nucleoprotein plays a central role in this complex by coating the viral genome, thereby protecting it and providing a scaffold for the assembly of the replication machinery. In addition to its structural role, N interacts with other viral components and contributes to the regulation of transcription and replication. Thus, N is an essential factor for the BoDV-1 life cycle and a key determinant of viral RNA synthesis [4–7].

Despite its essential role in viral replication, the nucleoprotein has also been reported to exert an inhibitory effect on BoDV-1 infection when overexpressed. Ectopic expression of N can reduce the efficiency of viral infection, suggesting that an imbalance in the levels of viral ribonucleoprotein components interferes with the viral life cycle [4, 8]. However, the context and stage at which this suppression occurs, as well as the underlying molecular basis, remain poorly understood. These observations indicate that the function of N in the viral life cycle may depend on its expression level and cellular context.

Although the suppressive effect of nucleoprotein overexpression has been observed in BoDV-1 [8], the molecular determinants underlying this phenomenon remain largely undefined. Specifically, it is unclear which regions or specific amino acid residues within N contribute to the reduction of viral replication when expressed in excess. As a result, a systematic identification of candidate residues associated with nucleoprotein-mediated suppression in BoDV-1 is still lacking.

Endogenous bornavirus-like nucleoproteins (EBLNs), which are derived from ancient bornavirus infections, have been identified in the genomes of various vertebrate species [9–12]. Some EBLN sequences retain relatively long open reading frames with similarity to the viral nucleoprotein, and a subset has been reported to be expressed at the RNA or protein level [9, 10, 12]. Notably, an EBLN identified in the thirteen-lined ground squirrel has been shown to suppress BoDV-1 replication when expressed in cultured cells, whereas a human element does not necessarily exhibit such activity [11]. These observations suggest a potential relationship between sequence variation in nucleoprotein-like elements and their impact on viral replication, although the underlying sequence determinants remain unclear.

In this study, we performed a functional screening using a randomly mutagenized nucleoprotein library to identify candidate residues associated with nucleoprotein-mediated suppression of BoDV-1 replication. By separating cell populations based on their ability to maintain or reduce viral replication, we systematically compared mutation profiles between suppressive and non-suppressive conditions. This approach enabled the identification of amino acid residues potentially involved in suppression associated with N overexpression. Furthermore, we examined the distribution of these candidate residues in relation to known functional regions and compared them with endogenous nucleoprotein-like sequences. Through this analysis, we aimed to provide a comprehensive map of candidate residues associated with nucleoprotein-mediated suppression in BoDV-1. This approach enables a residue-level comparison of variants associated with distinct phenotypic outcomes.

Materials and Methods

Cell culture

Human oligodendroglioma (OL) cells and OL cells persistently infected with BoDV-GFP (a recombinant BoDV-1 He/80/Fct strain expressing GFP) [13, 14] were maintained in Dulbecco's modified Eagle's medium (DMEM; Gibco) supplemented with 5% fetal bovine serum (FBS; Gibco), 4.5 g/L D-glucose, and L-glutamine. All cells were maintained at 37°C in a humidified incubator with 5% CO_2. 293T cells were cultured in DMEM supplemented with 10% FBS.

Lentiviral vector production and titration

The coding sequences of BoDV-1 N and mCherry were cloned into the pLVSIN-CMV Puro vector (TaKaRa) to generate lentiviral expression plasmids (pLVSIN-N and pLVSIN-mCherry, respectively). These plasmids were transfected into 293T cells using TransIT-293 Transfection Reagent (TaKaRa) according to the manufacturer's instructions. After 72 h of culture, the supernatants were collected, centrifuged at 2,000 × g for 5 min at 4°C to remove cell debris, and filtered through a 0.45 μm membrane. The clarified supernatants were used as lentiviral vector stocks and stored at -80°C until use. Viral titers were determined using the mCherry-expressing lentiviral vector. Briefly, OL cells (1.0×10^5 cells) were infected

with 50 µL of a 1:1,000 dilution of the viral stock. At 72 h post-infection, mCherry fluorescence intensity (PE-A) and the percentage of mCherry-positive cells were measured using a CytoFLEX flow cytometer (Beckman Coulter, CA, USA). Viral titers were calculated as infectious units per milliliter (IU/mL), defined as the number of cells expressing the transgene per milliliter of viral stock. The titer of the N-expressing lentiviral vector was assumed to be equivalent to that of the mCherry control vector.

Expression of wild-type N in persistently infected cells

BoDV-GFP persistently infected OL cells (8.0×10^5 cells) were seeded in 6-well plates and cultured for 24 h prior to transduction. Cells were then transduced with lentiviral vectors expressing BoDV-1 N at a multiplicity of infection (MOI) of 0.1 (8.0×10^4 IU per well) to minimize multiple integrations per cell. The lentiviral vector pLVSIN-CMV Puro (TaKaRa) contains a puromycin resistance gene, allowing selection of transduced cells. At 48 h post-transduction, 20% of the cells were passaged, and puromycin (5 µg/mL; InvivoGen) was added 24 h later. Cells were then cultured under selection for up to 4 weeks. The proportion of GFP-positive cells was measured by flow cytometry at 14 and 28 days post-transduction. As a non-suppressive control, cells were transduced with a lentiviral vector expressing mCherry and processed in parallel. Uninfected OL cells were used as a GFP-negative control.

Generation of a mutagenized nucleoprotein library

Random mutations were introduced into the BoDV-1 N coding sequence (1113 bp) using the GeneMorph II Random Mutagenesis Kit (Agilent Technologies, CA, USA). Error-prone PCR was performed under conditions designed to introduce approximately 1–16 mutations per kilobase. Reaction conditions and reagent volumes followed the manufacturer's instructions, and primer sequences are listed in Table S1. The resulting PCR products were cloned into the pLVSIN-CMV Puro vector (TaKaRa) using NEBuilder (New England Biolabs). A total of 876 ng of the mutagenized N DNA was used for cloning. The ligation mixture containing 304 ng of plasmid DNA was introduced into MegaX DH10B T1R Electrocompetent Cells (Thermo Fisher Scientific) by electroporation. Transformants were recovered and pooled from approximately 6×10^6 colonies to generate the plasmid library (pLVSIN-N-library). Plasmid DNA was purified using the NucleoBond Extra Maxi kit (TaKaRa) according to the manufacturer's instructions. To assess mutation efficiency, six clones were randomly selected from the library and subjected to Sanger sequencing. Each clone contained 3–6 mutations or small deletions, confirming successful mutagenesis under the selected conditions.

Production and titration of lentiviral N library vectors

The lentiviral N library was produced by transfecting pLVSIN-N-library (16.5 µg) into Lenti-X 293T cells (1.5×10^7 cells) using TransIT-293 Transfection Reagent (TaKaRa) according to the manufacturer's instructions. After 72 h of culture, the supernatants were collected, centrifuged at 2,000 × g for 5 min at 4°C to remove cell debris, and filtered through a 0.45 µm membrane. The clarified supernatants were

used as lentiviral stocks and stored at -80°C until use. Viral titers were determined using a GFP-expressing lentiviral vector processed in parallel. Briefly, OL cells (1.0×10^5 cells) were infected with 250 µL of a $1:10^5$ dilution of the GFP-expressing vector. At 72 h post-infection, GFP-positive cells were quantified by fluorescence microscopy. Viral titers were calculated as infectious units per milliliter (IU/mL), defined as the number of transduced cells per milliliter of viral stock. The titer of the N library vector was assumed to be equivalent to that of the GFP control vector.

Transduction of the mutagenized N library

BoDV-GFP persistently infected OL cells (6.0×10^7 cells) were seeded into two T225 flasks (225 cm^2 each) and cultured for 24 h prior to transduction. Cells were then infected with the lentiviral N library at a MOI of 0.1 (6.0×10^6 IU in total) to minimize multiple integrations per cell. At 48 h post-transduction, cells were expanded by passaging into six flasks, and puromycin (5 µg/mL; InvivoGen) was added 24 h later to select for transduced cells. Cells were maintained under selection for 14 days following transduction.

Isolation of suppressive and non-suppressive cell populations by cell sorting

Because BoDV-GFP is replication-competent, cell sorting was performed using a CytoFLEX SRT cell sorter (Beckman Coulter) installed within a biosafety cabinet. BoDV-GFP–infected OL cells were harvested by trypsinization and resuspended in PBS supplemented with 2% FBS for sorting. Cell populations were first gated based on forward scatter (FSC-A) and side scatter (SSC-A) to define the viable cell population. GFP fluorescence intensity (FITC-A) was then used to classify cells. Based on the fluorescence profiles of control samples (mCherry-expressing cells as a non-suppressive control and uninfected OL cells as a GFP-negative control), a threshold of 2×10^4 was set to distinguish GFP-positive and GFP-negative populations. Cells with fluorescence above this threshold were defined as GFP-positive (non-suppressive, Pos), whereas those below the threshold were defined as GFP-negative (suppressive, Neg). A total number of cells exceeding the library complexity (at least 6.0×10^7 cells) was collected to ensure sufficient coverage of the mutagenized library.

Preparation of sequencing samples

Genomic DNA was extracted from sorted cell populations using the Blood & Cell Culture DNA Maxi Kit (QIAGEN, Hilden, Germany) according to the manufacturer's instructions. The number of cells used for DNA extraction was $7.3 \times 10^7 \pm 2.0 \times 10^7$ for the non-suppressive (Pos) population and $3.0 \times 10^7 \pm 1.1 \times 10^7$ for the suppressive (Neg) population (mean ± SD, n = 3). The corresponding DNA yields were 134 ± 78 µg (Pos) and 44 ± 21 µg (Neg). The target region for sequencing was a 1195 bp fragment comprising the full-length BoDV-1 N coding sequence (1113 bp) flanked by 34 bp at the 5' end and 48 bp at the 3' end. These flanking sequences are derived from the pLVSIN-CMV Puro vector (TaKaRa) and are not present in

the human genome, allowing specific amplification of integrated N library sequences. For each sample, 3.2 μg of genomic DNA was amplified using PrimeSTAR MAX DNA Polymerase (TaKaRa) in 50 μL reaction mixtures. PCR reactions were performed according to the manufacturer's instructions, and a total reaction volume of 800 μL was prepared for each population. Primer sequences are listed in Table S1. PCR products were separated by agarose gel electrophoresis, and fragments corresponding to the expected size (approximately 1195 bp) were excised and purified using the NucleoSpin Gel and PCR Clean-up kit (MACHEREY-NAGEL, Düren, Germany). The purified DNA yields were 9.1 ± 2.9 μg (Pos) and 9.6 ± 1.2 μg (Neg) (mean ± SD, n = 3).

Next-generation sequencing and data processing

Sequencing of the N library DNA was performed using a commercial next-generation sequencing service (Rhelixa Inc., Tokyo, Japan). Genomic DNA samples were fragmented by ultrasonication, followed by end repair and 3' A-tailing. Sequencing adapters were ligated, and libraries were purified using the NEBNext Ultra II DNA Library Prep Kit (New England Biolabs) according to the manufacturer's instructions. Sequencing was conducted on an Illumina NovaSeq X Plus platform using paired-end 150 bp reads (PE150). Approximately 1 G bases per sample were generated, corresponding to ~ 6.7 million reads per sample (3.3 million read pairs). Raw sequencing data were processed using a standard bioinformatics pipeline. Read quality was assessed using FastQC (v0.11.7), and adapter trimming and quality filtering were performed using Trimmomatic (v0.38) [15]. Filtered reads were mapped to the reference sequence using BWA (v0.7.17-r1188) [16]. Duplicate reads were removed using Picard (v2.18.11), and variants were called using FreeBayes (v1.3.4) [17].

Analysis of nucleotide mutation frequencies

Mapped sequencing reads (BAM files) from the suppressive (Neg) and non-suppressive (Pos) populations were analyzed using R (version 4.5.1). Only bases with a mapping quality of at least 20 and base quality of at least 20 were considered. Insertions and deletions were excluded from the analysis. For each nucleotide position, mutation frequency was calculated as the number of observed variant bases divided by the total number of bases sequenced at that position. Mutation frequencies were then averaged across three biological replicates for each population.

Classification of mutation types

All nucleotide substitutions were classified into three categories: synonymous, nonsynonymous (excluding stop codons), and nonsense (stop codon–introducing) mutations. For each nucleotide position, counts of mutations belonging to each category were aggregated. These values were then combined at the amino acid level by summing mutation counts across codons corresponding to the same residue. The proportion of each mutation category at each amino acid position was calculated and averaged across three biological replicates.

Comparative analysis of amino acid substitutions

The frequencies of synonymous and nonsynonymous substitutions were compared between the Neg and Pos populations for each amino acid position of the 370-residue N protein. For nonsynonymous substitutions, the ratio of mutation frequency in the Pos population to that in the Neg population was calculated for each residue. Residues with a ratio of at least 1.9 were defined as Pos-enriched (non-suppressive–associated) mutations, whereas those below this threshold were considered shared between the two populations.

Distribution of Pos-enriched mutations on N and EBLN sequences

The amino acid sequence of the BoDV-1 N protein (accession no. CAC70640.1) was used as the reference sequence, and the sequences of itEBLN (XP_005342534.1) and hsEBLN-1 (NP_001381686.1) were aligned using MAFFT with the --addfragments option. The alignment strategy was automatically selected using the --auto option [18].

Pairwise co-occurrence analysis of amino acid substitutions

To assess whether combinations of mutations were enriched in the Pos population, pairwise co-occurrence analysis was performed using sequencing reads. Among reads containing nucleotide substitutions, those harboring multiple amino acid substitutions within a window of 50 amino acids were extracted. For each mutation pair, the co-occurrence frequency was calculated as the number of reads containing both substitutions divided by the total number of reads. Values were averaged across three biological replicates for each population. Enrichment was evaluated by calculating the $\log_2$ ratio of co-occurrence frequencies between the Pos and Neg populations. To avoid division by zero, a pseudo-count of 1×10^{-6} was added to all values prior to ratio calculation.

Results

Suppression of BoDV-1 infection by nucleoprotein overexpression in persistently infected cells

Previous studies have shown that overexpression of the nucleoprotein can inhibit BoDV-1 infection in the context of newly infected cells [8, 11]; however, its effect on established persistent infection remains less clearly defined. To examine whether overexpression of the nucleoprotein affects virus infection levels in persistently infected cells, BoDV-GFP persistently infected OL cells were transduced with a lentiviral vector expressing wild-type N. Cells were maintained under puromycin selection, and the proportion of GFP-positive cells was monitored over time by flow cytometry. As shown in Fig. 1A, expression of N resulted in a marked decrease in the percentage of GFP-positive cells, reaching 3.26 ± 0.50% at 14 days post-transduction (mean ± SD, n = 3). In contrast, cells transduced with a control vector expressing

mCherry maintained high levels of GFP positivity throughout the observation period. These results indicate that overexpression of N reduces BoDV-1 infection levels in persistently infected cells.

Establishment of a functional screening system using a mutagenized nucleoprotein library

To identify residues associated with this suppressive effect, we established a functional screening system using a randomly mutagenized N library (Fig. 1B). Mutations were introduced into the N coding sequence by error-prone PCR, and the resulting library was cloned into a lentiviral expression vector to generate pLVSIN-N-library with an estimated diversity of 6×10^6 variants. The library was transduced into BoDV-GFP persistently infected OL cells at a low multiplicity of infection (MOI = 0.1), followed by puromycin selection to ensure stable expression.

Cells were cultured for 14 days after transduction, a time point selected based on the kinetics of wild-type N-mediated reduction in GFP-positive cells (Fig. 1A). Flow cytometry analysis revealed heterogeneous GFP expression within the N library-transduced population (Fig. 1C). Specifically, 29.27 ± 4.79% (mean ± SD, n = 3) of cells exhibited reduced GFP expression, whereas 70.73 ± 4.79% retained GFP expression. In contrast, cells expressing mCherry showed uniformly high GFP positivity (98.81 ± 0.29%). Based on GFP intensity, cells were classified into two populations: a suppressive N population (Neg), characterized by reduced GFP expression, and a non-suppressive N population (Pos), in which GFP expression was maintained. These populations were subsequently isolated for comparative mutation profiling.

Global mutation landscape of the nucleoprotein library in suppressive and non-suppressive populations

To characterize the overall mutation landscape of the N library, we analyzed the distribution of nucleotide substitution frequencies in the suppressive (Neg) and non-suppressive (Pos) populations (Fig. 2A). Mutations were detected across all nucleotide positions in both populations, indicating broad coverage of the mutagenized library. Notably, regions with elevated mutation frequencies were observed in both populations, particularly at positions 657–719 and 1070, where mutation frequencies exceeded 0.01. The median mutation frequency was 0.00297 in the Pos population and 0.00362 in the Neg population. These results indicate that the overall mutation distribution is largely comparable between the two populations, with shared regions exhibiting higher mutation rates.

To assess the functional consequences of these mutations, nucleotide substitutions were classified into synonymous, nonsynonymous, and nonsense mutations, and their relative proportions were calculated at the amino acid level (Fig. 2B). In the Pos population, synonymous, nonsynonymous, and nonsense mutations accounted for 26.2 ± 12.6%, 70.7 ± 12.6%, and 3.1 ± 6.3% (mean ± SD, n = 370), respectively. Similarly, in the Neg population, the corresponding proportions were 28.7 ± 13.8%, 69.5 ± 13.3%, and 1.8 ± 3.8%. The overall composition of mutation types was comparable between the two populations,

suggesting that differences in suppressive activity are not attributable to global differences in mutation frequency or mutation type.

Identification of candidate residues associated with loss of suppressive activity

To systematically identify amino acid residues associated with nucleoprotein-mediated suppression, we performed comparative mutation profiling between the suppressive (Neg) and non-suppressive (Pos) populations. Mutation counts were first calculated at each nucleotide position from mapped sequencing reads, and individual substitutions were classified as synonymous, nonsynonymous, or nonsense mutations. For each amino acid position, mutation frequencies were obtained by aggregating mutations of the same type and averaging across three biological replicates for each population.

Comparison of nonsynonymous mutation frequencies between the Pos and Neg populations revealed that most residues exhibited similar mutation levels, as shown by their distribution along the diagonal in the scatter plot (Fig. 3A). However, a subset of residues showed increased mutation frequencies in the Pos population, with log base 1.9 (Pos/Neg) values of at least 1. These residues were highlighted and defined as candidate positions potentially associated with the loss of suppressive activity. When visualized along the amino acid sequence, these candidate residues were not uniformly distributed but instead showed a clustered pattern (Fig. 3B, upper panel). Notably, many of these residues were located within regions corresponding to previously defined functional domains, including the P-binding regions (amino acid positions 51–100 and 131–158). In contrast, synonymous substitutions showed minimal bias between the Pos and Neg populations, with only a limited number of residues exceeding the same threshold (Fig. 3B, lower panel), indicating that the observed differences are primarily associated with amino acid–altering mutations.

To further illustrate the magnitude of these differences, mutation frequencies at the identified candidate residues were plotted individually for the Pos and Neg populations (Fig. 3C). Consistent with the comparative analysis, most of the 23 residues exhibited higher mutation frequencies in the Pos population than in the Neg population. These results indicate that enrichment of nonsynonymous mutations at specific amino acid positions is associated with the non-suppressive phenotype, and that these effects are localized rather than distributed across the entire sequence.

Comparison with EBLN sequences and analysis of mutation co-occurrence patterns

To further examine the potential relevance of the identified candidate residues, we compared their conservation with EBLNs. Previous studies have shown that an EBLN derived from the thirteen-lined ground squirrel (itEBLN) can suppress BoDV-1 replication, whereas a human-derived EBLN-1 (hsEBLN-1) does not exhibit such suppressive activity. Alignment of BoDV-1 N with itEBLN and hsEBLN-1 sequences revealed that all 23 candidate positions enriched for nonsynonymous substitutions in the non-suppressive population retained the wild-type BoDV-1 N residues in itEBLN. In contrast, 5 of these

positions showed amino acid substitutions in hsEBLN-1 (Fig. 4 upper panel). Mutation frequencies at these positions are shown in Fig. 4 (lower panel). As expected from the selection criteria, substitutions at these residues were more frequently observed in the Pos population than in the Neg population. Notably, multiple distinct amino acid substitutions were detected at each position, with a limited number of substitutions accounting for the majority of the observed variants (Table S2). These substitutions included chemically diverse amino acids, indicating that loss of suppression is associated with a range of alternative residue states rather than a single specific substitution. Taken together, these results indicate that residues associated with the non-suppressive phenotype in the screening analysis coincide with positions that are altered in a non-suppressive EBLN sequence, while being conserved in a suppressive EBLN.

To explore whether combinations of mutations contribute to the non-suppressive phenotype, we performed a pairwise co-occurrence analysis of amino acid substitutions using sequencing reads. Mutation pair frequencies were calculated as the proportion of reads containing both substitutions relative to the total number of reads, and values were averaged across three biological replicates for each population. Comparative analysis revealed that certain mutation pairs were more frequently observed in the Pos population than in the Neg population (Table S3). However, these Pos-enriched mutation pairs showed limited overlap with the candidate residues identified by single-residue analysis (Fig. 3) and with residues highlighted by comparison with EBLN sequences (Fig. 4). This suggests that pairwise enrichment captures a distinct feature of the Pos population that is not directly reflected in residue-level mutation enrichment. In addition, regions exhibiting high pairwise mutation frequencies broadly overlapped with regions showing elevated overall mutation rates, indicating that co-occurrence is influenced, at least in part, by local mutation density. Taken together, these results suggest that while specific amino acid substitutions are associated with the non-suppressive phenotype at the single-residue level, subsets of N variants in the Pos population also harbor multiple substitutions within the same molecule. The functional contribution of such mutation combinations remains to be determined.

Discussion

In this study, we investigated the effect of N overexpression on BoDV-1 infection and performed a functional screening to identify amino acid residues associated with this phenotype. We found that overexpression of N reduces infection levels in persistently infected cells (Fig. 1A), extending previous observations made in newly infected systems [8, 11]. Using a randomly mutagenized N library, we separated variants into suppressive and non-suppressive populations and conducted comparative mutation profiling. This analysis identified a set of candidate residues enriched for nonsynonymous substitutions in the non-suppressive population, which were clustered in specific regions of the protein (Fig. 3C). All identified residues were conserved in itEBLN, whereas several were altered in hsEBLN-1 (Fig. 4). In addition, pairwise analysis revealed that subsets of non-suppressive variants harbor multiple substitutions within the same molecule (Table S3). Together, these findings provide a residue-level map associated with nucleoprotein-mediated suppression of BoDV-1.

Previous studies have demonstrated that overexpression of N can inhibit BoDV-1 infection [8, 11], particularly in the context of newly infected cells. In these systems, the suppressive effect is observed relatively rapidly following ectopic expression of N, suggesting interference with early steps of the viral life cycle. In contrast, in the present study using persistently infected cells, a marked reduction in infection levels was observed only after prolonged expression of N (Fig. 1A). This difference in kinetics suggests that the suppressive effect of N is influenced by the cellular context and the stage of infection. One possible explanation is that, in newly infected cells, incoming vRNP complexes may be more susceptible to perturbation by excess N, whereas in persistently infected cells, pre-existing viral complexes are more stable and less immediately affected [4, 7]. Under such conditions, the impact of N overexpression may become apparent only over time, potentially through gradual turnover or reorganization of viral components. Although the precise mechanism remains unclear, these observations indicate that N-mediated suppression is not a simple immediate inhibitory effect but may instead reflect context-dependent modulation of viral processes.

Comparative mutation profiling revealed that nonsynonymous substitutions were enriched at a defined subset of amino acid positions in the non-suppressive population (Fig. 3B upper panel). These residues were not randomly distributed but instead formed clusters within specific regions of the nucleoprotein. In contrast, synonymous substitutions showed minimal bias between the two populations (Fig. 3B lower panel), indicating that the observed differences are primarily associated with amino acid–altering mutations rather than global mutation bias.

Many of the candidate residues were located within or in proximity to previously defined P-binding regions (Fig. 3C) [19]. This observation suggests a potential relationship between these regions and the suppressive effect of N. However, the distribution of candidate residues was not strictly confined to these regions, indicating that the suppressive phenotype cannot be explained solely by perturbation of P-binding sites. These observations suggest that certain positions in N are sensitive to amino acid substitutions, such that changes at these sites are associated with the loss of the suppressive phenotype. Importantly, this does not imply that individual substitutions are sufficient to abolish suppression, but rather that these positions may represent sites where specific amino acid states are required to maintain the suppressive effect. Taken together, the clustering of candidate residues across multiple regions, including but not limited to P-binding domains, suggests that the suppressive effect of N may depend on the integrity of multiple structural or functional elements rather than a single localized interaction.

Comparison with EBLNs provided an additional perspective on the candidate residues identified in this study. All 23 candidate positions enriched for nonsynonymous substitutions in the non-suppressive population retained the wild-type BoDV-1 N residues in itEBLN, whereas several of these positions exhibited amino acid substitutions in hsEBLN-1 (Fig. 4). Given that itEBLN has been reported to suppress BoDV-1 replication, while hsEBLN-1 does not exhibit such activity [11], this pattern is consistent with a potential relationship between residue conservation and the suppressive phenotype. Importantly, the present data do not establish a direct causal link between individual residues and suppressive activity.

However, the complete conservation of these positions in itEBLN, together with their alteration in a non-suppressive EBLN, suggests that these sites may be less tolerant to amino acid substitutions in the context of suppression. In this sense, the overlap between residues identified by functional screening and those conserved in itEBLN is consistent with the idea that the suppressive effect of N is associated with residue-level constraints that are, at least in part, reflected in evolutionary conservation. At the same time, only a subset of these residues differed in hsEBLN-1, indicating that loss of suppression in this sequence cannot be explained by a single substitution or a simple pattern of variation. This observation further supports the notion that the suppressive phenotype may depend on the combined contribution of multiple residues rather than a single determinant.

In addition to residue-level analysis, pairwise co-occurrence analysis revealed that certain mutation pairs were relatively enriched in the non-suppressive population (Table S3). These enriched pairs were not limited to the candidate residues identified by single-residue analysis and instead showed only partial overlap with positions highlighted in Fig. 3B, C and Fig. 4. This suggests that pairwise enrichment captures a distinct aspect of sequence variation that is not fully represented by residue-level mutation frequencies. At the same time, regions exhibiting elevated co-occurrence frequencies broadly overlapped with regions showing higher overall mutation rates, indicating that local mutation density may contribute to the observed patterns. Therefore, while these results suggest that subsets of non-suppressive variants harbor multiple amino acid substitutions within the same molecule, the present data do not distinguish whether such combinations reflect functional interactions between mutations or simply the accumulation of tolerated substitutions. Further studies will be required to clarify the functional significance of these mutation combinations.

Although the present study does not directly address the molecular mechanism underlying nucleoprotein-mediated suppression, the observed patterns provide some clues. The enrichment of nonsynonymous substitutions at specific residues, together with their clustered distribution, suggests that the suppressive effect of N may depend on the integrity of multiple regions within the protein. Given that N functions as a core component of the vRNP complex [4, 6, 7], it is possible that alterations at these positions influence the stability, assembly, or dynamic behavior of viral complexes. The partial overlap of candidate residues with previously defined P-binding regions further raises the possibility that interactions with viral or host factors may be involved. However, the fact that candidate residues are not restricted to these regions indicates that no single interaction can fully account for the observed phenotype. Instead, the suppressive effect of N may reflect a collective property of the protein that emerges from the combined contribution of multiple residues. In this context, one plausible interpretation is that overexpression of N perturbs the balance of viral components, and that this perturbation is modulated by the amino acid composition of N. Changes at specific residues may alter how N participates in this balance, thereby affecting the outcome of viral replication. Further experimental analyses will be required to determine how these residue-level effects are translated into changes in viral behavior.

Several limitations of this study should be noted. First, the present analysis is based on population-level sequencing data, and the functional impact of individual substitutions has not been directly tested. As such, the candidate residues identified here represent positions associated with the non-suppressive phenotype rather than residues whose individual contributions have been experimentally validated. In addition, the analysis of mutation combinations was exploratory, and the current data do not distinguish between functional interactions among mutations and the accumulation of tolerated substitutions.

Despite these limitations, this study provides a systematic and unbiased map of amino acid residues associated with nucleoprotein-mediated suppression of BoDV-1. By integrating functional screening with quantitative mutation profiling, we identified specific positions that are sensitive to amino acid substitutions and linked these positions to phenotypic outcomes. Importantly, the correspondence between residues identified in this screening and their conservation in itEBLN suggests that residue-level constraints observed in an experimental system may, at least in part, reflect constraints acting over evolutionary timescales. Taken together, these findings establish a framework for understanding how variation at specific amino acid positions is associated with differences in the suppressive behavior of N. Rather than attributing the phenotype to a single determinant, our results highlight the importance of distributed residue-level features, providing a basis for interpreting functional variation in bornavirus nucleoproteins.

Declarations

Conflict of interest

The authors declare that they have no competing interests.

Ethical approval

This study did not involve human participants or animal experiments requiring ethical approval.

Author Contribution

K.N., H.H., and M.S. performed experiments and data analysis. K.N. curated the data and prepared the figures. J.K. contributed to data analysis. K.T. contributed to data interpretation and funding acquisition. A.M. supervised the study, secured funding, and wrote the manuscript. All authors reviewed and approved the final manuscript.

Acknowledgements

This work was supported in part by JSPS KAKENHI, grant numbers JP26H02606 (KT), JP24K22131 (KT), JP18K05991 (AM), and JP26K10043 (AM); MEXT KAKENHI, grant number JP25H01316 (AM); AMED, grant numbers JP20wm0325011 (AM) and JP23bm1223017 (KT); JSPS Core-to-Core Program,

JPJSCCA20240006 (AM); 2021 Kaketsuken Research grant (KT); and the Joint Usage/Research Center Program on Institute for Life and Medical Sciences, Kyoto University

Data Availability

The sequencing data generated in this study are available from the corresponding author upon reasonable request. Processed mutation data supporting the findings of this study are provided within the article and its supplementary information files.

References

1. Rubbenstroth D, Briese T, Dürrwald R, Horie M, Hyndman TH, Kuhn JH et al (2021) ICTV Virus Taxonomy Profile: Bornaviridae. J Gen Virol 102(7):001613. 10.1099/jgv.0.001613
2. Matsumoto Y, Hayashi Y, Omori H, Honda T, Daito T, Horie M et al (2012) Bornavirus closely associates and segregates with host chromosomes to ensure persistent intranuclear infection. Cell Host Microbe 11(5):492–503. 10.1016/j.chom.2012.04.009
3. Schneemann A, Schneider PA, Lamb RA, Lipkin WI (1995) The remarkable coding strategy of Borna disease virus: a new member of the nonsegmented negative strand RNA viruses. Virology 210(1):1–8. 10.1006/viro.1995.1311
4. Schneider U, Naegele M, Staeheli P, Schwemmle M (2003) Active borna disease virus polymerase complex requires a distinct nucleoprotein-to-phosphoprotein ratio but no viral X protein. J Virol 77(21):11781–11789. 10.1128/JVI.77.21.11781-11789.2003
5. Rudolph MG, Kraus I, Dickmanns A, Eickmann M, Garten W, Ficner R (2003) Crystal structure of the borna disease virus nucleoprotein. Structure 11(10):1219–1226. 10.1016/j.str.2003.08.011
6. Hock M, Kraus I, Schoehn G, Jamin M, Andrei-Selmer C, Garten W et al (2010) RNA induced polymerization of the Borna disease virus nucleoprotein. Virology 397(1):64–72. 10.1016/j.virol.2009.11.016
7. Schwemmle M, Salvatore M, Shi L, Richt J, Lee CH, Lipkin WI (1998) Interactions of the borna disease virus P, N, and X proteins and their functional implications. J Biol Chem 273(15):9007–9012. 10.1074/jbc.273.15.9007
8. Geib T, Sauder C, Venturelli S, Hässler C, Staeheli P, Schwemmle M (2003) Selective virus resistance conferred by expression of Borna disease virus nucleocapsid components. J Virol 77(7):4283–4290. 10.1128/JVI.77.7.4283-4290.2003
9. Horie M, Honda T, Suzuki Y, Kobayashi Y, Daito T, Oshida T et al (2010) Endogenous non-retroviral RNA virus elements in mammalian genomes. Nature 463(7277):84–87. 10.1038/nature08695
10. Horie M, Kobayashi Y, Suzuki Y, Tomonaga K (2013) Comprehensive analysis of endogenous bornavirus-like elements in eukaryote genomes. Philos Trans R Soc Lond B Biol Sci 368(1626):20120499. 10.1098/rstb.2012.0499

11. Fujino K, Horie M, Honda T, Merriman DK, Tomonaga K (2014) Inhibition of Borna disease virus replication by an endogenous bornavirus-like element in the ground squirrel genome. Proc Natl Acad Sci U S A 111(36):13175–13180. 10.1073/pnas.1407046111
12. Fujino K, Horie M, Kojima S, Shimizu S, Nabekura A, Kobayashi H et al (2021) A human endogenous bornavirus-like nucleoprotein encodes a mitochondrial protein associated with cell viability. J Virol 95(14):e02030–e02020. 10.1128/JVI.02030-20
13. Daito T, Fujino K, Honda T, Matsumoto Y, Watanabe Y, Tomonaga K (2011) A novel Borna disease virus vector system that stably expresses foreign proteins from an intercistronic noncoding region. J Virol 85(23):12170–12178. 10.1128/JVI.05554-11
14. Schneider U, Ackermann A, Staeheli P (2007) A Borna disease virus vector for expression of foreign genes in neurons of rodents. J Virol 81(13):7293–7296. 10.1128/JVI.02467-06
15. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15):2114–2120. 10.1093/bioinformatics/btu170
16. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25(14):1754–1760. 10.1093/bioinformatics/btp324
17. Garrison E, Marth G (2012) Haplotype-based variant detection from short-read sequencing. arXiv:1207.3907 [q-bio.GN]
18. Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780
19. Schneider U, Blechschmidt K, Schwemmle M, Staeheli P (2004) Overlap of interaction domains indicates a central role of the P protein in assembly and regulation of the Borna disease virus polymerase complex. J Biol Chem 279(53):55290–55296. 10.1074/jbc.M408913200

Figures

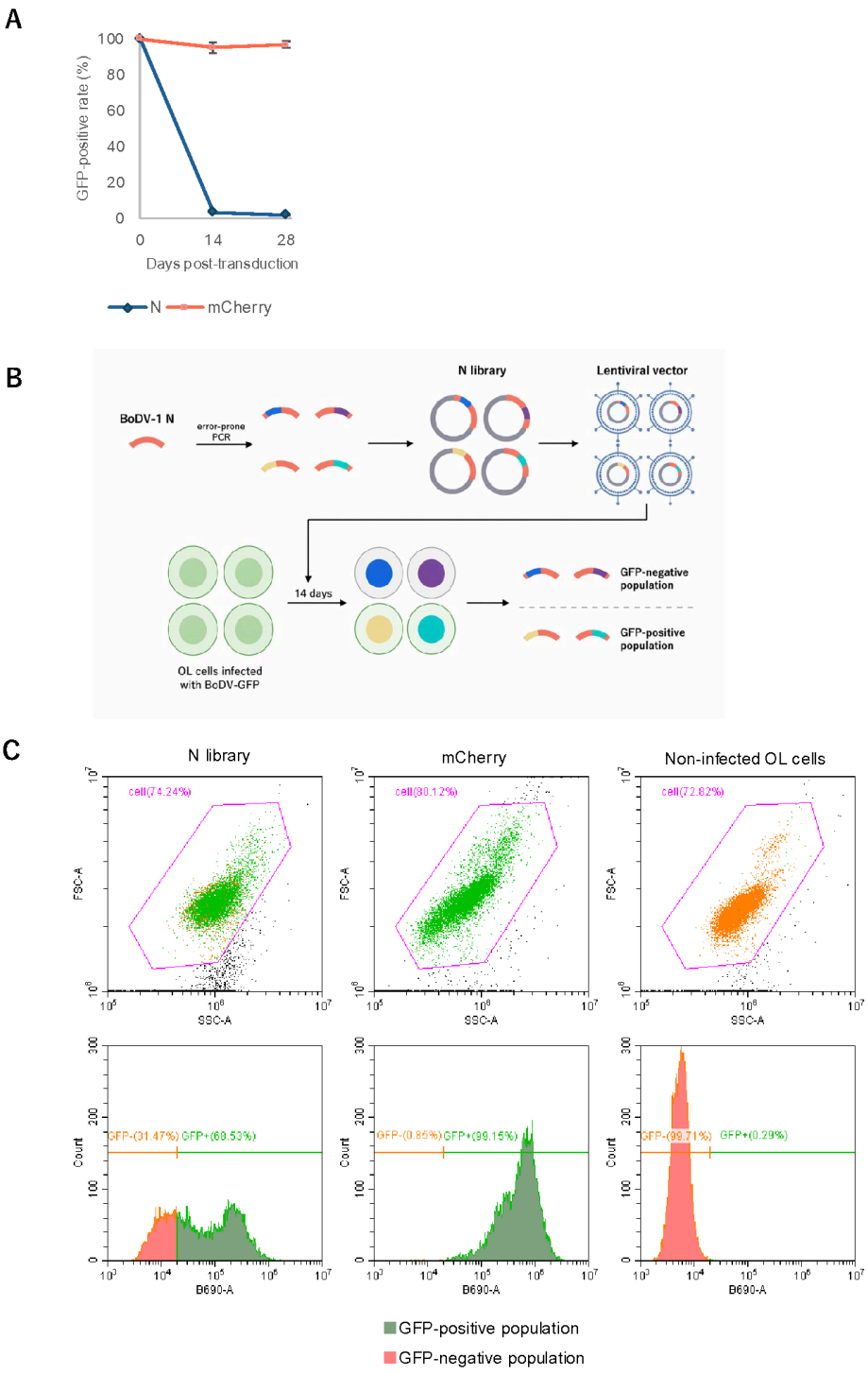


Figure 1

Figure 1

Functional screening identifies nucleoprotein variants associated with suppression of BoDV-1 replication in persistently infected cells.

(A) Time course of BoDV-1 infection levels following overexpression of wild-type N. BoDV-GFP persistently infected OL cells were transduced with a lentiviral vector expressing BoDV-1 nucleoprotein

(N) or a control vector expressing mCherry. Cells were cultured under puromycin selection, and the percentage of GFP-positive cells was measured by flow cytometry at 14 and 28 days post-transduction. Data represent mean ± SD (n = 3).

(B) Schematic representation of the functional screening strategy. A randomly mutagenized N library was generated by error-prone PCR and expressed in BoDV-GFP persistently infected OL cells using a lentiviral system. After 14 days of culture, cells were classified based on GFP expression into suppressive (Neg) and non-suppressive (Pos) populations.

(C) Flow cytometry analysis of cells expressing the N library. Upper panels show gating of the cell population based on forward scatter (FSC-A) and side scatter (SSC-A). Lower panels show GFP fluorescence intensity (B690-A). Cells with reduced GFP expression were defined as the suppressive N population (Neg, orange), whereas cells maintaining GFP expression were defined as the non-suppressive N population (Pos, green). Cells expressing mCherry and uninfected OL cells are shown as controls.

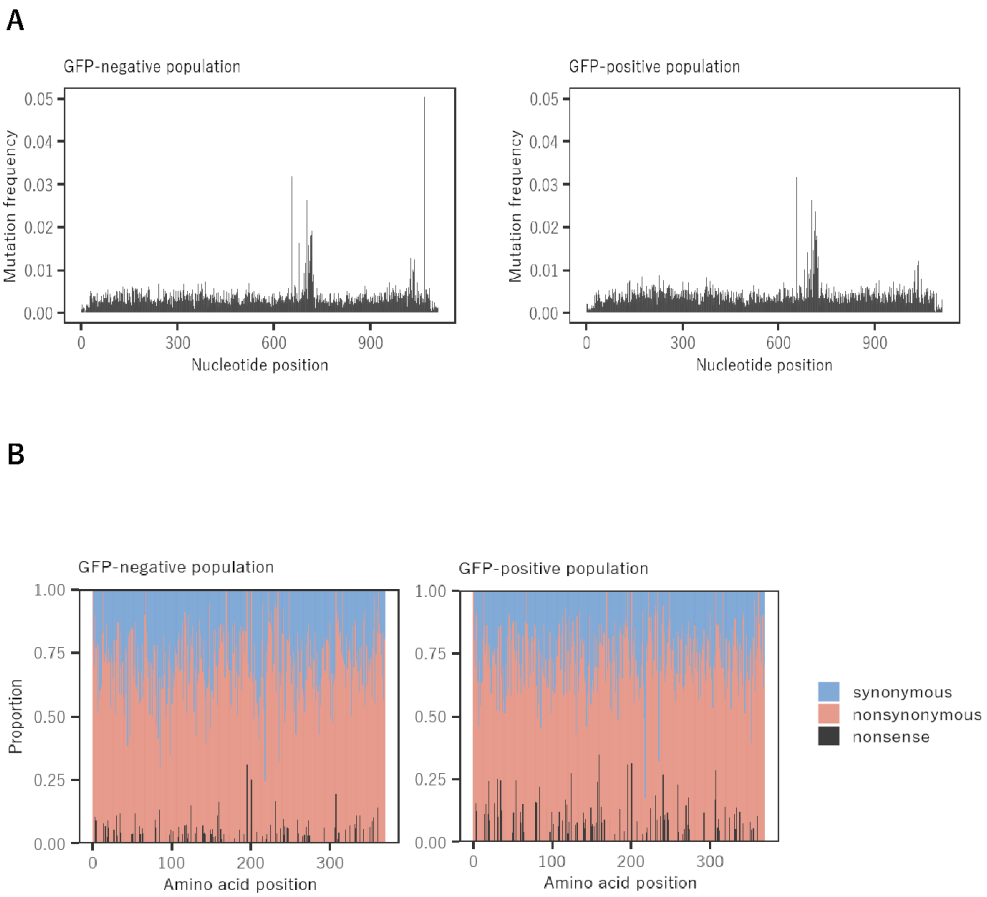


Figure 2

Figure 2

Global mutation profiles of suppressive and non-suppressive N populations.

(A) Distribution of nucleotide substitution frequencies across the N coding sequence in the suppressive (Neg) and non-suppressive (Pos) populations. Mutation frequencies were calculated for each nucleotide

position based on mapped sequencing reads. Regions with elevated mutation frequencies were observed in both populations.

(B) Classification of mutations at the amino acid level. Detected mutations were categorized as synonymous, nonsynonymous, or nonsense substitutions, and their proportions were calculated for each amino acid position. The overall distribution of mutation types was comparable between the suppressive (Neg) and non-suppressive (Pos) populations.

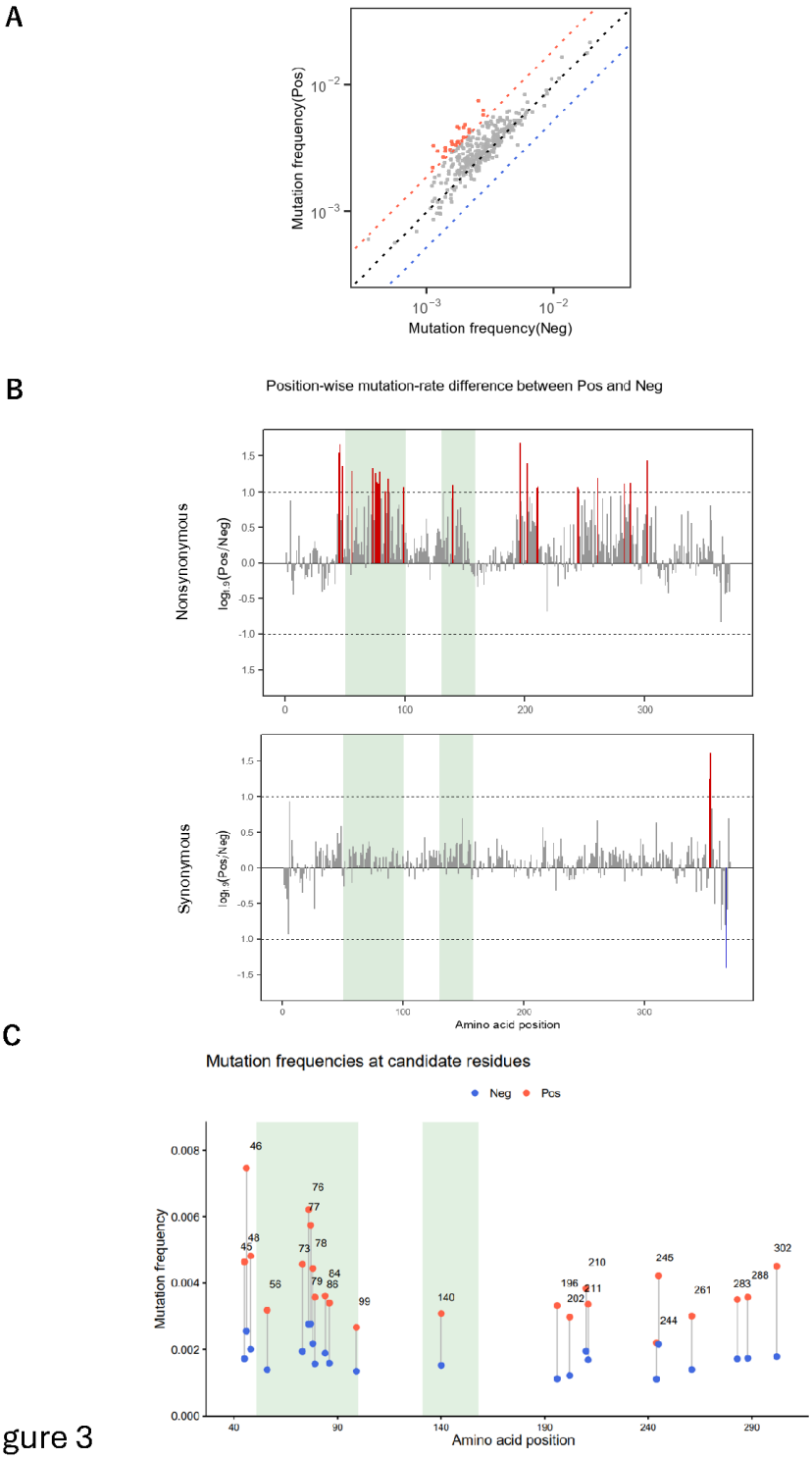


Figure 3

Figure 3

Identification of candidate residues associated with nucleoprotein-mediated suppression based on comparative mutation profiling.

(A) Scatter plot of nonsynonymous mutation frequencies in the suppressive (Neg) and non-suppressive (Pos) populations. Each point represents an amino acid position. Most residues are distributed along the diagonal, indicating comparable mutation frequencies between the two populations. Residues with increased mutation frequencies in the Pos population ($\log_{1.9}$(Pos/Neg) at least 1) are highlighted in red.

(B) Distribution of $\log_{1.9}$(Pos/Neg) values across the amino acid sequence. The upper panel shows nonsynonymous substitutions, and the lower panel shows synonymous substitutions. Residues exceeding the threshold ($\log_{1.9}$(Pos/Neg) values of at least 1 or at most −1) are highlighted. Shaded regions indicate P-binding regions (amino acid positions 51–100 and 131–158).

(C) Mutation frequencies at candidate residues identified in (A). Each point represents the mutation frequency at a given amino acid position in the Pos (red) and Neg (blue) populations. The majority of candidate residues show higher mutation frequencies in the Pos population compared to the Neg population.

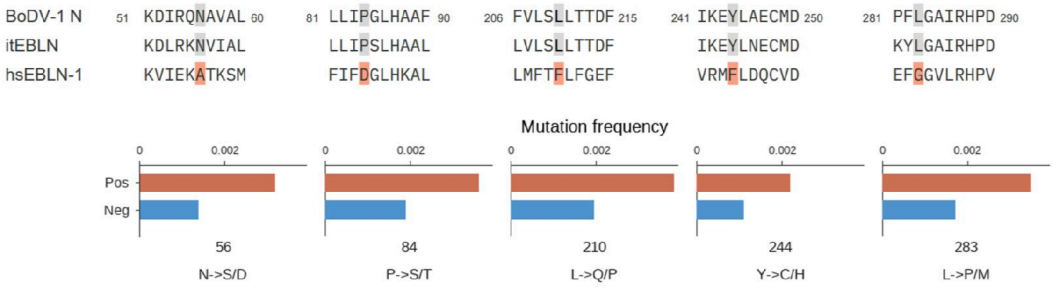


Figure 4

Figure 4

Comparison of selected candidate residues with EBLNs.

Upper panel: Alignment of representative regions of BoDV-1 nucleoprotein (N) with EBLN sequences from the thirteen-lined ground squirrel (itEBLN) and human (hsEBLN-1). Among the 23 candidate

residues identified in Fig. 3, five positions that are not conserved in hsEBLN-1 are shown. Gray and red shading indicates residues in hsEBLN-1 that differ from the corresponding wild-type BoDV-1 N residues.

Lower panel: Mutation frequencies at the five selected candidate residues in the suppressive (Neg) and non-suppressive (Pos) populations. Bars indicate the frequencies of nonsynonymous substitutions at each position. The most frequently observed substitutions are indicated below each position. As expected from the selection criteria, mutation frequencies are higher in the Pos population than in the Neg population at these positions.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS1.xlsx
- TableS2.xlsx
- TableS3final.csv