Genomic and in silico Structural Characterization of Dobrava-Belgrade orthohantavirus Isolate from European Side of Turkey

Mert Erdin (mert.erdin@helsinki.fi)
University of Helsinki, Faculty of Medicine, Medicum, Department of Virology  https://orcid.org/0000-0001-7285-1654

Ceylan Polat
Hacettepe University, Faculty of Medicine, Department of Medical Microbiology

Teemu Smura
University of Helsinki

Sercan Irmak
Balıkesir University, Science and Technology Application and Research Center

Ortac Cetintas
Bülent Ecevit University, Faculty of Arts and Science, Department of Biology

Muhsin Cogal
Bülent Ecevit University, Faculty of Arts and Science, Department of Biology

Faruk Colak
Bülent Ecevit University, Faculty of Arts and Science, Department of Biology

Ahmet Karatas
Ömer Halisdemir University, Faculty of Arts and Sciences, Department of Biology

Mustafa Sozen
Bülent Ecevit University, Faculty of Arts and Science, Department of Biology

Ferhat Matur
Dokuz Eylül University, Faculty of Science, Department of Biology

Tarja Sironen
University of Helsinki  https://orcid.org/0000-0002-2344-2755

Ibrahim Mehmet Ali Oktem
Dokuz Eylül University, Faculty of Medicine, Department of Medical Microbiology

Article

Keywords: genome sequencing, Dobrava-Belgrade orthohantavirus (DOBV), Apodemus flavicollis’ lung tissue, phylogenetic analysis
Genomic and in silico Structural Characterization of Dobrava-Belgrade orthohantavirus Isolate from European Side of Turkey

Mert Erdin1,2, Ceylan Polat3, Teemu Smura1, Sercan Irmak4, Ortac Cetintas5, Muhsin Cogal5, Faruk Colak5, Ahmet Karatas6, Mustafa Sozen7, Ferhat Matur7, Tarja Sironen1* and Ibrahim Mehmet Ali Oktem8*

1. University of Helsinki, Faculty of Medicine, Medicum, Department of Virology, Helsinki, Finland
2. Dokuz Eylül University, Institute of Health Sciences, Department of Microbiology and Clinical Microbiology, Izmir, Turkey
3. Hacettepe University, Faculty of Medicine, Department of Medical Microbiology, Ankara, Turkey
4. Balıkesir University, Science and Technology Application and Research Center, Balıkesir, Turkey
5. Bülent Ecevit University, Faculty of Arts and Science, Department of Biology, Zonguldak, Turkey
6. Ömer Halisdemir University, Faculty of Arts and Sciences, Department of Biology, Niğde, Turkey
7. Dokuz Eylül University, Faculty of Science, Department of Biology, Izmir, Turkey
8. Dokuz Eylül University, Faculty of Medicine, Department of Medical Microbiology, Izmir, Turkey

*These co-authors share equal contributions as having jointly supervised the work.

Corresponding Authors:

mert.erdin@helsinki.fi

ORCID ID:

Mert Erdin: 0000-0001-7285-1654
Ceylan Polat: 0000-0003-1511-4177
Teemu Smura: 0000-0002-9187-3151
Sercan Irmak: 0000-0002-1577-8208
Ortac Cetintas: 0000-0001-7601-2540
Muhsin Cogal: 0000-0002-7749-3566
Faruk Colak: 0000-0003-3985-7864
Ahmet Karatas: 0000-0002-5985-2094
Mustafa Sozen: 0000-0002-1911-605X
Ferhat Matur: 0000-0001-9488-1408
Tarja Sironen: 0000-0002-2344-2755
Ibrahim Mehmet Ali Oktem: 0000-0002-3185-8355
ABSTRACT

Orthohantaviruses are transmitted to humans mostly through small mammals that are the reservoirs of these viruses. Because orthohantaviruses show high genetic variability through geographic regions, the genetic characterization of these viruses with whole genome sequencing is of great importance to clarify the molecular epidemiology and track their genetic changes in the reservoir hosts. We have previously reported the presence of Dobrava-Belgrade orthohantavirus (DOBV) in the Igneada region, Kirkkareli province by showing antibodies against the virus in rodents and by sequencing partial genomes of the virus. Here we report the whole genome sequencing of DOBV Igneada strain directly from Apodemus flavicollis’ lung tissue by next-generation sequencing followed by phylogenetic analyses. In addition, viral protein structures of DOBV Igneada strain were modelled, and in silico prediction analyses of amino acid changes on viral protein function and stability were performed.

The phylogenetic analysis showed a close relation between the DOBV Igneada strain from Turkey and DOBV Ano-Poroia strain from Greece. Similarity plot analysis revealed also similarities between DOBV Igneada strain and other DOBV strains from the Balkans such as Greece, Croatia, and Slovenia. Additionally, in silico prediction suggested that G318E, Y322H, and S324P mutations on Gn glycoprotein are deleterious, and all amino acid changes decrease the stability of both Gn and Gc glycoproteins.

In conclusion, full orthohantaviral genomes can be obtained directly from rodent lung tissues allowing detailed genetic and structural analyses of orthohantaviruses. The DOBV Igneada strain shows great similarity to the prototype Ano-Poroia strain, yet it was predicted that DOBV Igneada strain may have some changes on its pathogenicity and its structure warranting further research.

INTRODUCTION

Orthohantaviruses are enveloped, negative-sense, single stranded, and tri-segmented RNA viruses and members of the Hantaviridae family. While rodents and small mammals are the main reservoirs for these viruses, humans are accidental hosts. Humans can be infected via inhalation of viral particles from the secretions of reservoir animals or via direct contact with the reservoirs [1].

The distribution of the orthohantaviruses depends on the geographic distribution of their reservoir hosts, and orthohantaviruses are separated into two main groups: New World and Old World Orthohantaviruses. The New World Orthohantaviruses are mainly distributed through North and South America, and they cause “Hantavirus Pulmonary Syndrome” (HPS) in humans [2]. The Old World Orthohantaviruses are mostly distributed in Europe, Asia, and Africa, and they are the causative agents for “Hemorrhagic Fever with Renal Syndrome” (HFRS) in human infections. Additionally, a member of the Old World Orthohantaviruses, Puumala orthohantavirus (PUUV), causes a milder type infection in humans, which is called “Nephropathia Epidemica” (NE), [2].

Some rodent species identified as reservoirs for family Hantaviridae, genus Orthohantavirus, species DOBV, PUUV, Tula orthohantavirus (TULV), and Seoul orthohantavirus (SEOV), are widely present in Turkey, including Myodes glareolus (PUUV), Apodemus flavicollis (DOBV), Apodemus agrarius (DOBV), Microtus arvalis (TULV), and Rattus norvegicus (SEOV) [3-6]. The initial detection of orthohantaviruses in Turkey was in Zonguldak and around provinces in 2009, and later, it also detected in Giresun province and in the western Black
Sea region, European cost of Istanbul, Kirklareli province and Erzurum province with molecular and serological techniques from both humans and rodents [7-12].

The Kirklareli province became an important region in point of orthohantaviruses after the detection of positive rodents in our previous study [11]. Orthohantavirus positivity was initially reported serologically and then confirmed as DOBV by sequencing partial S-, M- and L- segments followed by phylogenetic analysis that showed the genetic relation between DOBV Igneada strains and other DOBV strains [11]. The study suggested that the DOBV Igneada strains were similar to strains from Balkans. This particular region can be seen as a barrier between Europe and Asia due to its various geographical properties and its location. In addition, the reservoir host *Apodemus flavicollis* is distributed in both European side and Anatolia (Asian side of Turkey), and it shows different genetic properties on these sides [26-27]. Therefore, orthohantaviruses that exist in this region, might have genetic properties of viruses from both of these continents [11, 25]. Thus, there is great importance to obtain whole genome sequences of DOBV from this region as these might help us to understand the genetic relation between orthohantaviruses from both Asia and Europe. Furthermore, very limited data is available on DOBV whole genomes and additional sequences could provide detailed information about genetic similarities and differences between DOBV strains as well as inform of structural and functional properties of different strains.

In the past, it was challenging to detect whole genome of viral zoonotic agents directly from rodents or small mammals which are carriers for many different and important zoonosis agents, yet it has become easy to obtain whole genome sequences with the second-generation sequencing systems such as Illumina systems over the past decade [13]. On these days, the biggest challenge for the sequencing of such viruses is the low abundance of viral particles in their host tissues [13]. Additionally, RNA viruses have non-coding regions (NCR) on their 5’ and 3’ ends of their genomes, and these NCRs can’t be covered if cDNA synthesis is performed before the library preparation for next generation sequencing [13-14]. Thus, RNA library preparation kits with specific modifications ought to be used in order to obtain whole genome sequences of RNA viruses. Here, we report the discovery of the whole genome sequence of DOBV Igneada strain from lung tissue for the first time from Turkey as a natural representative of the virus that wasn’t taken to the viral cell culture in order to increase the viral load. Also, we report detailed genetic characterization and in silico prediction of structural properties of DOBV Igneada strain.
MATERIALS AND METHODS

Samples

In 2009, 73 rodents were trapped via Sherman-type traps and DOBV positivity was shown in 8 *Apodemus flavicollis* mice (confirmed with Sanger sequencing of cytochrome-b gene) [11]. In this study, one archived lung tissue sample out of 8 positive samples was chosen as a representative and taken for obtaining the whole genome sequence of DOBV Igneada strain. RNA extraction was performed from the lung tissue via TRIzol Reagent (Invitrogen, Carlsbad, CA, USA), followed by library preparation as described below.

Library Preparation and Sequencing

Host rRNA was removed from the total RNA by NEBNext rRNA depletion kit (Human/Mouse/Rat), and NEBNext Ultra II RNA library prep kit for Illumina (New England Biolabs, Ipswich, Massachusetts, USA) was used for library preparation. Library preparation was performed according to manufacturer’s guide. The libraries were sequenced in Illumina MiSeq system with the MiSeq Reagent Kit v3 (600-cycles) (Illumina, San Diego, CA, USA).

In order to clean host sequences from the raw data, a dataset was created from the NCBI GenBank with the sequences of host rodent *Apodemus flavicollis*. After that, the sequences that belonged to the host were cleaned from the raw data using mirabait tool on MIRA 5.0 program [24]. After cleaning, the remaining sequence reads were *de novo* assembled with MIRA 5.0 program. Later, *de novo* assembled contig was used as reference in UGENE tool for remapping to confirm the accuracy of *de novo* assembly.

The assembled sequences were checked on BLAST tool from NCBI GenBank and compared with both nucleotide database and protein database. Consequently, all clinically important viruses that were existing at reservoir rodents, were screened and it was confirmed that DOBV sequences had the highest number and obtained accurately.

Finally, a dataset which contained all available complete DOBV sequences, was created from the NCBI GenBank database. The obtained sequence of DOBV Igneada strain was added to that dataset and later, this dataset was aligned with ClustalW in MEGA X tool to see the gaps on the nucleotide sequence of DOBV Igneada strain. New primers were designed according to the gaps (Supplementary Table 1). The extracted RNA was reverse-transcribed with random hexamers by RevertAid Premium reverse transcriptase (Thermo Fisher Scientific, Waltham, MA, USA). Synthesized cDNA was used to perform PCR with designed primers to fill the gaps remaining after the MiSeq sequencing. Amplified PCR products were later sequenced with Sanger sequencing to fill the gaps. In the end, the gaps on the nucleotide sequences of DOBV Igneada strain were filled with the sequences that were obtained from Sanger sequencing.

Phylogeny and Genetic Analysis

Construction of phylogenetic trees were performed in MEGA X tool. Aligned datasets were analyzed with “find the best DNA/Protein model” for the model that fits for the sequence datasets to construct phylogenetic trees, in MEGA X. Later, phylogenetic trees were constructed in 1000 bootstraps. Constructed trees were viewed and edited in FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).
After constructing phylogenetic trees, open reading frames (ORFs) were determined with NCBI ORFfinder tool (https://www.ncbi.nlm.nih.gov/orffinder/). Default settings were used to determine the ORFs on each three segments. After determining the ORFs, percentage similarities of both nucleotide and protein sequences between DOBV Igneada strain and other DOBV strains were calculated. In order to perform that, MEGA X tool was used by computing pairwise distances.

Finally, SimPlot analysis was performed as a final phylogenetic analysis for each three segments separately. DOBV Igneada strain’s nucleotide sequences of all three segments were compared with other DOBV strains which showed close relation with Igneada strain, and later, the similarities and differences were shown on a graph position-by-position on nucleotide sequences with SimPlot analysis. SimPlot analysis was done by SimPlot v.3.5.1. tool, and the parameters were: gap strip was open, bootstrap was 1000, model was F84 Maximum Likelihood, T/t was 2.0 [32].

**Protein Modelling and Analysis**

The protein sequences of DOBV Igneada and Ano-poroia strains were uploaded to SWISS-MODEL tool (https://swissmodel.expasy.org/) for 3D modelling [28-29]. First, the protein sequences were analyzed for finding templates. After the selection of templates, 3D protein models were constructed. The models with the best statistical scores were chosen and compared in SWISS-MODEL tool. In the compare section, two models that are one from each strain, aligned to each other to see the amino acid changes in both strains. Modelled proteins were visualized in RasMol v.2.7.5.2. Amino acid changes on the protein models were selected and spacefilled with different colors in order to make them visible. Finally, protein models were compared with each other to see whether there is any conformational change in their structure.

In addition, in silico prediction analysis for function of glycoproteins was conducted using PROVEAN v1.1.3 (provean.jcvi.org/index.php) [30]. Amino acid substitutions on the glycoproteins may not affect the conformational structure, but they might affect the function of proteins. During the analysis, results were considered as deleterious or neutral. The cut off value for the determination of protein change affection for being neutral or deleterious, was set to −2.282 (Deleterious < -2.282 > Neutral) [23]. In order to perform the predictive analysis of protein stability, I-Mutant v3.0 (gpcr2.biocomp.unibio.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi) software was used [31]. Default settings were set 7.0 and 25 for pH and temperature (°C), respectively. Predictive stability analysis results were obtained in ternary classification (Large Decrease -0.5=<DDG=<0.5 Large Increase).
RESULTS

The complete coding sequence of DOBV Igneada strain was obtained with next generation sequencing, yet there are some missing sequences at the ends of the segments. The details about the segments were shown at Table 1.

Table 1. Genome details of DOBV Igneada strain.

| Segment | Length (nt) | Number of missing nucleotides on the 5’ end | Number of missing nucleotides on the 3’ end | Accession Number on NCBI GenBank database |
|---------|-------------|---------------------------------------------|---------------------------------------------|------------------------------------------|
| S       | 1633        | 32                                          | 9                                           | MW055919                                 |
| M       | 3639        | 15                                          | -                                           | MW055918                                 |
| L       | 6328        | 180                                         | 24                                          | MW055917                                 |

ORFs were determined on each segment’s nucleotide sequences. The details about the ORFs on each segment were shown at Table 2.

Table 2. ORF details from each segment of DOBV Igneada strain.

| Segment | Polarity | Start (nucleotide) | Stop (nucleotide) | Length (amino acid) | Product                                      |
|---------|----------|--------------------|-------------------|---------------------|----------------------------------------------|
| S       | +        | 5                  | 1294              | 429                 | Nucleocapsid protein                         |
| M       | +        | 26                 | 3433              | 1135                | Gn and Gc glycoprotein precursors            |
| L       | +        | 77                 | 6313              | 2078                | RNA dependent RNA polymerase enzyme          |

Phylogenetic trees were constructed for each three segments separately, and these phylogenetic trees were shown at Figure 1, Figure 2, and Figure 3 for S-, M-, and L- segments, respectively.
Figure 1. Figure represents the phylogenetic tree of S segment of DOBV strains. Phylogenetic tree was constructed with 1293 nucleotides for each sequence. Analysis parameters: model was “T92 + G + I”, bootstrap was 1000, and Maximum Likelihood tree was used. Seoul orthohantavirus was used as an outgroup in the analysis.

Figure 2. Figure represents the phylogenetic tree of M segment of DOBV strains. The phylogenetic tree was constructed with 3407 nucleotides for each sequence. Analysis parameters: model was “GTR + G + I”, bootstrap was 1000, and Maximum Likelihood tree was used. Seoul orthohantavirus was used as an outgroup in the analysis.
Figure 3. Figure represents the phylogenetic tree of L segment of DOBV strains. The phylogenetic tree was constructed with 6310 nucleotides for each sequence. Analysis parameters: model was “GTR + G + I”, bootstrap was 1000, and Maximum Likelihood tree was used. Seoul orthohantavirus was used as an outgroup in the analysis.

After ORFs were determined and phylogenetic trees were constructed, percentage identities were calculated with both nucleotide and protein sequences by MAFFT. These percentage identities for DOBV Igneada strain were shown at Table 3.

Table 3. The details of both nucleotide and amino acid percentage identities of DOBV Igneada strain.

| Segment | DOBV Igneada strain that has the highest percentage identity | Percentage identity of nucleotide sequences | Percentage identity of amino acid sequences |
|---------|-------------------------------------------------------------|---------------------------------------------|-------------------------------------------|
| S       | Greece / Ano-Poroia strain                                  | 96,26%                                      | 99,77%                                    |
| M       | Greece / Ano-Poroia strain                                  | 93,64%                                      | 98,50%                                    |
| L       | Greece / Ano-Poroia strain                                  | 93,71%                                      | 99,09%                                    |

The DOBV Igneada strain showed high similarity with the DOBV Ano-Poroia strain (Greece) in both phylogenetic trees and percentage identities. After all these phylogenetic analysis, SimPlot analysis was also done to observe nucleotide differences position-by-position. The S segment of the DOBV Igneada strain showed great similarity with strains from Greece, Croatia and Slovenia (Supplementary Figure 1). The M and L segment of the DOBV Igneada strain presented high similarity with strains from Greece and Slovenia (Supplementary Figure 2 and Supplementary Figure 3).
Protein models were constructed based on the sequences of Igneada and Ano-Poroia strains. The main reason for the selection of Ano-Poroia strain is the close phylogenetic relation between two strains. Even though there is a high similarity for both amino acid and nucleotide sequences, there are still some amino acid changes between two strains. Modelling of the nucleocapsid protein from S segment showed that there is only one amino acid change between two strains, yet this change doesn’t affect the 3D conformational structure (Figure 4).

Modelling of Gn glycoproteins from both strains illustrated that there are 10 amino acid changes in total while there are only three amino acid changes on the Gc glycoprotein. However, none of them are likely to affect the 3D conformational structures (Figure 5 and 6).

**Figure 4.** A 3D protein model of the nucleocapsid protein of both Igneada and Ano-Poroia strains. At the position 112, there is one amino acid change, identified also on the figure. Template protein ID: 6i2n.1.A. A) Protein model of Igneada strain. GMQE score: 0.69; QMEAN score: -1.29. B) Protein model of Ano-Poroia strain. GMQE score: 0.69; QMEAN score: -1.21.
Figure 5. A 3D protein model of the Gn glycoprotein from both Igneada and Ano-Poroia strains. Template protein ID: 6y6p.1.A. A) Gn glycoprotein model of Igneada strain. GMQE score: 0.38; QMEAN score: -1.71. B) Gn glycoprotein model of Ano-Poroia strain. GMQE score: 0.37; QMEAN score: -2.02.

Figure 6. A 3D protein model of the Gc glycoprotein from both Igneada and Ano-Poroia strains. Template protein ID: 6z06.1.C. A) Gc glycoprotein model of Igneada strain. GMQE score: 0.72; QMEAN score: -1.76. B) Gc glycoprotein model of Ano-Poroia strain. GMQE score: 0.72; QMEAN score: -1.57.
According to *in silico* prediction analysis results of amino acid substitution effect on biological function of a protein, while 3 amino acid substitutions (G138E, Y322H, and S324P), on the Gn glycoprotein were found to be deleterious, the rest of the amino acid substitutions on both Gn and Gc glycoproteins were found neutral (Figure 7).

**Figure 7.** The results of the *in-silico* prediction of the amino acid substitution effect on biological function of a protein by PROVEAN. The cut off value for deleterious substitutions is −2.282. **A)** Bar chart shows the results of DOBV Igneada strain Gn glycoprotein. **B)** Bar chart illustrates the results of DOBV Igneada strain Gc glycoprotein.
Furthermore, the results of the *in-silico* prediction of both glycoproteins’ stability analysis showed that all of the amino acid substitutions on both Gn and Gc glycoproteins are largely decreasing the stability of glycoproteins (Figure 8).

**Figure 8.** Bar charts illustrate the results of the *in-silico* prediction of the stability of DOBV Igneada strain glycoproteins. According to ternary classification, DDG (Free Energy Change) values lower than −0.5 were considered as large decrease, while DDG values higher than 0.5 were considered as large increase. On the other hand, DDG values between −0.5 and 0.5 were considered as neutral. Reliability index is a scale between 1 to 10, which 1 is the lowest and 10 is the highest. **A)** Bar chart shows the results of DOBV Igneada strain’s Gn glycoprotein. **B)** Bar chart represents the results of DOBV Igneada strain’s Gc glycoprotein.
In addition, genetic and geographical distances were analyzed to search for a correlation between them. The DOBV-Ignea strain and Kirklareli province were set as reference points for genetic distance and geographical distance, respectively. A slight positive correlation between the genetic and geographical distance was observed (Figure 9).

Figure 9. Charts represent the correlation between genetic distance and geographical distance. A) The chart shows the correlation results for DOBV’s S segment. B) The chart illustrates the correlation results for DOBV’s M segment. C) The chart represents the correlation results for DOBV’s L segment.
Lastly, positions that have amino acid changes on the M segment of the DOBV-Igneada strain were compared with other DOBV strains from different regions of Europe. It was observed that while deleterious changes at the positions 318 and 324 were unique to DOBV-Igneada strain, amino acid change at the position 322 was shared with the DOBV-Lipetsk strain (Figure 10).

| Country       | Strain          | Region        | AA Change | AA Change | AA Change | AA Change | AA Change | AA Change |
|--------------|------------------|---------------|-----------|-----------|-----------|-----------|-----------|-----------|
| 1. Turkey     | DOBV-Igneada     | Southern Europe | V         | E         | X         | K         | R         | V         |
| 2. Greece     | DOBV/As-Fernalla/AB9/1999/342554.1 | Southern Europe | A         | G         | Y         | S         | R         | I         |
| 3. Slovenia   | DOBV/Slo/A49R/ADP2/2934 | Central Europe | V         | V         | Y         | S         | R         | I         |
| 4. Slovenia   | DOBV/Slo/Stansforsu/SHL/AL598016.1 | Central Europe | V         | V         | Y         | S         | R         | I         |
| 5. Slovenia   | DOBV/Slo/A132777.1 | Central Europe | V         | V         | Y         | S         | R         | I         |
| 6. Slovenia   | DOBV/Slo/A132777.1 | Central Europe | V         | V         | Y         | S         | R         | I         |
| 7. Austria    | DOBV/AT/KS18/1012_QK053232.1 | Central Europe | V         | V         | Y         | S         | R         | I         |
| 8. Austria    | DOBV/AT/KS12/1827_QK053231.1 | Central Europe | V         | V         | Y         | S         | R         | I         |
| 9. Germany    | DOBV/Ger/AA/1993297.1 | Central Europe | V         | V         | Y         | S         | R         | I         |
| 10. Germany   | DOBV/GER/04/118A_ADA68852.1 | Central Europe | V         | V         | Y         | S         | R         | I         |
| 11. Germany   | DOBV/GER/04/118A_ADA68851.1 | Central Europe | V         | V         | Y         | S         | R         | I         |
| 12. Germany   | DOBV/GER/04/118A_ADA68850.1 | Central Europe | V         | V         | Y         | S         | R         | I         |
| 13. Germany   | DOBV/GER/04/118A_ADA68850.1 | Central Europe | V         | V         | Y         | S         | R         | I         |
| 14. Germany   | DOBV/GER/04/118A_ADA68850.1 | Central Europe | V         | V         | Y         | S         | R         | I         |
| 15. Russia    | DOBV/118A/Bulgarian_01_ABY64907.1 | Eastern Europe | V         | V         | Y         | S         | R         | I         |
| 16. Russia    | DOBV/118A/Bulgarian_01_ABY64907.1 | Eastern Europe | V         | V         | Y         | S         | R         | I         |
| 17. Estonia   | DOBV/Saarbrueck/000V_CA000002.2 | Northern Europe | V         | V         | Y         | S         | R         | I         |

**Figure 10.** The figure shows the amino acid diversity between different DOBV strains from different regions of Europe at the positions of DOBV-Igneada strain specific amino acid changes. Arrows at the figure represents the positions that are deleterious amino acid changes.

**DISCUSSION**

An HFRS outbreak occurred in Bartın and Zonguldak provinces in Turkey in 2009, and that was the first time that DOBV was detected in Turkey with a field study to determine the causative virus in those outbreaks [15]. A year later, an individual from Istanbul developed HFRS, and was diagnosed with DOBV infection [10]. Another HFRS outbreak occurred in 2010 in the Giresun province and in the western Black Sea region, and the causative viruses were determined as DOBV and PUUV in this outbreak from human samples by serological assays [8]. After these outbreaks and diagnoses, a field study was performed by collecting *Apodemus flavicollis* mice from the Kirklareli province by our team, and the circulation of DOBV was confirmed in this species of mice both serologically and molecularly [11]. Studies done to date have shown that DOBV causes severe form of HFRS in humans, and the mortality rate is higher in DOBV infections than the other HFRS causing orthohantaviruses. The mortality rate varies between 5 to 15% establishing DOBV as an important virus to study in Turkey [15, 16-18]. This variation on mortality rate is caused by the pathogenicity of different DOBV genotypes that varies greatly and hence knowing the sequence of the local DOBV variant is of great importance [33].

All the outbreaks caused by DOBV in Turkey were reported from the Black Sea region, in where the reservoir host *Apodemus* spp. is widely distributed [5]. Also, this Black Sea region of Turkey and specifically the Balkan countries are endemic regions for DOBV infections with an average number of 10,000 – 12,000 cases of HFRS diagnosed every year. The causative viruses are mostly DOBV and PUUV in this HFRS cases [19]. Thus,
it is important to obtain more whole-genome sequences of DOBV which is common both in Balkan countries and Turkey, and determine the genetic characteristics of the virus particularly as so few complete genomes of DOBV are yet available in public databases.

The whole genome sequence of DOBV Ano-Poroia strain from Greece is the reference sequence for DOBV representing the Balkan Region and a neighbor for Turkey in European Region [20]. The DOBV Igneada strain reported here showed high identity to DOBV Ano-Poroia strain on both nucleotide and protein sequences. A close relation was confirmed between these two strains also with both phylogenetic and SimPlot analyses. However, there are only 10 whole-genome sequences of DOBV on GenBank databases are available, and this limits the analysis specifically as no complete genomes are available from Bulgaria close to the Kirklareli province. Furthermore, as the Igneada strain is so far the only whole genome DOBV sequence from Turkey, the question still remains on its relation to the DOBV that caused an outbreak in the Black Sea Region of Turkey. In addition to this, most of the whole-genome sequences on the GenBank database are sequences from viral cell culture or a clone, neither of which represent the wild type of DOBV [34-35]. As a consequence, the data used in this study were very limited, and more whole-genome sequence data obtained directly from tissue samples are needed for DOBV.

The prediction of viral structural protein models was performed in order to identify increased emerging potential of DOBV Igneada strain, and also, to provide preliminary data for further study such as determining the active sites of the proteins, or determination of epitope regions of viral glycoproteins. The amino acid changes on the M- segment might change the pathogenicity, or affect the pathogenesis and diagnosis [21-22]. In this study, even though some amino acid changes were identified on the protein sequences, none of them really affected the 3D conformational structure (Figure 4, Figure 5, Figure 6). We also acknowledge that the viral protein prediction using only bioinformatics tools is challenging, yet these models are the first models that belong to DOBV and they might be guiding models for further and better study such as de novo viral protein modelling.

PROVEAN software is a tool that gives prediction on the effect of different amino acid substitutions [30]. The substitution might be deleterious meaning that such amino acid substitution may affect the function of a protein. In terms of virology, deleterious predictions might be affecting the pathogenicity of a virus. Even though the default setting of PROVEAN software for threshold value is -2.5 for determining whether an amino acid change is deleterious or not, in this study, the threshold value was set to -2.282, since this value gives the maximum of minimized values of both specificity and sensitivity. When -2.282 was set as a threshold value for single amino acid substitution, both specificity and sensitivity have the highest of minimized values with 79.11 and 78.39, respectively [23]. According to PROVEAN predictions, the substitutions G318E, Y322H, and S324P are deleterious and somehow, they affect the function of Gn glycoprotein and hence the pathogenicity of this virus might be affected due to these substitutions. Even though there are some amino acid substitutions, it seems that DOBV Igneada strain is likely similar to DOBV Ano-Poroia strain with some additional changes on its structure. These results are however only predictions that need to be then confirmed by in vitro or in vivo assays.

I-Mutant is a software that predicts both the direction (the ΔΔG sign) of the protein stability changes and the ΔΔG associated values. The tools prediction accuracy is 80 and 77% for structure-base and protein sequence-base prediction, respectively [31]. The correlation between predicted values and experimental database is 0.71 and 0.62 for structure-base and sequence-base prediction, respectively [31]. According to I-Mutant tool, all the amino
acid substitutions decreased the stability of the glycoproteins, although the reliability index of some predictions were either really low or zero. The question remains if this stability decreasing amino acid substitutions cause the viral particle lifetime in the environment to be shorter, again requiring experimentation.

CONCLUSION

In conclusion, the DOBV-Igneada strain is a close relative of DOBV-Ano-Poroia strain in all three segments with unique deleterious amino acid substitutions, and with glycoproteins that have decreased stability, predictively. Further \textit{in vitro} and \textit{in vivo} experiments are needed to support these preliminary predictions.

Acknowledgments

This study was performed in the laboratories of the University of Helsinki, Faculty of Medicine, Department of Virology, Viral Zoonosis Research Unit. All the authors are grateful to Olli Vapalahti for opening the viral zoonosis laboratories to use as a leader of Viral Zoonosis Research Unit.

Funding

Mert Erdin was funded with Research and Training Grant in order to perform this study by Federation of European Microbiological Societies (FEMS).

Ethical approval

The design of this study, animal handling and experimental procedures were approved by the Dokuz Eylül University Local Ethical Committee of Animal Experiments (No: 03/2020).

REFERENCES

1. Sola-Riera C, Gupta S, Ljunggren HG, Klingström J (2019) Orthohantaviruses belonging to three phylogroups all inhibit apoptosis in infected target cells. Sci. Rep. 9:834. \url{https://doi.org/10.1038/s41598-018-37446-1}

2. Vaheri A, Strandin T, Hepojoki J, Sironen T, Henttonen H, Makela S, Mustonen J (2013) Uncovering mysteries of hantavirus infections. Nat. Rev. Microbiol. 11:539-550. \url{https://doi.org/10.1038/nrmicro3066}

3. Kefelioglu H, Tez C, Gunduz I (2003) The taxonomy and distribution of Apodemus agrarius (Pallas, 1771) (Mammalia: Rodentia) in the European part of Turkey. Turk J Zool 27:141-6

4. Yigit N, Colak E, Sozen M, Karatas A (2006) Rodents of Türkiye: Türkiye Kemiricileri. Meteksan Press Ankara.

5. Krystufek B, Vohralik V (2007) Distribution of field mice (Apodemus) (Mammalia: Rodentia) in Anatolia. Zool Middle East 42:25-36. \url{https://doi.org/10.1080/09397140.2007.10638243}

6. Krystufek B, Vohralik V (2009) Mammals of Turkey and Cyprus. Rodentia II: Cricetinae, Muridae, Spalacidae, Calomyscidae, Capromyidae, Hysteridae, Castoridae. Koper, Slovenia.

7. Ertek M, Buzgan T (2009) An outbreak caused by hantavirus in the Black Sea Region of Turkey. Eurosurveillance 14:1-2.
8. Kaya S, Yilmaz G, Erensoy S, Yager Çaglayık D, Uyar Y, Koksal I (2010) Hantavirus infection: two case reports from a province in the Eastern Black Sea Region, Turkey. Mikrobiyol. Bul. 44(3):479-87.

9. Celebi G, Oztoprank N, Oktem IMA, Heyman P, Lundkvist A, Wahilstrom M, Kotkurt F, Piskin N (2019) Dynamics of Puumala hantavirus outbreak in Black Sea Region, Turkey. Zoonoses Public Health 66:783-797. https://doi.org/10.1111/zph.12625

10. Oncul O, Atalay Y, Onem Y, Turhan V, Acar A, Uyar Y, Caglayik DY, Ozkan S, Gorenek L (2011) Hantavirus infection in Istanbul, Turkey. Emerg. Infect. Dis.17(2):303-304. https://doi.org/10.3201/eid1702.100663

11. Polat C, Sironen T, Plyusemina A, Karatas A, Sozen M, Matur F, Vapalahti O, Oktem IMA, Plyusnin A (2018) Dobrava hantavirus variants found in Apodemus flavicollis mice in Kırklareli province, Turkey. J Med Virol. 90:810-818. https://doi.org/10.1002/jmv.25036

12. Polat C, Ergunay K, Irnak S, Erdin M, Brinkmann A, Cetintas O, Cogal M, Sozen M, Matur F, Nitsche A, Oktem IMA (2019) A novel genetic lineage of Tula orthohantavirus in Altai voles (Microtus obscurus) from Turkey. Infect. Genet. Evol. 67:150-158. https://doi.org/10.1016/j.meegid.2018.11.015

13. No JS, Kim WK, Cho S, Lee SH, Kim JA, Lee D, Song DH, Gu SH, Jeong ST, Wiley MR, Palacios G, Song JW (2019) Comparison of targeted next-generation sequencing for whole-genome sequencing of Hantaan orthohantavirus in Apodemus agrarius lung tissues. Sci. Rep. 9:16631. https://doi.org/10.1038/s41598-019-53036

14. Radford AD, Chapman D, Dixon L, Chantrey J, Darby AC, Hall N (2012) Application of next-generation sequencing technologies in virology. J. Gen. Virol. 93:1853-1868. https://doi.org/10.1099/vir.0.043182-0

15. Oktem IMA, Uyar Y, Dincer E, Gozalan A, Schlegel M, Babur C, Celebi B, Sozen M, Karatas A, Ozkazanc NK, Matur F, Korukluoglu G, Ulrich RG, Ertek M, Ozkul A (2014) Dobrava-Belgrade virus in Apodemus flavicollis and A. uralensis mice, Turkey. Emerg. Infet. Dis. 21(1):121-125. https://doi.org/10.3201/eid2001.121024

16. Jonsson CB, Figueiredo LTM, Vapalahti O (2010) A global perspective on Hantavirus ecology, epidemiology, and disease. Clinical Microbiology Reviews 23(2):412-441. https://doi.org/10.11128/CMR.00062-09

17. Linderholm M, Elgh F (2001) Clinical characteristics of hantavirus infections in the Eurasian continent. Curr Topics Microbiol Immunol 256:135-51.

18. Vaheri A, Henotton H, Voutila main L, Mustonen J, Sironen T, Vapalahti O (2013) Hantavirus infections in Europe and their impact on public health. Rev. Med. Virol. 23:35-49. https://doi.org/10.1002/rmv.1722

19. Klemek B, Tkachenko EA, Dzagurova TK, Yunicheva YV, Morozov VG, Okulova NM, Slyusareva GP, Smirnov A, Kruger DH (2008) Hemorrhagic fever with renal syndrome cause by 2 lineages of Dobrava hantavirus, Russia. Emerg. Infect. Dis. 14(4):617-625. https://doi.org/10.3201/eid1404.071310

20. Papa A, Nemirov K, Henotton H, Niemimaa J, Antoniadis A, Vaheri A, Plyusnin A, Vapalahti O (2001) Isolation of Dobrava virus from Apodemus flavicollis in Greece. J. Clin. Microbiol. 39(6):2291-2293. https://doi.org/10.1128/JCM.39.6.2291-2293.2001

21. Hepojoki J, Strandin T, Vaheri A, Lankinen H (2010) Interactions and oligomerizations of hantavirus glycoproteins. J Virol. 84(1):227-242. https://doi.org/10.1128/JVI.00481-09
22. Kukkonen SKJ, Vaheri A, Plyusnin A (2005) L protein, the RNA-dependent RNA polymerase of hantaviruses. Arch. Virol. 150:533-556. https://doi.org/10.1007/s00705-004-0414-8

23. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP (2012) Predicting the functional effect of amino acid substitutions and indels. PLoS One. 7(10): e46688 https://doi.org/10.1371/journal.pone.0046688

24. Chevreux, B., Wetter, T. and Suhai, S. (1999) Genome Sequence Assembly Using Trace Signals and Additional Sequence Information. Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB) 99, pp. 45-56.

25. Nieberding CM, Durette-Desset MC, Vanderpoorten A, Casanova JC, Ribas A, Deffontaine V, Feliu C, Morand S, Libois R, Michaux JR. (2008) Geography and host biogeography matter for understanding the phylogeography of a parasite. Mol. Phylogen. Evol. 47, 538-554. https://doi.org/10.1016/j.ympev.2008.01.028

26. Michaux JR, Libois R, Paradis E, Filippucci MG. (2004) Phylogeographic history of the yellow-necked fieldmouse (Apodemus flavicollis) in Europe and in the Near and Middle East. Mol. Phylogen. Evol. 32, 788-798. https://doi.org/10.1016/j.ympev.2004.02.18

27. Bilgin R. (2011) Back to the Suture: The Distribution of Intraspecific Genetic Diversity in and Around Anatolia. Int. J. Mol. Sci. 12, 4080-4103. https://doi.org/10.3390/ijms12064080

28. Studer G, Rempfer C, Waterhouse AM, Gumieny R, Haas J, Schwede T. (2020) QMEANDisCo-distance constraints applied on model quality estimation. Bioinformatics. 36(6), 1765-1771. https://doi.org/10.1093/bioinformatics/btz828

29. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumieny R, Heer FT, Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T. (2018) SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 46(W1), W296-W303. https://doi.org/10.1093/nar/gky427

30. Choi Y, Chan AP. (2015) PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics. 31(16), 2745-2747. https://doi.org/10.1093/bioinformatics/btv195

31. Capriotti E, Fariselli P, Casadio R. (2005) I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Res. 33(W1), W306-W310. https://doi.org/10.1093/nar/gki375

32. Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, Ingersoll R, Sheppard HW, Ray SC. (1999) Full-length Human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. Virol. J. 73(1), 152-160.

33. Klempa B, Avsic-Zupanc T, Clement J, Dzagurova TK, Henttonen H, Heyman P, Jakab F, Kruger DH, Maes P, Papa A, Tkachenko EA, Ulrich RG, Vapalahti O, Vaheri A. (2013) Complex evolution and epidemiology of Dobrava-Belgrade hantavirus: definition of genotypes and their characteristics. Arch Virol. 158(3):521-9. https://doi.org/10.1007/s00705-012-1514-5

34. Lundkvist A, Cheng Y, Sjolander KB, Niklasson B, Vahteri A, Plyusnin, A. (1997) Cell culture adaptation of Puumala hantavirus changes the infectivity for its natural reservoir, Clethrionomys glareolus, and leads to accumulation of mutants with altered genomic RNA S segment. J. Virol. 71, 9515–9523. https://doi.org/10.1128/vi.71.12.9515-9523.1997
Supplementary Table 1. The table illustrates the primer sequences that were designed according to gaps on sequences of DOBV-Igneada strain.

| PRIMER          | SEQUENCE                        | SEGMENT | LENGTH | POSITION   |
|-----------------|---------------------------------|---------|--------|------------|
| Forward_S_(G1) | ACAACCACGAAGGCAACTG             | S       | 20     | 70–89      |
| Reverse_S_(G1) | TGTCCTGTAGTGCTCATCAATGTC        | S       | 23     | 357–379    |
| Forward_S_(G2) | GATATGAGGAATACCATCATGGA         | S       | 23     | 1046–1066  |
| Reverse_S_(G2) | CCTAGTGCAATACATCCACCAA          | S       | 23     | 1410–1430  |
| Forward_M_(G1) | GAGACAACATCAAGTGAGGCTAA         | M       | 23     | 331–352    |
| Reverse_M_(G1) | GAAACAATCTCTGCGCTATAACG         | M       | 23     | 1048–1069  |
| Forward_M_(G2) | GGTGTACCGAGCATATAACCTCT        | M       | 23     | 1904–1924  |
| Reverse_M_(G2) | CAGGATTACAGCCCCCACTG            | M       | 20     | 2322–2342  |
| Forward_L_(G1) | GAGGGATTGGTTATCAAAAAGCC         | L       | 23     | 778–800    |
| Reverse_L_(G1) | GTGGGATTTCTTTATGCTTGCA         | L       | 23     | 1319–1341  |
| Forward_L_(G2) | CGAAGTCCTCAGGTTTAGCTAA         | L       | 22     | 2445–2466  |
| Reverse_L_(G2) | GTTCAATAAAGCTCTCCCCAGA         | L       | 22     | 2876–2897  |
| Forward_L_(G3) | GAAGGCTGTGCTGTATCAATAC         | L       | 22     | 3566–3587  |
| Reverse_L_(G3) | TGCATGTAACCTAAAAGTGCC          | L       | 21     | 3910–3930  |
| Forward_L_(G4) | GAGGTAACTCAAGAGATCTTG          | L       | 22     | 4385–4379  |
| Reverse_L_(G4) | GAAGGTCACCTTCATAGAGC           | L       | 20     | 5206–5225  |
| Forward_L_(G5) | CCCCTGCTGCATACTCATTA           | L       | 21     | 5763–5783  |
| Reverse_L_(G5) | CTTTTTGAATACCCAGGACTG          | L       | 22     | 6366–6387  |
**Supplementary Figure 1.** SimPlot analysis represents the nucleotide changes at the S segment sequence of DOBV Igneada strain in contrast to other close related DOBV strain’s S segment sequences. Analysis parameters were as followed: Gap Strip was open, bootstrap was 1000, model was F84 Maximum Likelihood, and T/t ratio was 2.0. X axis represents the nucleotide position on the sequences, while Y axis shows percentage similarity.
Supplementary Figure 2. SimPlot analysis represents the nucleotide changes at the M segment sequence of DOBV Igneada strain in contrast to other close related DOBV strain’s M segment sequences. Analysis parameters were as followed: Gap Strip was open, bootstrap was 1000, model was F84 Maximum Likelihood, and T/t ratio was 2.0. X axis represents the nucleotide position on the sequences, while Y axis shows percentage similarity.
Supplementary Figure 3. SimPlot analysis represents the nucleotide changes at the L segment sequence of DOBV Igneada strain in contrast to other close related DOBV strain’s L segment sequences. Analysis parameters were as followed: Gap Strip was open, bootstrap was 1000, model was F84 Maximum Likelihood, and T/t ratio was 2.0. X axis represents the nucleotide position on the sequences, while Y axis shows percentage similarity.