Virology 535 (2019) 45–58

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/virology

Molecular evolution and selection pressure analysis of infectious hematopoietic necrosis virus (IHNV) revealed the origin and phylogenetic relationship of Iranian isolates in recent epidemics in Iran

Seyed Amir Hossein Jalali[a,b,*], Rezvan Mohammadinezhad[a], Ashraf Mohammadi[c], Mohamad Hassan Latifian[d], Majid Talebi[a,d], Sabihe Soleimanin-Zad[a,e], Pouran Golkar[a], Farhid Hemmatzadeh[f]

[a] Research Institute for Biotechnology and Bioengineering, Isfahan University of Technology, Isfahan, 84156-83111, Iran
[b] Department of Natural Resources, Isfahan University of Technology, Isfahan, 84156-83111, Iran
[c] Human Viral vaccine Department, Razi Vaccine and Serum Research Institute (RVSRI), Hessark Karadj Agricultural Research, Education and Extension Organization (AREEO), Tehran, Iran
[d] Department of Agricultural Biotechnology, College of Agriculture, Isfahan University of Technology, Isfahan, 8415683111, Iran
[e] Department of Food Science and Technology, College of Agriculture, Isfahan University of Technology, Isfahan, Iran
[f] School of Animal and Veterinary Sciences, The University of Adelaide, Roseworthy, SA, Australia

ARTICLE INFO

Keywords:
Infectious hematopoietic necrosis virus
Molecular epidemiology
Molecular evolution
Phylogenetic analysis

ABSTRACT

Infectious hematopoietic necrosis virus (IHNV) is the causative agent for a lethal salmonid disease. In this study, we surveyed the IHNV's epidemiology, diversity and the origin of infection in Iran. Phylogenetic analysis revealed that Iranian isolates belonged to one of the two lineages of E genogroup. Subsequently, a combination of phylogenetic, antigenic and structural analysis was performed to investigate the evolution of E genogroup lineages. Site-specific analysis of the viral glycoprotein showed different co-evolving and positively selected sites in each lineage. Most of these sites were mapped to the predicted antigenic patches of the glycoprotein. Further characterization revealed E lineages can be differentiated, in part, by specific mutations at positions 91 and 130, which are located in the structurally flexible regions of the glycoprotein, suggesting a key adaptative role for these sites. These data may assist in better monitoring the emerging isolates in regions infected to IHNV from E genogroup.

1. Introduction

Infectious hematopoietic necrosis virus (IHNV) is an important viral pathogen which causes a lethal disease in wild and cultured salmonid species (Nichol et al., 1995; Adel et al., 2016). Due to its very contagious nature, IHNV may lead to cumulative mortality of 100% depending on the species and size of fish and environmental condition. Currently, the IHN disease is prevalent in a broad geographical range from North America to Europe and Asia and is notifiable to the World Organization for Animal Health (OIE) (Abbadi et al., 2016; OIE, 2016; Ahmadivand et al., 2017) as a major threat to the aquaculture industry. The causative agent of the diseases (assigned to the type species *Salmonid novirhabdovirus*) is a member of *Novirhabdovirus* genus within the *Rhabdoviridae* family (ICTV, 2014). Its bullet shape virion encapsidates the non-segmented and negative sense single-stranded RNA genome, which encodes for five structural proteins including nucleoprotein (N), phosphoprotein (P), matrix protein (M), envelope glycoprotein (G), and RNA polymerase (L) as well as a unique non-virion protein (NV) (Morzunov et al., 1995; Kurath, 2012). The surface glycoprotein is the only structural protein of the IHNV capable of eliciting neutralizing antibodies in the host cells (Engelking and Leong, 1989; Boudinot et al., 1998; Verjan et al., 2008). This protein has greater genetic diversity compared to other viral proteins and has been subjected to intense studies for either phylogenetic analysis or identifying antigenic constituents for developing vaccines (LaPatra et al., 2008; Alonso and Leong, 2013; Bellec et al., 2017). According to the partial or full-length sequence of this gene, IHNV isolates are currently clustered into five geographically distinct groups, including U, M, L, E, and J. The genogroups U, M, and L represent the geographical distribution of IHNV isolates in upper, middle and lower regions of the west coast of North

* Corresponding author. Research Institute for Biotechnology and Bioengineering, Isfahan University of Technology, Isfahan, 84156-83111, Iran.
E-mail address: sahjalali@cc.iut.ac.ir (S.A.H. Jalali).

https://doi.org/10.1016/j.virol.2019.06.012
Received 7 March 2019; Received in revised form 19 June 2019; Accepted 19 June 2019
Available online 27 June 2019
0042-6822/

America, respectively (Purcell et al., 2009; Ammayappan et al., 2010). The E genogroup, derived from M genogroup, compromise of isolated evolved in European countries (Abbadi et al., 2016; Cieslak et al., 2017). The J genogroup, which shares a common origin to U genogroup and exhibits high genetic divergence, mainly includes isolates from Japan and South Korea (Park et al., 2010; Cha et al., 2012).

The viral glycoprotein also plays a pivotal role in shaping the evolution pattern of IHNV variants (He et al., 2013). Mutations in this protein may result in amino acid changes with negative, neutral or positive impacts in terms of virus fitness and adaptation to host or environmental conditions. Mutations with enhanced fitness in the epitopic regions of glycoprotein can alter the virus antigenicity, known as antigenic drift, which may lead to displacement of a new antigenic variant with previous dominant virus genotype and eventually evolution and differentiation of lineages within circulating populations in a given geographical region (Cattoli et al., 2011; Sandbulte et al., 2011). Displacement events which occur at relatively high rates in RNA viruses such as IHNV may consequently diminish the efficiency of therapies and vaccination (Dormitzer et al., 2011; Rajao et al., 2018).

Determination of the relationship between the antigenic and genetic characteristics is helpful for understanding the evolutionary mechanisms employed by virus isolates for effective adaptation (Liao et al., 2013). Several studies have experimentally mapped the genetic and antigenic evolution of different viruses such as foot-and-mouth disease virus (FMDV) (Xu et al., 2018), vesicular stomatitis Indiana virus (VSV) (Pauszek et al., 2011), bovine viral diarrhea virus (BVDV) (Neil et al., 2019) and Influenza (Kratsch et al., 2016). For IHNV, however, a comprehensive description of the correlation between genetic diversity, amino acid properties, and antigenic diversity remains unclear, largely due to the lack of antigenic data for isolates from different geographical regions, and the fact that not all antigenic alterations can be experimentally determined. Using cross-neutralization assay, Huang et al. (1996) identified eight amino acid positions on IHNV glycoprotein where the mutations affect virus antigenicity. Despite specifying sites prone to antigenic drift, this study did not provide information regarding the association between genetic and antigenic variations in IHNV isolates.

The advent of sophisticated computational programs, however, has rectified the limitations in drawing the link between genetic and antigenic diversity (Sun et al., 2013; Peng et al., 2017; Rajao et al., 2018). Amino acid positions with the potential of antigenic drift can be computationally detected by estimating the degree of positive selection pressure on each site. These positions are identifiable by significant exceed of non-synonymous (dN) to synonymous (dS) mutation rate (Sironi et al., 2015; Hu et al., 2016). The higher amino acid variation at positions under positive selection indicates the occurrence of more evolutionary events in a given population, which results in a more branched phylogenetic tree and divergence of new lineages (Pontremoli et al., 2016). However, it is imperative to note that genetic and antigenic diversity is not always linearly correlated. Sometimes, few but key amino acid mutations may cause biologically remarkable antigenic variations (Ping et al., 2008; Harvey et al., 2016; Chrzastek et al., 2018). In line with this view, the key antigenic sites can be more precisely detected by considering the 3-D structure of the protein, as it better illustrates the flexible positions on the interacting surface of virus antigen and host receptors or antibodies. Furthermore, far apart amino acids in the linear sequence can be in a close proximate in the 3-D structure of a protein to form antigenic patches with the collective effect for escaping from or interacting with host proteins (LaPatra et al., 2008; Kratsch et al., 2016; Glavina et al., 2018). Such data will be beneficial for predicting the possible antigenic shift in the virus population and subsequently for selecting appropriate virus isolates to design antigenically relevant vaccine, especially in regions that show wide circulation of new genotypes.

Iran is one of the top producers of freshwater rainbow trout in the world with a production yield of approximately 163000 tons per year (FAO, 2016). In recent years, the profitability of the trout industry in Iran has remarkably decreased due to the spread of viral diseases including IHNV (Adel et al., 2016; Ahmadivand et al., 2017). Despite the significant economic loss, the real situation of IHNV prevalence and diversity in Iran has not been studied yet. To fill this knowledge gap and as the first goal of this study, we performed a comprehensive investigation to determine IHNV epidemiology, genetic diversity, and phylogeny using samples collected during natural outbreaks of the virus in Iran over a 3.5 years period of time. Using the full IHNV glycoprotein as the target sequence, we showed that all Iranian isolates are exclusively related to one of the two main clades of E genogroup (referred to as E-I vs. E-II). To better understand the underlying evolutionary processes shaping Iranian IHNV population, we then sought to determine the genetic and antigenic characteristics of E-I and E-II lineages using computational approaches combine the phylogenetic, antigenic and 3-D structural information. This information may assist in developing more effective prevention and surveillance strategies to reduce the incidence of new emerging isolates in countries exposed to IHNV infection from E genogroup, such as Iran.

2. Material and methods

2.1. Specimen collection

In the present study, the distribution of IHNV infection in freshwater rainbow trout farms in Iran was monitored in three independent surveys upon the occurrence of suspected outbreaks of viral infection from Dec 2014 to June 2017. In the first survey, sampling was conducted from trout rearing farms of two non-neighboring provinces, Chaharmahal va Bakhtiari and Hamedan, in the west of Iran in Dec 2014. During the second survey, in a broad geographical evaluation, fish samples were collected from 11 top rainbow trout producer provinces in Iran from Apr to May 2015 (Table 1). In the third survey, sampling was purposefully performed from rearing farms of Chaharmahal va Bakhtiari province from May 2016 to June 2017. In all surveys, 6 farms in each province were randomly selected and 10 fish samples weighing 1–20 g were collected and transferred to the diagnostic laboratory at Research Institute for Biotechnology and Bioengineering, Isfahan University of Technology, for virologic and molecular examinations.

2.2. Virus isolation and antigen verification

The tissue samples were prepared in accordance with the guidelines of OIE (OIE, 2016). In brief, a pool of about 0.5 g of virus target organs (heart, spleen, and anterior kidney) was homogenized in Dulbecco's modified Eagle's medium (DMEM) (Gibco, Grand Island, NY, USA), followed by centrifugation at $2000 \times g$ (4 °C). The supernatant of homogenized tissues was used to inoculate monolayers of epithelioma papulosum cyprini (EPC) cells grown in 24-well cell culture plate. The inoculated cells were incubated at 15 °C for two weeks and were examined daily for cytopathic effects (CPE). The supernatants of positive CPE samples were tested thereafter by IHNV specific ELISA (enzyme-linked immunosorbent assay) detection kit (Bio-X Diagnostics, Jemelle, Belgium).

2.3. RNA extraction, RT-PCR amplification, and sequencing

Tissue samples homogenized in DMEM were used directly for RNA extraction to eliminate undesired mutations during cell culture passages. Total RNA was extracted using the Qiagen RNeasy Mini kit according to the manufacturer's recommendations (Qiagen, Hilden, Germany) and quantified with Pico200 spectrophotometer (Picodrrop, Hinxton, UK).

Specific primers to amplify the complete glycoprotein gene were designed according to the published sequences of IHNV glycoprotein in

Table 1
Sampling information during the first (Dec 2014 to May 2015), second (Apr and May 2015) and third (May 2016 to June 2017) surveys for monitoring IHNV infection in Iran trout rearing farms.

	Collection site (province)	Time of outbreak	Infected farms (%)	Isolate name (representative genotype) [a]	Gene bank Sequence ID	Clinical Sign [b]				
						SD	EX	HS	PG	YF
First survey	Chaharmahal va Bakhtiari	Dec 2014	66.6	IRIHNV–CH–14 (IRIHN-14)	MG753776	+	+	–	+	–
	Hamedan	Dec 2014	33.3	IRIHNV-HA-14 (IRIHN-14)	MK376324	+	–	–	+	+
Second survey	Chaharmahal va Bakhtiari	Apr 2015	66.6	IRIHNV–CH–15 (IRIHN-15)	MG753783	+	+	+	+	+
	FARS	May 2015	33.3	IRIHNV-FA-15 (IRIHN-15)	MG753784	+	–	–	+	+
	Isfahan	May 2015	33.3	IRIHNV–IS–15 (IRIHN-15)	MG753777	+	+	+	+	+
	West Azerbaijan	May 2015	33.3	IRIHNV-WE-15 (IRIHN-15)	MG753778	+	+	–	+	–
	Hamedan	Apr 2015	33.3	IRIHNV-HA-15 (IRIHN-15)	MG753779	+	–	–	+	+
	East Azerbaijan	May 2015	0	–						
	Mazandaran	May 2015	0	–						
	Kermanshah	May 2015	0	–						
	Kohgiloye va Boyerahamd	May 2015	0	–						
	Lorestan	May 2015	0	–						
Third survey	Chaharmahal va Bakhtiari	May 2016	33.3	IRIHNV–CH–16 (IRIHN-16)	MG753780	+	–	–	+	–
	Chaharmahal va Bakhtiari	Mar 2017	33.3	IRIHNV-CH1-17 (IRIHN-1-17)	MG753781	–	+	+	+	–
	Chaharmahal va Bakhtiari	June 2017	33.3	IRIHNV-CH2-17 (IRIHN-2-17)	MG753782	+	+	+	+	–

[a] All IHNV isolates that shared a similar sequence in their mid-glycoprotein region were grouped as a genotype, according to a demarcation criteria proposed by MEAP-IHNV database. See the text for more information.
[b] SD: skin darkening; EX: exophthalmia; HS: hemorrhage of skin lesions, PG: pale gills; YF: yellowish fluid in the intestine.

NCBI using Oligo 7.56 software (forward primer: AGAACGCAACTCGC AGAGA, reverse primer: GGTGAAGATTGAGGTCCTTTAGG. Reverse transcription reaction was performed using 1 µg of RNA in 20 µl reaction volumes containing random hexamer and gene-specific reverse primers according to the Thermo Scientific RevertAid First Strand cDNA Synthesis kit protocol (Thermo Fisher Scientific, Waltham, USA). The full length of IHNV glycoprotein was then amplified in PCR reaction containing 2 µl of cDNA, 10 pmol of specific forward and reverse primers, 1.5 mM $MgCl_2$, 0.2 mM dNTPs, 1 × Exprime *Taq* buffer and 1 unit of high fidelity Exprime *Taq* DNA polymerase (Genet bio, Daejeon, South Korea) in a total volume of 30 µl. PCR reaction was performed in a thermocycler (Eppendorf 6325 Mastercycler Pro S, Hamburg, Germany), with the following program: one cycle at 94 °C for 5 min, 35 cycles at 94 °C for 45 s, 56 °C for 45 s, and 72 °C for 90 s, followed by a final elongation cycle at 72 °C for 10 min. The RT-PCR products were separated by agarose gel electrophoresis. At least three positive PCR products for glycoprotein genes from each infected farm in each province were then used for sequencing. For this end, the amplified glycoprotein fragments were gel purified (QIAquick PCR & Gel Cleanup Kit, Qiagen, Hilden, Germany), cloned into pTG19 T/A cloning vector (Vivantis, Selangor, Malaysia) and transformed into *E. coli* strain DH5α. The recombinant plasmids were extracted (QIAprep Spin Miniprep kit, Qiagen, Hilden, Germany) and the sequence of the inserts was determined on both strands (Bioneer, Daejeon, South Korea).

Sequence data were assembled and edited in CLC main workbench software (Qiagen, Hilden, Germany). The identity of the assembled nucleotide sequences was verified by BLAST search (http://www.ncbi.nlm.nih.gov/BLAST/), and the sequences were submitted to GenBank under accession numbers as denoted in Table 1.

2.4. Sequence analysis of Iranian IHNV isolates

The glycoprotein sequences of Iranian IHNV isolates were aligned using Clustal W tool. Nucleotide and amino acid diversity were calculated by pairwise comparison in MEGA version 5 (Tamura et al., 2011) using Kimura two-parameter (K2p) model (Kimura, 1980) for nucleotide data and pairwise distance (p-distance) model for amino acid data.

2.5. Phylogenetic analyses

For phylogenetic analysis, 191 unique complete coding sequences of the glycoprotein gene (1524 bp without stop codon) with defined isolation date were retrieved from NCBI. Sequences with ambiguity or frameshift were not included in datasets. A list of sequence accession numbers is provided in Table S1 in the supplemental material. The phylogenetic and sequence diversity analyses were performed using MEGA 5. Sequence alignment was performed using the MUSCLE tool. The best-fit nucleotide substitution model was selected using the Bayesian Information Criterion score (BIC) via the Find Best DNA Model tool in MEGA 5 (Table S2). The selected best-fit model was general time

Table 2
Genetic distances (%) between nucleotide and amino acid glycoprotein sequences of Iranian IHNV genotypes.

				Amino acid sequence distance (%)	
	IRIHN-14	**IRIHN-15**	**IRIHN-16**	**IRIHN-1-17**	**IRIHN-2-17**
IRIHN-14	-	0.98	0.98	0.79	0.79
IRIHN-15	0.59	-	0.78	0.98	0.98
IRIHN-16	0.79	0.73	-	0.98	0.59
IRIHN-1-17	0.79	0.86	1.06	-	0.79
IRIHN-2-17	0.59	0.53	0.73	0.86	-
	Nucleotide sequence distance (%)				

reversible (Lanave et al., 1984) with invariable sites and gamma distribution. Maximum likelihood (ML) tree was constructed in MEGA 5 using the above parameter settings. To evaluate phylogenetic robustness, bootstrap resampling was undertaken using 1000 replicates.

2.6. Phylodynamic analysis

The BEAST software package v 2.5.1 was used to estimate the mean evolutionary rates and the times to the most recent common ancestor (tMRCA) using the Bayesian Markov chain Monte Carlo (MCMC) method (Bouckaert et al., 2014). First, the input file of data sets was generated in BEAUti v2.5.1. General time reversible with gamma + invariant sites was selected as the best substitution model via bModelAnalyzer (implemented in BEAUTi 2, see Fig. S1) (Rasmussen et al., 2015) and used along with an uncorrelated relaxed lognormal molecular clock type model (Bouckaert and Drummond, 2017). The strict clock type model was rejected because 95% highest probability density (HPD) interval of clock rate variance statistics did not span zero (coefficient of variation (CoV) = 0.78, ucld.stdev = 0.96). A coalescent constant size was set as tree prior. Chain length was run for 50 million generations with sampling every 100 generations (with 10% burn-in) to obtain good convergence (effective sample size value > 200 for each parameter). The convergence of BEAST output file was assessed in TRACER v1.6 (from BEAST package) in which statistical uncertainty was provided by the 95% HPD values. The maximum clade credibility (MCC) tree was generated using Tree Annotator v 2.5.1 and visualized using FigTree v1.4.3.

2.7. Selection pressure analysis

To assess selection pressure acting on the codons of the glycoprotein gene, the ratios of non-synonymous (dN) and synonymous (dS) substitution rates were estimated using the HyPhy open source software package available at the datamonkey web server (https://www.datamonkey.org/) (Delport et al., 2010). Detection of positive pressure was performed using four different codon-based maximum likelihood approaches (Yang, 2000) including single likelihood ancestor counting (SLAC), fixed effects likelihood (FEL), mixed effects model of evolution (MEME), and fast unbiased Bayesian approximation (FUBAR). To further test for possible selection, Tajima's D was employed using the MEGA software (Tajima, 1989).

2.8. Determination of co-evolving pair sites

Co-evolving sites in the IHNV glycoprotein sequences were predicted implying two approaches:1) Bayesian graphical model (BGM)-Spidermonkey implemented in HyPhy package (http://classic.datamonkey.org/dataupload.php) (Poon et al., 2007). In this tool, codons are considered as nodes of a network and the significant connections between amino acid position are determined by edges of the network. 2) the mutual information server to infer coevolution (MISTIC) (http://mistic.leloir.org.ar/index.php) which detects co-evolving sites by estimating their mutual information (MI) (Simonetti et al., 2013). MI is a statistical measure of the dependency between two distinct positions and indicates how much information of amino acid identity at one site is useful for inferring the identity of the amino acid at the second site. Amino acid positions were considered to be co-evolved if they showed a posterior probability > 0.95 in (BGM)-Spidermonkey and displayed an MI score greater than the 95th percentile, which calculated using all MI scores of aligned amino acid sequences in MISTIC.

2.9. Antigenicity and tertiary structure prediction

The antigenicity plots were constructed using James and Wolf method in Protean software (DNASTAR Package, Lasergen). To evaluate the structural properties of amino acid changes, the location of respective sites was mapped to the predicted 3D-structure of IHNV glycoprotein. The B chain of VSV glycoprotein (5i2m.B) was used as the highest-scoring template to construct the homology-based 3D structure of IHNV glycoprotein using the SWISS model web server (https://swissmodel.expasy.org/) (Waterhouse et al., 2018). The predicted structures covered residues 42 to 436 of IHNV glycoprotein.

3. Results

3.1. IHNV prevalence and distribution in Iran

The first goal of the present study was to evaluate the spread of IHN disease in Iran rainbow trout rearing provinces. For this end, IHN infection distribution was monitored over a 3.5-year period and in three distinct surveys.

3.1.1. First survey

In Dec 2014, the first bunch of the samples was collected from fish farms of two geographically far apart provinces in the west of Iran (Chaharmahal va Bakhtiari, and Hamedan). The presence of IHNV was confirmed in ELISA and RT-PCR, targeting the glycoprotein gene. The positive samples were inoculated in EPC cells culture (Fig. S2) and virus propagation was confirmed by observation of cytopathic effects (CPE) and ELISA. In this survey, 66.6% (4 out of 6) of the rainbow trout farms in Chaharmahal va Bakhtiari and 33.3% (2 out of 6) of the rainbow trout farms in Hamedan province showed IHN disease.

3.1.2. Second survey

To obtain an estimate of the prevalence of the IHN disease in different trout rearing regions of Iran and also to assess the genetic diversity of viral strains in Iran, the second survey was conducted with a broader geographical range of evaluation. A total of 660 samples from 66 farms of 11 top rainbow trout producer provinces (~70% of total rainbow trout production of Iran) were randomly collected and screened for IHNV infection during Apr and May 2015.

Fish rearing farms of five provinces, including Chaharmahal va Bakhtiari, Isfahan, Fars, West Azerbaijan, and Hamedan were shown to be IHNV positive (Fig. 1). According to ELISA and RT-PCR tests, none of the IHNV-positive samples were co-infected with infectious pancreatic necrosis virus (IPNV) or viral hemorrhagic septicemia (VHSV). In all infected provinces, more than one farm was detected as IHNV positive. Similar to the previous survey, Chaharmahal va Bakhtiari province with 66.6% of infected farms (4 of 6) showed the highest infection rate in studies provinces (Table 1). On average, 18.8% of studied farms (12 of 66) and 14% of all tested samples were positive for IHNV during the second survey. Fish samples at IHNV positive sites showed clinical signs typical for IHN disease, including skin darkening, abdominal distension, exophthalmia, hemorrhage of skin lesions, and pale gills (Table 1).

3.1.3. Third survey

Considering the high mortality and infection rate in Chaharmahal va Bakhtiari province, the third survey (May 2016 to June 2017) was conducted to monitor the occurrence of IHNV infection in rearing farms of this province, as well as for further assessing Iranian virus diversity. During the third survey, IHNV was detected in suspected virus outbreaks in May 2016, Mar 2017 and June 2017 (Table 1). In all three outbreak times, a lower number of infected farms was detected as compared to the previous surveys (33.3% vs. 66%). However, the infection rates were still high and caused significant economic damage to trout rearing farms of Chaharmahal va Bakhtiari province.

3.2. Genetic diversity of Iranian IHNV isolates

The positive PCR products for glycoprotein genes from the selected locations were sequenced in both directions and submitted to NCBI

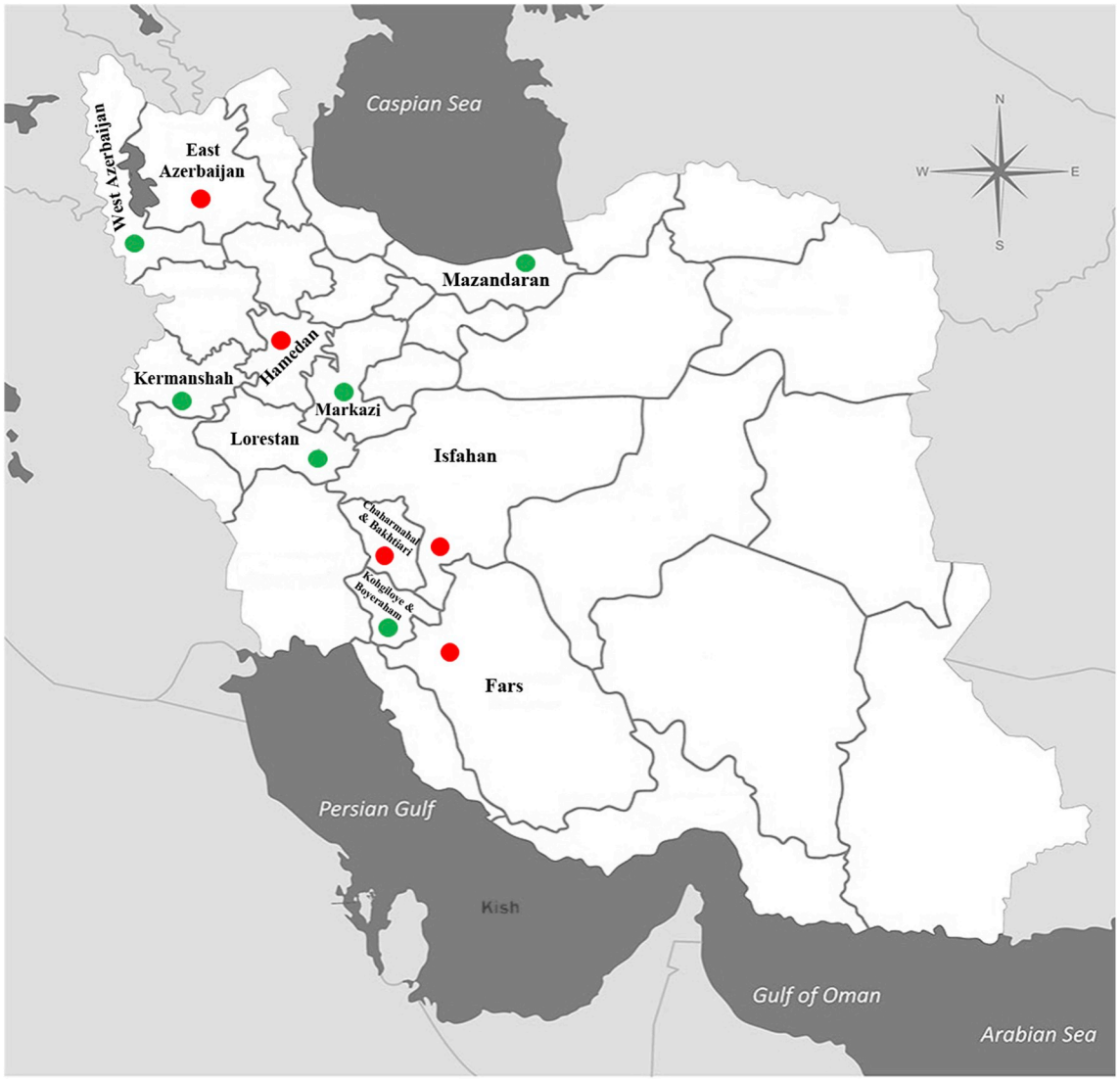

Fig. 1. Map of Iran showing the geographical locations where rainbow trout were sampled. In the first survey (Dec 2014), farms of two non-neighboring provinces, Hamedan and Chaharmahal va Bakhtiari were evaluated. In the second survey (Apr to May 2015), Trout rearing farms from 11 provinces in the west (Hamedan, Markazi, Kermanshah, Isfahan, Chaharmahal va Bakhtiari, Kohgiloye va Boyerahamd, Fars), north (Mazandaran) and northwest (West Azerbaijan, East Azerbaijan) of Iran have been monitored for IHNV infection. In the third survey, fish farms of Chaharmahal va Bakhtiari were further evaluated during 2016–2017. Provinces that proved to be positive for IHNV are represented by red dots while negative ones are indicated by green dots.

(Table 1). Nucleotide sequence alignment of glycoprotein genes revealed that the Iranian isolates obtained during the same outbreak in each survey were identical, irrespective of the geographical distances. Adopting to a defined demarcation criterion proposed by MEAP-IHNV database (http://gis.nacse.org/ihnv/), all IHNV isolates with similar mid-G sequence (a variable 303 nucleotide region in the middle of the glycoprotein gene) could be grouped as a genotype. Based on this criterion, all Iranian IHNV isolates collected during five outbreaks (Dec 2014, May/Apr, 2015; May 2016, Mar 2017 and June 2017) were grouped into 5 unique genotypes designated as IRIHN-14, IRIHN-15, IRIHN-16, IRIHN-1-17, IRIHN-2-17.

Pairwise sequence comparisons revealed low levels of diversity among the obtained glycoprotein sequences. The five genotypes were different in a total of 28 nucleotide and 10 amino acid polymorphic sites, corresponding to 0.53–1.06% nucleotide and 0.59–0.98% amino acid diversity (Table 2). Using the BLAST search tool, all Iranian isolates revealed to share > 99% nucleotide similarity to some Italian (KU878359, KU878360, KU878361) and one German (LN897535) isolates. Most noticeably, the sequence similarity of each Iranian IHNVs to these European IHNVs was greater than its similarity to the rest of Iranian isolates. As an example, the observed nucleotide similarity of IRIHN-2-17 genotype (collected during June 2017) with Italian (KU878359) and German (LN897535) isolates was 99.75%, corresponding to 4 nucleotides difference, whereas similarities to IRIHN-1-17 (collected during Mar 2017) and IRIHN-16 (collected during May 2016) genotypes were 99.14% (corresponding to 13 nucleotides difference) and 99.27% (corresponding to 11 nucleotides differences), respectively. This observation implies that the dominant IHNVs in rearing farm of Iran at each outbreak time are probably not due to the maintenance and diversification of previously established isolates, but rather due to the sequential introduction of IHNV isolates from European countries.

3.3. Phylogenetic analysis and tracking the origin of Iranian IHNV isolates

Phylogenetic relationship of Iranian genotypes was determined with 191 unique glycoprotein sequences belonging to the all five known IHNV genogroups (E, J, L, M, and U), covering a time span from 1966 to 2014 (Fig. 2). All major genogroups in the maximum likelihood (ML) phylogenetic tree were resolved as expected and supported with strong

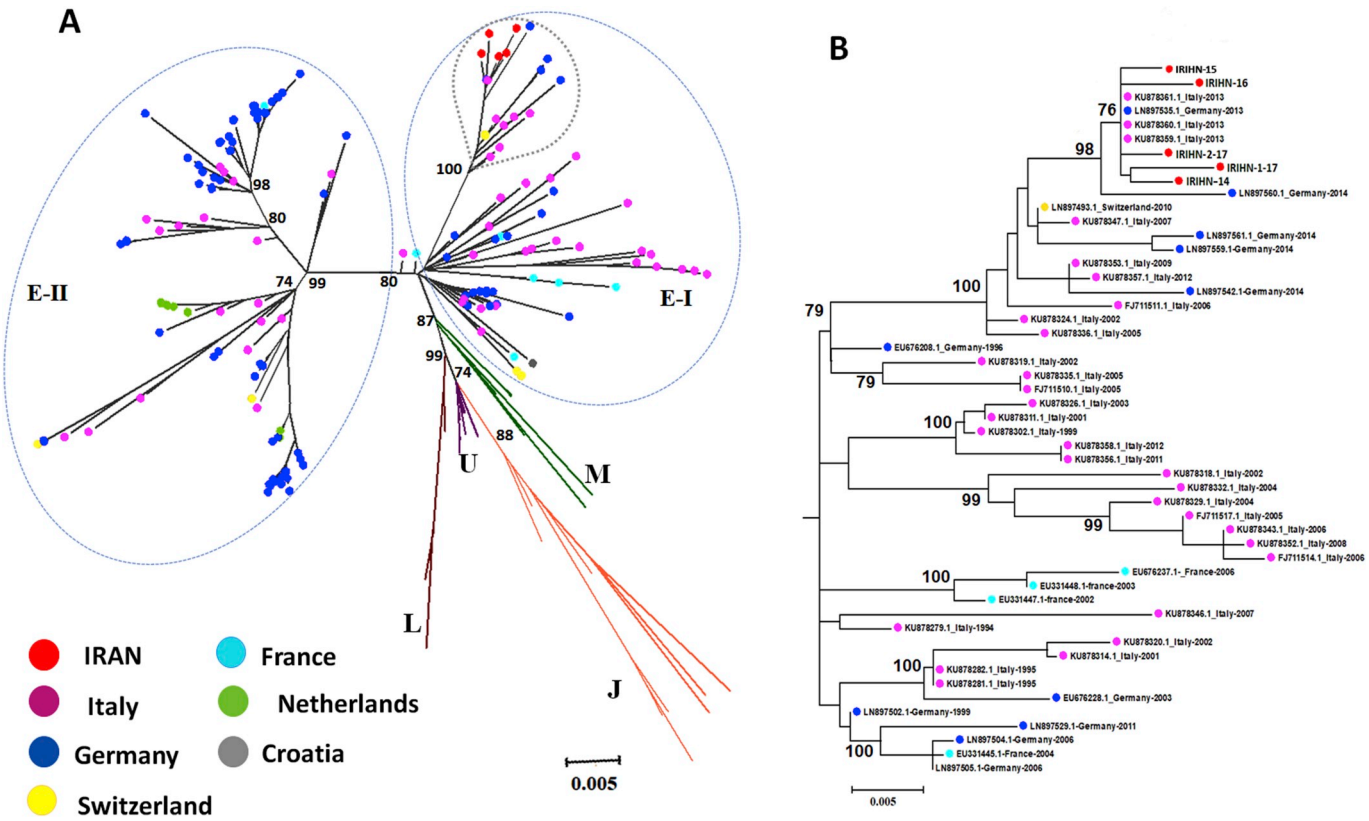

Fig. 2. Phylogenetic trees showing the relationship of Iranian and the 191 other unique full-length IHNV G gene sequences. The trees were constructed in MEGA 5 software using maximum likelihood (ML) method and GTR + G + I as the best fit model. (A) Radial phylogenetic tree format that includes 162 isolates from E, 6 isolates from M, 10 isolates from J, 8 Isolates from U and 5 isolates from L genogroups (GenBank accession numbers are listed in Table S1). Geographic origins of E genogroup isolates are indicated by colored circles. Splitting of E genogroup into two main lineages and bootstrap values of the main nods are shown (B) a rectangular phylogenetic tree to show the relationship of isolates that are indicated by a grey dotted line within the E-I lineage in the radial tree. Bootstrap values were obtained from 1000 resampled datasets and nodes with bootstrap values less than 70% have been collapsed in. Scale bars in each tree represent the expected genetic distance (numbers of substitutions per site).

Table 3
Estimates for genetic diversity and divergence of IHNV E genogroup according to the full sequence of glycoprotein.

Tajima test	dN-dS	tMRAC (95% HPD)	Rate (95% HPD)	Within group aa mean distance (p-distance)	Within group nt mean distance (k2p)	Group
−1/68	0.34	33.95 (32.3–40.37)	8.36 (7.14–9.67)	3/36	2/88	E
−1/44	0.34	32.3 (28.7–38.17)	8.7 (7.2–9.86)	2/95	2/32	E-I
−1/42	0.38	22.49 (20.35–25.09)	12/2 (9.6–15.2)	2/73	2/31	E-II

nt: Nucleotide; aa: amino acid; HPD: highest probability density; TMRCA, time to the most recent common ancestor; dN: number of non-synonymous mutations; dS: number of synonymous mutations.

bootstrap values (77–100%). The E genogroup was divided in two main clades (referred to as E-I and E-II clades in this study) with a bootstrap value of 80% in ML tree, as well as few genetic groups in them, no name has been assigned. The E-I clade was in a close connection to the M genogroup node while the E-II showed a more genetic distance to the ancestral nod. Each main clade was further divided into some sub-clades, but not with strong bootstraps. Maximum clade credibility (MCC) tree with similar phylogenic divisions was also constructed using BEAST software (Fig. S3). All Iranian isolates were clustered within E-I clade in a monophyletic branch with Italian and German isolates.

In order to track the route of Iranian isolates, the phylogeographic patterns of E genogroup clades were investigated. No clear trend could be inferred towards European countries in each clade since a mix of IHNV isolates from different countries was present in both clades. This observation might dismiss the possibility of splitting E-I and E-II clades due to the geographical differentiation of IHNV strains. However, E-I clade, which includes Iranian isolates, was mostly dominated by Italian IHNVs (54.5%) isolated from 1992 to 2013. The E-II clade was mostly dominated by German IHNVs (~65%) isolated from 1998 to 2015. Also, all Netherland isolates were clustered in this clade.

Given the absence of a clear geographical distinction between European IHNV isolates in E genogroup, we tempted to speculate what evolutionary forces have driven the diversification of E-I and E-II lineages and whether the genetic diversity in E genogroup lineages is correlated to antigenic variation? The clarification of these questions is pivotal for broaden our understanding of genetic and antigenic properties of Iranian IHNV isolates which shares the same origin to E-I clade. Accordingly, and as the second goal of this study, we evaluate the genetic diversity, evolution pattern and structural properties of glycoprotein in both E-I and E-II subgroups.

3.4. Genetic variation and divergence of European IHNV genogroup as the origin of Iranian isolates

Mean nucleotide and amino acid distances within E genogroups and clades are depicted in Table 3 When analyzed together, European IHNV

isolates showed 2.88% nucleotide and 3.36% amino acid diversity. When E-I and E-II clades were analyzed as separate datasets, an almost monotonic genetic diversity was observed within E-I and E-II subgroups (i.e. 2.3% nucleotide and 2.95 and 2.73% amino acid diversity, respectively).

The evolutionary rate of the glycoprotein gene was inferred via Bayesian coalescent method in BEAST software. The mean evolution rate of European isolates was estimated to be 8.36×10^{-4} subs/site/year (95% HPD, 7.14–9.67 $\times 10^{-4}$ subs/site/year). When analyzed separately, the hypothesis for the presence of a molecular clock for E genogroup was rejected indicating inconstant substitution rate within this genogroup. E-II clade with 12.2×10^{-4} sub/site/year (95% HPD, 9.6–15.2 $\times 10^{-4}$ subs/site/year), was given 1/4 faster evolution rate compared to E-I subgroup with 8.7 subs/site/year (95% HPD, 7.2–9.86 $\times 10^{-4}$ subs/site/year). The estimated evolution rates were subsequently used to identify the time to the most recent ancestor (tMRCA). In total, analyzed European IHNV isolates were estimated to possess a tMRCA of approximately 33.95 years (~1983). Within European clade, E-I Subclade was estimated to be emerged around 32.3 years before 2017, while tMRCA of E-II group was more recent (22.49 years before 2017).

3.5. Different site-specific evolutionary pattern of European IHNV isolates

To assess whether European IHNV subgroups differ in selection pressure, which might be correlated with differences in their divergent patterns and antigenicity, we calculate the ratio of nonsynonymous (dN) to synonymous (dS) substitutions per each codon using four different methods. In general, the dN/dS ratios estimated for E main group (0.34) and E-I and E-II subgroups (0.34 and 0.38, respectively) were all less than 1.0, suggesting that what they were not under positive selection but purifying selection, congruent with results of Tajima's D test (Table 3).

Negatively selected sites in each clade were searched using SLAC, FEL and FUBAR methods. Amino acid positions which selected by at least two methods were used to calculate the percentage of negatively selected sites in every 50 amino acids of the glycoprotein sequence (Fig. 3A). This analysis showed that the action of negative selection was not evenly distributed along the glycoprotein sequence of E-I and E-II clades. For E-I, the strongest selection constraint was observed in the mid-region to C-terminal part of the protein (aa 300–450), while for E-II clade the strongest purifying selection was observed in the *N*-terminal region (aa 1–200) of the protein. Noteworthy, serine amino acid at position 498 (in the cytoplasmic part of glycoprotein) showed to be under strong negative selection in both E clades as well as in U, L, M and J main genogroups (data not shown) which implies an important function for this codon.

Despite the action of negative selection on the glycoprotein, a number of codons were indicated to be under positive selection. Table 4 shows the amino acid positions selected by at least 3 models ($P \leq 0.1$ or posterior probability ≥ 0.9). When all of the European sequences were analyzed together, 13 amino acid sites were identified to be under positive selection including 4, 98,168, 247, 257, 270, 276, 277, 284, 442 and 475. When two subclades of E group were analyzed separately, it was found that these subgroups are not similar in terms of positive selection sites, which could be contributed to the diversifying events leading to each clade. Isolates of E-I clade were found to have five amino acid sites under positive pressure, including 247, 252, 277, 376 and 475 whereas E-II clad showed eight positively selected sites (4, 24, 247, 252, 256, 257, 270, 276). We then adapted the dN-dS plot (obtained with SLAC method) to the consensus antigenicity plot of glycoprotein to investigate whether antigenic sites are also exposed to selective pressure (Fig. 3B). Most of the positively selected sites were matched to the predicted antigenic regions, especially substitutions at positions 24, 257, 277 and 376 were in prominent antigenic-stretches. However, position 247 was clearly matched to non-antigenic sites. Also, positive selections at positions 4 and 475 were not related to the antigenicity of glycoprotein as they were located in the signal peptide and transmembrane domain parts of the protein, respectively.

Co-evolved positions were also determined using the aligned amino acid sequence sets of each lineage. According to the probability values and MI scores of BGM-Spidermonkey and MISTIC tools, seven and eight different pairs of co-evolved positions were identified for E-I and E-II lineages, respectively (Table 5, Fig. 3A). While there were no identical co-evolved pair sites between the two lineages, amino acid positions 247 and 376 were involved in shaping the co-evolving network of both E-I and E-II groups. Furthermore, despite high amino acid polymorphism at positively selected sites, most of the co-evolved pairs were not correspond to these sites. The exception was of (247,376) co-evolved pair in E-I lineage where glutamine amino acid at position 247 was synchronously evolved with arginine at position 376 ($p > 0.999$). Noteworthy, amino acid tyrosine at position 437 of E-I lineage was synchronously evolved with four other sites (e.i. 26, 79, 272 and 435). This may reflect a central role for this position in the modulation of intra- or inter-protein interactions.

3.6. Biochemical and structural properties of the evolutionary important sites within the E genogroup

We also analyzed the amino acid substitutions at identified positively selected sites as well as at positions that were not under selection, but different between each clade (91, 130, 436 and 491). Amino acid substitutions, their frequencies and biochemical properties corresponding to un-polar (*), un-charged-polar (°) and charged (±) categories are listed in Table 6.

In some positions, substitutions were only observed between amino acids of the same biochemical class, including positively charged amino acid substitutions at position 130 and positive and negative charged amino acids at position 274. For three sites (4, 277 and 491) substitutions were between un-polar and un-charged-polar amino acids, whereas other positions exhibited more flexibility to accept amino acid substitutions from different biochemical categories.

The amino acids at positions 91 and 436 were distinctly different between the two subgroups. In E-I clads, position 91 was occupied by aspartate or asparagine. However, in clad E-II, it was occupied by alanine and glutamate. The position 436 in subgroup E-I was fixed with un-polar leucine amino acid, while it was fixed by positively-charged histidine in E-II subgroups. The dominant amino acids at positions 130, and in particular 491 of clade E-I were also different compared to amino acids of E-II clade.

In order to evaluate the impact of amino acid substitutions on protein properties, the consensus amino acid sequence (based on maximum frequency of amino acids at each site) of each subgroup was used to construct the antigenicity plot as well as to predict the 3D structure of glycoprotein. Interestingly, at all positions which differ between the two consensus sequences (i.e. 91, 130, 247, 252, 257, 276, 277, 436 and 491), E-II subgroups showed higher antigenicity than E-I subgroup (Fig. 4). The higher antigenicity was more prominent at positions 91, 276–277 and 436.

The 3D structure of the consensus sequences of E-I and E-II lineages were predicted in a homology-based manner using the B-chain of VSV glycoprotein as the template. The ectodomain part of glycoprotein in this family consists of three main domains including, central domain (CD), fusion domain (FD) and pleckstrin homology domain (PHD) and five hinge regions (R1-R5) (Fig. 5A). The main domains and hinge regions on the predicted 3D structures of IHNV glycoproteins were determined using their homologous counterparts of VSV glycoprotein (Fig. S4). All positively selected sites were highly exposed or at least accessible on the surface of PHD and CD domains (Fig. 5B). None of the amino acid changes at positively selected sites resulted in structural

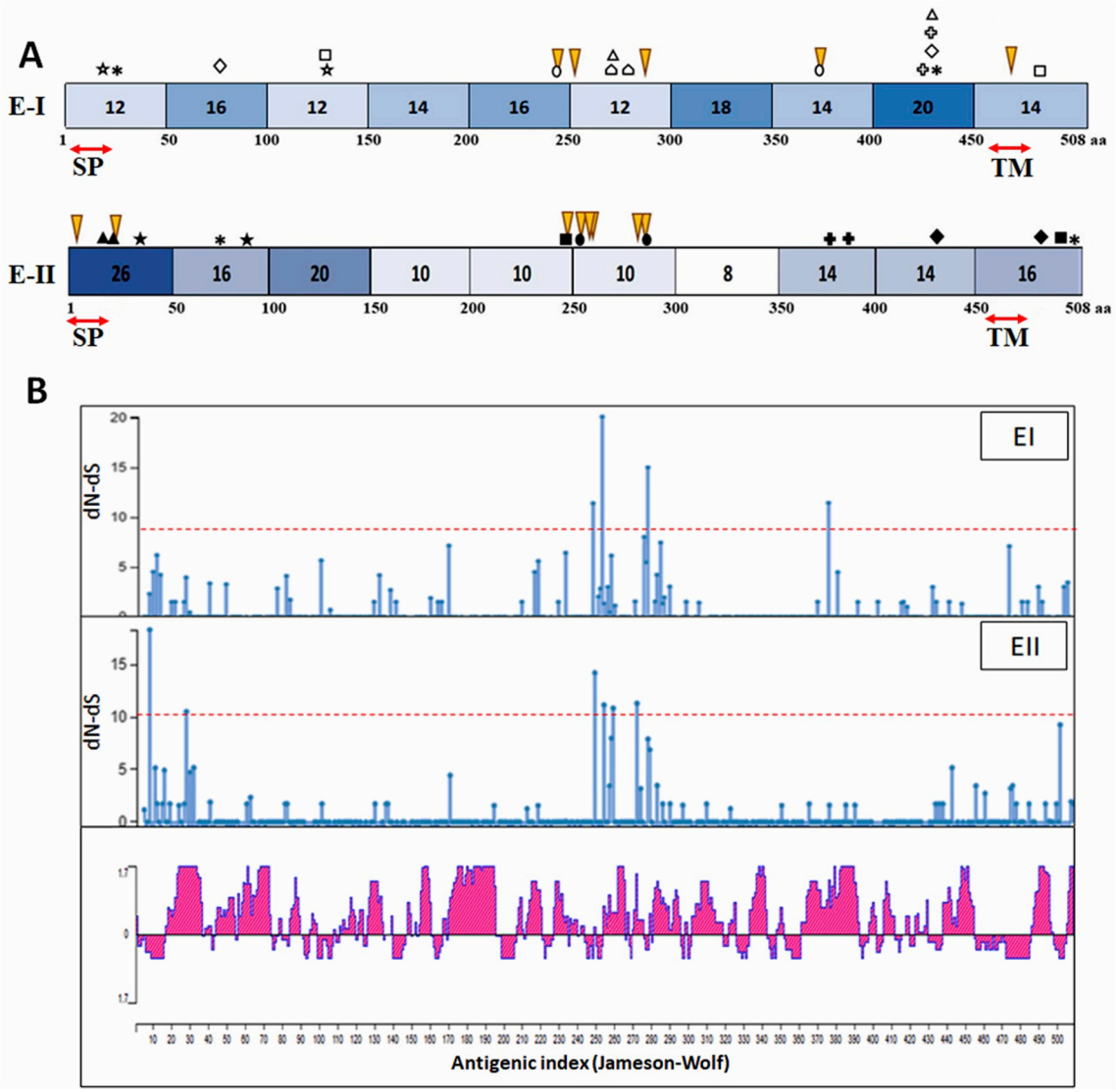

Fig. 3. A) Schematic representation of IHNV glycoprotein in E-I and E-II lineages plotted in linear diagrams (SP: signal peptide, TM: transmembrane domain). Each window of 50 amino acids is colored in hues of blue based on the percentage of negatively selected sites, which reported inside the windows. Positively selected sites are represented by yellow triangles. Each co-evolving pair site in E-I and E-II lineages are represented by similar black filled or white filled symbols, respectively. Co-evolving symbols are corresponding to the symbols indicated in Table 5 (B) Mapping the location of positively selected sites of E-I and E-II lineages on the predicted antigenicity plot of IHNV glycoprotein. dN/dS ratio graphs were obtained using SLAC method. The x-axises show amino acid positions along with the IHNV glycoprotein. The red dot lines on dN/dS graphs indicate the thresholds at which amino acid positions were identified to be under positive pressure by at least three models (p-value < 0.05 or posterior probability > 0.95, see the material and method for more detail). Antigenicity plot was constructed using the Jameson-Wolf model in PROTEAN software (DNAstar package, Lasergene).

changes in the protein. The amino acids at positions 247, 252 and 257 participated in the formation of two antiparallel sheets, while amino acids at sites 276 and 277 were presented in a highly exposed α-helix in both lineages, suggesting structural conservation at these sites. In contrast, variations in the positions that were not under positive selection caused changes in the predicted 3D structure of E-I and E-II lineages' glycoprotein. The presence of alanine at position 91 in the consensus sequence of E-I lineage resulted in the formation of a small α-helix in its following stretch, while aspartic acid at this position in E-I subgroup was present in a disordered loop. On the other hand, the fixed lysine at position 130 in E-I lineage caused a loop in the protein structure, while replacement of lysine with arginine in E-II subgroup overwhelmed the loop in favor of the formation of a short and very exposed α-helix. Since in the present study amino acid positions 42 to 436 were mapped to the homology-based constructed models, the local secondary structure for position 436 (at the end of the 3D model) was not reliable and therefore excluded from the analysis.

4. Discussion

The first report for IHNV detection in Iran goes back to 2003 (Fallahi et al., 2003). However, very few studies have been conducted so far for determination the virus spread and diversity in Iran. Furthermore, the few available studies mostly presenting virus distribution in two or three provinces targeting the partial sequence of viral glycoprotein (Adel et al., 2016; Ahmadivand et al., 2017).

The present study, which was carried out within the scope of vaccine development for controlling IHN disease in Iran, represents the first large-scale evaluation of IHN disease spread and virus diversity in Iran using a total of 960 specimens from 11 top rainbow trout producer provinces over a 3.5-year period. Furthermore, given the fact that IHNV

Table 4
Analysis of selection pressure on the glycoprotein of IHNV E-genogroup and E-I and E-II lineages using SLAC, FEL, MEME, and FUBAR methods. Only positions that found to be under positive selection by at least three out of four methods are shown.

Group or subgroup	Amino acid sites	SLAC	FEL	MEME	FUBAR
		P [dN/dS > 1][a]	p-value (β)[b]	p-value (β+)	Prob[α < β][c]
EUROPE	4	0.036	0.055	0.05	0.95
	24	0.040	0.009	0.01	0.99
	28	0.049	0.05	0.05	–
	98	0.041	0.049	0.04	0.96
	168	0.05	0.014	0.02	0.99
	247	0.004	0.001	0.00	0.99
	252	0.002	0.000	0.00	1.00
	257	0.008	0.004	0.01	0.99
	276	0.067	0.015	0.02	0.99
	277	0.004	0.003	0.00	0.99
	284	0.056	0.05	0.04	0.95
	442	0.098	0.050	0.01	–
	475	0.0838	0.011	0.02	0.96
E-I	252	0.01	0.001	0.001	1
	277	0.018	0.009	0.01	0.997
	376	0.047	0.039	0.043	0.98
	247	0.039	0.04	0.048	0.97
	475	–	0.037	0.05	0.96
E−1I	4	0.036	0.014	0.02	0.99
	24	0.037	0.034	0.05	0.98
	247	0.047	0.012	0.02	0.99
	252	0.083	0.05	0.04	0.98
	256		0.042	0.03	0.97
	257	0.05	0.030	0.04	0.98
	270	0.046	0.030	0.01	0.99
	276	–	0.047	0.1	0.97

[a] Amino acid sites with $p < 0.05$ in SLAC, $p < 0.05$ in FEL, $p < 0.1$ in MEME methods or posterior probability > 0.95 in FUBAR methods have been selected.
[b] β indicates the non-synonymous substitution rate.
[c] α indicates the synonymous substitution rate.

Table 5
Identified co-evolved sites on the glycoprotein of E-I and E-II lineages of IHNV E genogroup.

Lineage	Co-evolved sites[a]	Co-evolved amino acids[b]	posterior probability in (BGM)-Spidermonkey[c]	MI score in MISTIC[d]
E-I	272, 437 (△)	N ↔ Y K/q ↔ H	0.96	47.64
	26, 437 (✳)	D ↔ Y	0.96	47.3
	247, 376 (◯)	Q ↔ R	0.99	41.41
	79, 437 (◇)	S ↔ Y	0.93	39.57
	435, 437 (✲)	A ↔ Y V ↔ H	0.97	35.93
	272, 286 (⌓)	N ↔ K K/n↔ N	0.96	32.54
	138, 491 (□)	P ↔ T L↔ S	0.98	31.03
	20, 138 (☆)	S ↔ T G ↔ S	0.97	25.84
E-II	37, 81 (★)	A ↔ K S ↔ R/q	0.99	64.2
	252, 277 (●)	D ↔ T	0.99	58.37
	15, 20 (▲)	A ↔ G T ↔ S	0.97	50.93
	67, 505 (❋)	D ↔ A N ↔ T	0.99	34.7
	247, 495 (■)	H ↔ S	0.99	33.75
	376, 381 (✚)	R/h ↔ T L ↔ M	0.99	24.07
	433, 484 (◆)	I ↔ C V ↔ Y	0.95	25.27

[a] Each co-evolving amino acid pair in E-I and E-II lineages is represented by a black filled or white filled symbol, respectively. Co-evolving symbols are corresponding to the symbols indicated in Fig. 3.
[b] For each position pair, the co-evolving amino acid at site 1 and 2 are indicated. In co-evolving pairs, an amino acid at position 1 appears always coupled with the same amino acid at position 2 (indicated by uppercase words). However, in some cases, the amino acid at position 1 may appear less frequently coupled with another amino acid at position 2 (indicated by lowercase words).
[c] Position pairs with posterior probability > 0.95 in (BGM)-Spidermonkey tool are presented.
[d] Position pairs with MI score greater than the 95th percentile, which calculated using all MI scores of aligned amino acid sequences in MISTIC tool are presented.

infection in Iranian farms is probably of European origin, herein we presented new evidence regarding the genetic characteristics and evolutionary patterns within the E genogroup. This information would be valuable for developing and implementing new preventive strategies for IHN disease in Europe and countries which receive exports from Europe such as Iran.

4.1. Virus prevalence and diversity in Iran

In Dec 2014, after receiving several reports about the mass mortalities in trout farms of Iran, the first survey was conducted to assess the prevalence of IHN disease in different provinces. In addition to the observation and recording the clinical signs, ELISA test, virus isolation in cell culture and molecular analysis have been employed to detect IHNV. Virus infection status was examined in Hamedan and Chaharmahal va Bakhtiari provinces, and a unique genotype was identified in these geographically distinct. In a previous study, Adel et al. (2016) also reported the isolation of one virus strain with identical sequence from Mazandaran province in the north and Chaharmahal va Bakhtiari province in the west of Iran during a same period of the time. These observations together raise this assumption that irrespective to the geographical distance, one genotype may be dominated in different regions of Iran at each outbreak time. To further examine this hypothesis, we evaluate the virus infection status in 11 top rainbow trout producer provinces of Iran in Apr/May 2015 (second survey of this study). Five out of 11 investigated provinces (45.45%) were identified as IHNV infected zones, including three neighboring and two geographically far apart provinces. In accordance with our small-scale investigation, one unique genotype was detected in all infected provinces, suggesting the occurrence of a genetic shift at each virus infection onset time. Both genotypes obtained in the first and second surveys shared the most similarity (99%) to some new Italian and German isolates (2013–2014).

Although the average time for the emergence of a new IHNV genotype has been reported to be around one year (Cieslak et al., 2017), the domination of one new genotype at each event in distinct geographical provinces suggests that the frequent IHNV outbreaks in Iran is not due to the previously established viruses, but rather due to the continual import or intra-country distribution of infected eggs or fish. The imported viruses will replace the previous isolates most probably due to the higher prevalence in the population or possibly due to their greater adaptability or competitive ability (Dixon et al., 2016; Abbadi et al., 2016). This implies that stricter biosecurity rules and programs should be applied by the veterinary organization to prevent the import of infected samples.

Chaharmahal va Bakhtiari was identified as the most IHNV infected zone in both studies. This province is always ranked among the major trout producer in Iran, though it has been faced with significantly reduced production since 2013, mainly due to the spread of viral disease (Ghorani et al., 2016). A high density of farms is presented in this

Table 6
Analysis of glycoprotein of E-I and E-II lineages of IHNV E genogroup for amino acid variations, frequencies and biochemical properties.

Site / Lineage	4[a]	24	91[d]	130	247	252	256	257	270	276	277	376	436	491
E-I	T°[b] (83/3)[c] M* (16/7)	K^+ (90/3) E^- (8/3) H^+ (1/4)	A* 69/5 E^- (30/5)	K^+ (53) R^+ (47)	Q° (61/1) R+ (31/9) H^+ (5/6) K^+ (1/4)	E^- (30/6) G° (25) A* (23/6) D^- (11/1) T° (6/9) N° (2/8)	K^+ (84/7) G° (8/3) N° (4/2) E^- (1/4)	Q° (95/6) P* (1/4) S° (1/4) D^- (1/4)	E^- (80/6) V* (16/6) A* (1/4) D^- (1/4)	E^- (87/5) K^+ (11/1) D^- (1/4)	I* (73/6) T° (25) V* (1/4)	R^+ (61/1) H^+ (20/8) L* (15/3) S° (2/8)	L* (100)	P* (84/7) L* (13/9) S° (1/4)
E-II	T° (46/7) M* (28/9) S° (21/1) A* (1/1) L* (1/1)	K^+ (45/5) Q° (40) P* (10) A* (2/2) T° (1/1) H^+ (1/1)	D^- (67) N° (33)	R^+ (100)	H^+ (50) Q° (38/9) R^+ (11/1)	D^- (53/3) E^- (31) V* (10) N° (1/1) T° (1/1)	K^+ (89) G° (4/4) T° (3/3) N° (1/1)	P* (67/7) S° (31/1) T° (1/1)	E- (84/4) D^- (10) A* (2/2) S° (1/1) T° (1/1) V* (1/1)	K^+ (56/6) E^- (41/1) R+ (1/1) D^- (1/1)	T° (45/5) I* (45) V* (2/2) A* (2/2)	R^+ (90) L* (6/7) H^+ (3/3)	H^+ (100)	S° (95) T° (5)

[a] Blue highlight indicates positively selected sites. Yellow highlight indicates non-positively selected sites which differ between E-I and E-II lineages of IHNV E genogroup.
[b] Underlines indicate positions where the consensus amino acids (based on the maximum frequency) are not similar between E-I and E-II lineages.
[c] Biochemical characteristics of amino acids are shown in superscript. Amino acid's polarity symbols are as follow: (*) un-polar, (°) un-charged-polar, (+) positively charged, (−) negatively charged.
[d] At each site, frequencies are shown in parenthesis and amino acids are listed according to decreasing frequency.

relatively small province and most of them share the same ownership and operation management which increase the probability of virus horizontal transmission. Considering the high infection rate in Chaharmahal va Bakhtiari province and the observation that one genotype is dominated in Iran trout farms at each outbreak time, we conducted our third survey by monitoring the Chaharmahal va Bakhtiari farms during 2016 and 2017. In this period, three main outbreaks were recorded and three different genotypes were isolated at each time. Similar to the previous surveys, the obtained genotypes shared the most similarity to the Italian and German isolates, confirming the European origin of IHNV isolates in Iran. However, as of this report, there was no available IHNV sequence in the Gene Bank which be isolated during 2016–2017. Therefore, we could not evaluate the similarity between Iranian and European isolates in the same period of time.

4.2. Phylogenetic analysis

According to the phylogenetic analysis, all Iranian isolates were clustered within a major lineage of E genogroup (E-I subgroup) which separated with significant bootstrap from another E lineage. The split of European isolates into two main lineages has been also reported by other researchers (Kolodziejek et al., 2008; Haenen et al., 2016; Cieslak et al., 2017). By analyzing the mid-reign of 18 unique glycoprotein sequences of European IHNV isolates, Kolodziejek et al. (2008) suggested that different geographic origin of the viruses may be the reason for the differentiation of the two European IHNV lineages. However, in our analysis with the full-length glycoprotein of 161 unique European isolates, no clear geographical correlation could be inferred within each subgroup since a mix of isolates from different European countries were presented in both subgroups. The exception was Netherland isolated which were exclusively clustered within E-II subgroups. The circulation of IHNV in the Europe Union, probably via extreme trout-trading, has been previously proposed as an obstacle for conclusive clarification of phylogenetic relationships within the E genogroup (Enzmann et al., 2005; He et al., 2013). The absence of a clear trend toward geographical distribution in each subgroup posed the question to us which evolutionary events have resulted in the splitting of European isolates? And what are the molecular characteristics of both subgroups, in particular, the E-I lineage which comprises of Iranian isolates? These questions led us to evaluate the genetic diversity, evolution pattern and structural properties of glycoprotein in both subgroups.

Despite almost similar nucleotide and amino acid diversity within E-I and E-II subgroups, the evolutionary pattern of each lineage was found to be different. According to Bayesian MCMC analysis, the adaptation of E-II isolates to trout host and their spread across European countries has been followed by a 1.4-fold increase in the mean evolutionary rates (substitutions/site/year) as compared to the isolates of E-I subgroups.

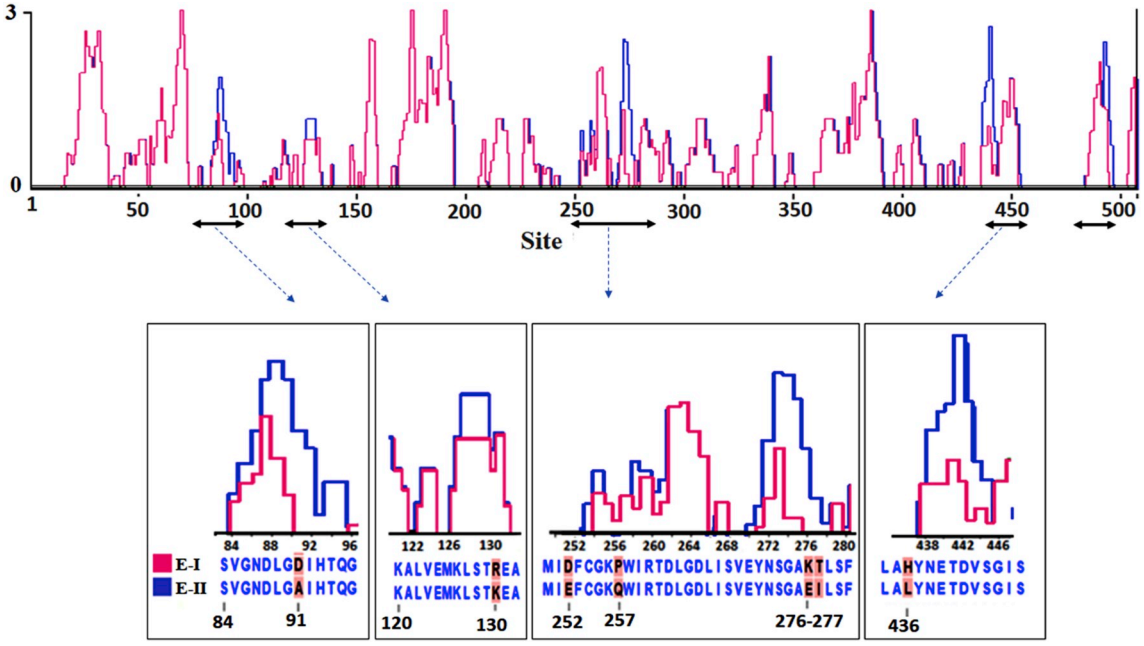

Fig. 4. Comparison of antigenicity profiles of IHNV glycoprotein in E-I and E-II lineages. Upper: antigenicity plots constructed using the consensus amino acid sequences of the full length of glycoproteins in PROTEAN software (DNAStar, Lasergene). In each E-I and E-II lineages, the consensus sequence was inferred from the maximum frequency of amino acids at each position (see Table 6). Lower: a close-up view of antigenicity profiles of E-I and E-II lineages at positions that differ between the two consensus sequences in the ectodomain of glycoprotein. Amino acid position 491 is in the cytoplasmic stretch of the protein.

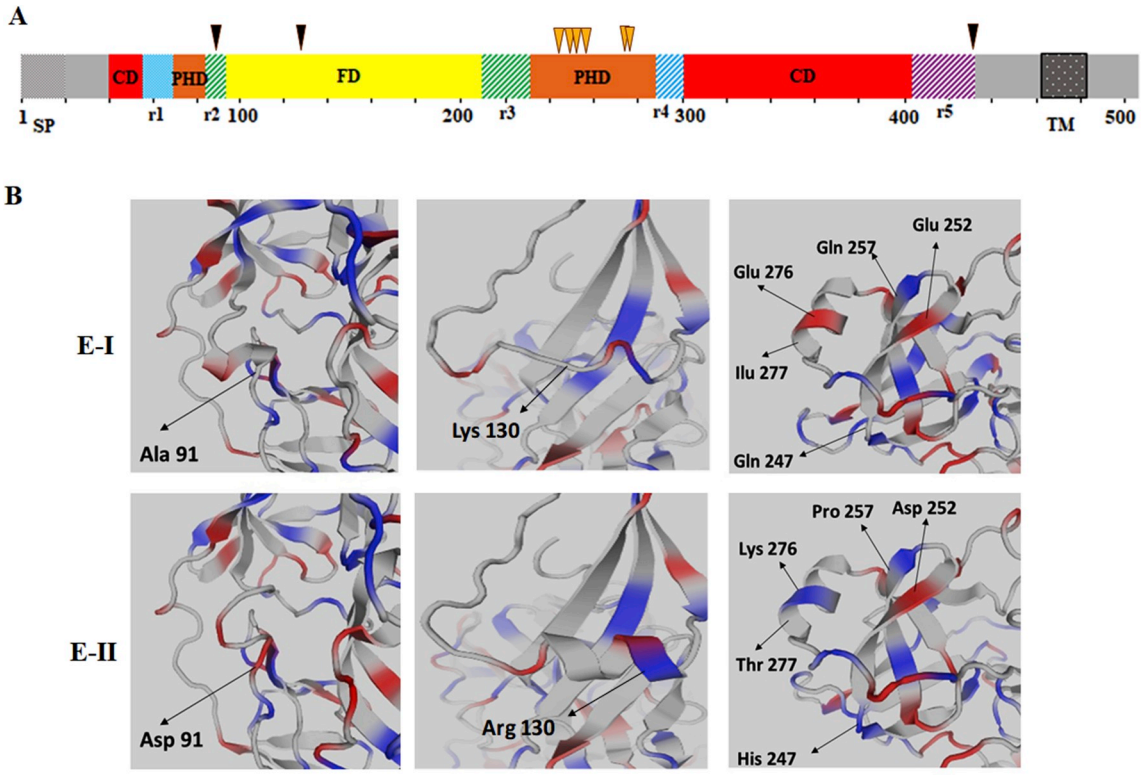

Fig. 5. A) linear plot representation of the main domains and hinge regions of IHNV glycoprotein as inferred from the homologous counterpart, VSV glycoprotein (see Fig. S4). Red boxes: central domain (CD) or trimerization domain, Yellow box: fusion domain (FD); Orange boxes: Pleckstrin homology domain (PHD); hatched blue boxes: hinge regions R1 and R4; Hatched green boxes: hinge regions R2 and R3; Hatched purple box: hinge region R5. Color coding and names are according to (Baquero et al., 2013). Amino acid positions that vary between the consensus sequence of E-I and E-II lineages are indicated by orange (positively selected sites) or black (lineage-specific sites) triangles, B) 3-D visualization of the possible locations of amino acid sites that vary between the consensus sequence of E-I and E-II lineages. In each E-I and E-II lineages, the consensus sequence was inferred from the maximum frequency of amino acids at each position. 3-D structures were generated by SWISS-MODEL web server using the VSV glycoprotein B chain (5i2m.B) as the homologous template.

The overall evolution rate of E genogroup and other IHNV main genogroups (Table S2) were all within the ranges of previously predicted rates based on the full sequence of G gene (He et al., 2013; Bellec et al., 2017), but lower than the rates ranges estimated based on the mid-glycoprotein gene sequence (Garver et al., 2003; Kurath et al., 2003). It worth to note that the mid-region of IHNV glycoprotein gene is more variable than the rest of the glycoprotein sequence (Kolodziejek et al., 2008; He et al., 2013) and may result in the false estimation of evolution rate.

Analysis of tMRAC in BEAST software suggests that E genogroup has possibly emerged around 1983 which approximately coincides with the first documentation of IHNV infection in Europe in 1987 (Bovo et al., 1987; Baudin-Laurencin, 1987). Also, the tMRCA of E-II lineage was estimated to be around 10 years later (~1995) than the E-I lineage. This result suggests that a differentiation event with a more selective fitness advantage may have occurred in the mid-1990s, which put greater evolutionary pressure on European IHNVs toward the E-II lineage since despite having less time to spread, this subgroup has evolved with higher evolution rate.

The other noticeable point that can be inferred from the phylogenetic analysis is that the genetic distance of European isolates (in special E-II lineage) from their ancestral node in M genogroup is continually increasing (Fig. 2). Therefore, the American strains (e.g. WRAC strain) which have been frequently used for developing vaccines against IHNV isolates from different genogroups (Corbeil et al., 2000; LaPatra et al., 2001; Garver et al., 2005; Lorito et al., 2011) may not be as efficient as previous against the current European IHNVs. This implies the necessity for applying relevant European isolates to develop more effective vaccines instead of the American strains.

4.3. Site-specific analysis of IHNV glycoprotein

To further evaluate the possible diversifying pattern in E genogroup, we then performed site-specific evolutionary analysis using different methods.

IHNV glycoprotein was shown to mostly undergo purifying events ($dN/dS < 1$), in agreement with reports of other researchers (He et al., 2013; Abbadi et al., 2016). Generally, purifying selection is a mechanism to decrease the accumulation of deleterious mutations in the coding sequence, which repeatedly occurs in RNA viruses due to error to prone nature of their RNA-dependent RNA polymerase (Elena and Sanjuán, 2005). Notably, the extent of functional constraint along the glycoprotein sequence of E-I and E-II lineages did not follow an equal distribution. Despite the action of purifying selection, 13 codons in European glycoprotein data set were found to be strongly under positive selection, as detected by all of the applied methods. When E-I and E-II lineages were analyzed separately, a clearly different pattern of positive selection was evident in each lineage, which may be contributing to their different evolution patterns. A similar observation was reported by LaParta et al. (2008) in the analysis of the selection pressure on different subgroups of IHNV isolates from Idaho. As expected, all putative positively selected sites in the ectodomain (252, 256, 270, 277, 376 and 436) were mapped to the predicted antigenic region of glycoprotein, except amino acid 247. Previous studies have also demonstrated the localization of neutralizing antigenic epitopes in the mid-part (270-336aa fragment) of IHNV glycoprotein (Xu et al., 1991, 2014). However, only one study has been conducted so far for experimental determination of antigenic amino acid positions in IHNV glycoprotein. In this experiment, Huang et al. (1996) reported that mutants with amino acid variations at positions 78, 81, 230–231, 272, 273, 275 and 276 can escape neutralizing monoclonal antibodies. Herein, we could not find selective pressure for these amino acid sites in European sequences except position 276. This contradiction may be explained by the fact that isolated used in Huang et al. study belonged to M genogroup. Furthermore, they performed consecutive virus passaging in cell culture to obtain antibody-escape mutants, which could result in the mutations that are not present in the wild-type populations (Novella et al., 2005; LaParta et al., 2008).

E genogroup lineages also showed a different pattern of co-evolving pair sites. Several studies have demonstrated that co-evolving amino acid positions which exert selective pressure on each other, can reveal structurally or functionally important sites, such as sites represent compensatory or epistatic effects (Kryazhimskiy et al., 2011; Bedhomme et al., 2015). For instance, epistatic interaction between two co-evolving positions in the glycoprotein of Chikungunya virus E1 enabled the virus to infect a new mosquito vector (Tsetsarkin et al., 2011). Therefore, the different co-evolving pattern may contribute to the evolutionary trajectories of viral strains to host shift or immune evasion (Forni et al., 2016).

In addition to different co-evolving and positively selected sites, four other positions (91, 130, 436 and 491) were demonstrated to be specific to E-I or E-II lineages, as they were completely (91 and 436) or nearly (130 and 491) fixed with different amino acids. Evaluation of biochemical properties of amino acids at positively selected or lineage-specific sites revealed that E-II lineage has possibly evolved toward having more polar or charged residues. Moreover, the predicted antigenicity for the consensus sequence of E-II lineage was greater than E-I at all different positions. Therefore, this hypothesis can be argued that the increased antigenicity of E-II lineage is a part of its adaptive evolution, which allowed it to spread with higher evolutionary rate and in a shorter period of time as compared to E-I lineage. Notably, Abbadi et al. (2016) reported a progressively increased virulence of European IHNV isolates in recent years. However, whether this putative increased antigenicity has provided a selective advantage to increase the pathogenicity of the virus requires further research, as the virulence involves a complex interaction of several factors (Galvani, 2003; Peñaranda et al., 2011).

In order to determine the possible location of positively selected sites and lineage-specific positions on the IHNV glycoproteins, the consensus 3D structure of each lineage was predicted using VSV glycoprotein as the template. Amino acids number 49 to 436 of E-I and E-II ectodomains were presented in the predicted structures. In total, Rhabdoviridae glycoproteins have been shown to share a high degree of structural homology despite the low level of amino acid similarity (Walker and Kongsuwan, 1999; Baquero et al., 2013). The ectodomain part of glycoprotein in this family folds into three distinct domains including 1) the central domain or trimerization domain, which involves in the attachment of three monomer glycoproteins, 2) the fusion domain, which triggers the fusion between viral and endosomal membranes at low pH and subsequently releases of viral nucleocapsid into the cytoplasm, and 3) the pleckstrin homology domain which is very accessible in the final trimer structure and contains the major antigenic regions (Roche et al., 2007; Baquero et al., 2013). The main three domains have been demonstrated to retain their 3D-structures during the transition from pre-fusion to post-fusion states, while they may undergo rearrangement in their orientation toward the membrane. On the other hand, several hinge regions between the main domains are not structurally conserved and can refold during the membrane fusion process (Roche et al., 2008; Albertini et al., 2012). Interestingly, almost all of the positions that were identified to be under positive selection in the ectodomain part were mapped to the conserved structures of PHD domain on the top of the glycoprotein surface. Amino acid differences between E-I and E-II lineages at these sites did not result in structural changes. Four of the putative positions (247, 252, 256 and 257) were involved in the formation of two anti-parallel sheets while amino acids 276 and 277 were presented in a very exposed α-helix. This helix corresponds to the A2 segment of VSV glycoprotein which is the most exposed region of the glycoprotein and constitutes the main antigenic region of VSV glycoprotein and several other rhabdovirus glycoproteins (Benmansour et al., 1991; Kongsuwan et al., 1998; Roche et al., 2007). The conservation of structure in this exposed and hypervariable region in different IHNV virus variant may indicate a structurally important

function for this region. Consequently, the possible role of this antigenic patch in escaping from the host immune system may be accomplished via changes in biochemical properties of amino acids rather than changes in physical and spatial interactions.

In contrast to the positively selected sites, the lineage-specific positions in the ectodomain were mapped to either main domains or hinge regions of the glycoprotein. Amino acid variations at sites 91 (in hinge region R2) and 130 (in the FD domain) resulted in structural changes in the glycoprotein of E-I and E-II lineages, possibly providing a fitness advantage for interaction with host components or membrane integration. Whether these structural changes are associated with different evolution pattern of European IHNV isolates remains to be determined. Several studies, however, have proved that even a single fixed mutation may exert a "gain of function" effect such as increased pathogenicity or host shift and consequently results in genotype differentiation. For instance, Kim et al. (2014) found eight conserved amino acid substitutions between VHSV strains from genotypes Ia and III that are pathogenic to trout, and non-pathogenic marine strains from genotypes Ib and IV. However, using the reverse genetic approach, they showed that a single mutation in virus RNA polymerase (I1102F) is responsible for the pathogenicity of Ia and III genotypes in trout host. It would be of great interest to generate recombinant virus using site-directed mutagenesis method such as CRISPR/CAS system, to empirically address the impact of identified sites on virulence and host adaptively of IHNV isolates from each lineage.

5. Conclusion

In the present study, a total of 960 specimens were examined over a 3.5-year period to determine the epidemiological status and genetic diversity of IHNV in Iran. Our large-scale analysis indicated an average of 18.8% infection rate in trout rearing farms of 11 top trout producer provinces of Iran. Based on the full sequence of virus glycoprotein, all Iranian IHNV isolates were clustered within E genogroup and showed to be originated from EI subgroup. This implies the need for stricter biosecurity criteria for monitoring programs and eradication plans. Moreover, regarding the European origin of the obtained IHNV isolates, we combined phylogenetic, evolutionary and structural information to reveal the genetic components shaping the evolution patterns of European E-I and E-II subgroups. Such detailed information will enhance our understanding of the possible evolutionary mechanisms employed by the virus to escape the host immune system and will be beneficial for selecting the correct viral strains for developing more effective vaccines.

Declarations of interest

None.

Acknowledgments

The authors wish to thank Iranian Vice-President for Science and Technology, Isfahan University of Technology(IUT) Research Council and Research Institute for Biotechnology and Bioengineering for supporting this work.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.virol.2019.06.012.

References

Abbadi, M., Fusaro, A., Ceolin, C., Casarotto, C., Quartesan, R., Dalla Pozza, M., Cattoli, G., Toffan, A., Holmes, E.C., Panzarin, V., 2016. Molecular evolution and phylogeography of co-circulating IHNV and VHSV in Italy. Front. Microbiol. 7, 1–13.

Adel, M., Amiri, A.B., Dadar, M., Breyta, R., Kurath, G., Laktarashi, B., Ghajari, A., 2016. Phylogenetic relationships of Iranian infectious hematopoietic necrosis virus of rainbow trout (*Oncorhynchus mykiss*) based on the glycoprotein gene. Arch. Virol. 161, 657–663.

Ahmadivand, S., Soltani, M., Mardani, K., Shokrpoor, S., Hassanzadeh, R., Ahmadpoor, M., Rahmati-Holasoo, H., Meshkini, S., 2017. Infectious hematopoietic necrosis virus (IHNV) outbreak in farmed rainbow trout in Iran: viral isolation, pathological findings, molecular confirmation, and genetic analysis. Virus Res. 229, 17–23.

Albertini, A.A.V., Baquero, E., Ferlin, A., Gaudin, Y., 2012. Molecular and cellular aspects of rhabdovirus entry. Viruses 4, 117–139.

Alonso, M.C., Leong, J.A., 2013. Licensed DNA vaccines against infectious hematopoietic necrosis virus (IHNV). Recent Pat. DNA Gene Sequences 7, 62–65.

Ammayappan, A., LaPatra, S.E., Vakharia, V.N., 2010. A vaccinia-virus-free reverse genetics system for infectious hematopoietic necrosis virus. J. Virol. Methods 167, 132–139.

Baudin-Laurencin, F., 1987. IHN in France. Bull. Eur. Assoc. Fish Pathol. 7, 104.

Baquero, E., Albertini, A.A., Vachette, P., Lepault, J., Bressanelli, S., Gaudin, Y., 2013. Intermediate conformations during viral fusion glycoprotein structural transition. Curr. Opin. Virol. 3, 143–150.

Bedhomme, S., Hillung, J., Elena, S.F., 2015. Emerging viruses: why they are not jacks of all trades? Curr. Opin. Virol. 10, 1–6.

Bellec, L., Louboutin, L., Cabon, J., Castric, J., Cozien, J., Thiéry, R., Morin, T., 2017. Molecular evolution and phylogeography of infectious hematopoietic necrosis virus with a focus on its presence in France over the last 30 years. J. Gen. Virol. 98, 2438–2446.

Benmansour, A., Leblois, H., Coulon, P., Tuffereau, C., Gaudin, Y., Flamand, A., Lafay, F., 1991. Antigenicity of rabies virus glycoprotein. J. Virol. 65, 4198–4203.

Bouckaert, R.R., Drummond, A.J., 2017. bModelTest: Bayesian phylogenetic site model averaging and model comparison. BMC Evol. Biol. 17, 42. https://doi.org/10.1186/s12862-017-0890-6.

Bouckaert, R., Heled, J., Kuhnert, D., Vaughan, T., Wu, C.H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 10, e1003537. https://doi.org/10.1371/journal.pcbi.1003537.

Boudinot, P., Blanco, M., de Kinkelin, P., Benmansour, A., 1998. Combined DNA immunization with the glycoprotein gene of Viral Hemorrhagic Septicemia Virus and Infectious Hematopoietic Necrosis Virus Induces double-specific protective immunity and nonspecific response in rainbow trout. Virology 249, 297–306.

Bovo, G., Giorgetti, G., Jørgensen, P.E.V., Olesen, N.J., 1987. Infectious hematopoietic necrosis: first detection in Italy. Bull. Eur. Assoc. Fish Pathol. 7, 124.

Cattoli, G., Fusaro, A., Monne, I., Coven, F., Joannis, T., El-Hamid, H.S.A., Hussein, A.A., Cornelius, C., Amarin, N.M., Mancin, M., Holmes, E.C., Capua, I., 2011. Evidence for differing evolutionary dynamics of A/H5N1 viruses among countries applying or not applying avian influenza vaccination in poultry. Vaccine 29, 9368–9375.

Cha, S.J., Jung, Y.H., Lee, H.Y., Jung, J.Y., Cho, H.J., Park, M.S., 2012. Genotype distribution of infectious hematopoietic necrosis virus (IHNV) in Korea. J. Fish. Pathol. 25, 143–150.

Chrzastek, K., Lee, D., Gharaibeh, S., Zsak, A., Kapczynski, D.R., 2018. Characterization of H9N2 avian influenza viruses from the Middle East demonstrates heterogeneity at amino acid position 226 in the hemagglutinin and potential for transmission to mammals. Virology 518, 195–201.

Cieslak, M., Wahli, T., Diserens, N., Haenen, O.L.M., Schütze, H., 2017. Phylogeny of the infectious hematopoietic necrosis virus in European aquaculture. PLoS One 12, 1–12.

Corbeil, S., LaPatra, S.E., Anderson, E.D., Kurath, G., 2000. Nanogram quantities of a DNA vaccine protect rainbow trout fry against heterologous strains of infectious hematopoietic necrosis virus. Vaccine 18, 2817–2824.

Delport, W., Poon, A.F.Y., Frost, S.D.W., Kosakovsky Pond, S.L., 2010. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics 26, 2455–2457.

Dixon, P., Paley, R., Alegria-Moran, R., Oidtmann, B., 2016. Epidemiological characteristics of infectious hematopoietic necrosis virus (IHNV): a review. Vet. Res. 47, 1–26.

Dormitzer, P.R., Galli, G., Castellino, F., Golding, H., Khurana, S., Giudice, G., Rappuoli, R., 2011. Influenza vaccine immunology. Immunol. Rev. 239, 167–177.

Elena, S.F., Sanjuan, R., 2005. Adaptive value of high mutation rates of RNA viruses: separating causes from consequences. J. Virol. 79, 11555–11558.

Engelking, H.M., Leong, J.C., 1989. The glycoprotein of infectious hematopoietic necrosis virus elicits neutralizing antibody and protective responses. Virus Res. 13, 213–230.

Enzmann, P.J., Kurath, G., Fichtner, D., Bergmann, S.M., 2005. Infectious hematopoietic necrosis virus: monophyletic origin of European isolates from North American Genogroup M. Dis. Aquat. Org. 66, 187–195.

Fallahi, R., Soltani, M., Kargar, R., Zorriehzahra, J., Shchelkunov, I., Hemmatzadeh, F., Nouri, A., 2003. Isolation and identification of the infectious haematopoietic necrosis virus (IHNV)-like agent from farmed rainbow trout (*Oncorhynchus mykiss*) from Iran. Arch. Razi. Inst. 56, 37–45.

FAO/Food and Agriculture Organization, Fisheries and Aquaculture Department, 2016. FishStatJ - software for fisheries statistical time series. Available online at: http://www.fao.org/fishery/statistics/software/FishStatJ/en.

Forni, D., Cagliani, R., Mozzi, A., Pozzoli, U., Al-Daghri, N., Clerici, M., Sironi, M., 2016. Extensive positive selection drives the evolution of nonstructural proteins. In: In: Perlman, S. (Ed.), Lineage C Betacoronaviruses, vol. 90. pp. 3627–3639 J. Virol.

Galvani, A.P., 2003. Epidemiology meets evolutionary ecology. Trends Ecol. Evol. 18, 132–139.

Glavina, J., Román, Espada, de Prat-Gay, Chemes, Lucía B., Sánchez, I.E., 2018. Interplay between sequence, structure and linear motifs in the adenovirus E1A hub protein. Virology 525, 117–131.

Garver, K.A., LaPatra, S.E., Kurath, G., 2005. Efficacy of an infectious hematopoietic

necrosis (IHN) virus DNA vaccine in *Chinook Oncorhynchus* tshawytscha and sockeye *O. nerka* salmon. Dis. Aquat. Org. 64, 13–22.
Garver, K.A., Troyer, R.M., Kurath, G., 2003. Two distinct phylogenetic clades of infectious hematopoietic necrosis virus overlap within the Columbia River basin. Dis. Aquat. Org. 55, 187–203.
Ghorani, M., Adel, M., Dadar, M., Langeroudi, A.G., Kamyabi, R., Vakharia, V.N., Einer-Jensen, K., 2016. Phylogenetic analysis of the glycoprotein gene of viral hemorrhagic septicemia virus from Iranian trout farms points towards a common European origin. Vet. Microbiol. 186, 97–101.
Haenen, O.L.M., Schuetze, H., Cieslak, M., Oldenburg, S., Spierenburg, M.A.H., Roozenburg-Hengst, I., Voorbergen-Laarman, M., Engelsma, M.Y., Olesen, N.J., 2016. First evidence of infectious hematopoietic necrosis virus (IHNV) in The Netherlands. J. Fish Dis. 39, 971–979.
Harvey, W.T., Benton, D.J., Gregory, V., Hall, J.P.J., Daniels, R.S., Bedford, T., Haydon, D.T., Hay, A.J., McCauley, J.W., Reeve, R., 2016. Identification of low- and high-impact hemagglutinin amino acid substitutions that drive antigenic drift of influenza A(H1N1) viruses. PLoS Pathog. 12, e1005526.
He, M., Ding, N.Z., He, C.Q., Yan, X.C., Teng, C.B., 2013. Dating the divergence of the infectious hematopoietic necrosis virus. Infect. Genet. Evol. 18, 145–150.
Hu, D., Lv, L., Gu, J., Chen, T., Xiao, Y., Liu, S., 2016. Genetic diversity and positive selection analysis of classical swine fever virus envelope protein gene E2 in East China under C-strain vaccination. Front. Microbiol. 7, 85. https://doi.org/10.3389/fmicb.2016.00085.
Huang, C., Chien, M.S., Landolt, M., Batts, W., Winton, J., 1996. Mapping the neutralizing epitopes on the glycoprotein of infectious hematopoietic necrosis virus, a fish rhabdovirus. J. Gen. Virol. 77, 3033–3040.
ICTV/International Committee on Taxonomy of Viruses. Available onlineat: http://www.ictvonline.org/virusTaxonomy.asp.
Kim, S.H., Thu, B.J., Skall, H.F., Vendramin, N., Evensen, O., 2014. A single amino acid mutation (I1012F) of the RNA polymerase of marine viral hemorrhagic septicemia virus changes in vitro virulence to rainbow trout gill epithelial cells. J. Virol. 88, 7189–7198.
Kimura, M., 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16, 111–120.
Kolodziejek, J., Schachner, O., Dürrwald, R., Latif, M., Nowotny, N., 2008. "Mid-G" region sequences of the glycoprotein gene of Austrian infectious hematopoietic necrosis virus isolates form two lineages within European isolates and are distinct from American and Asian lineages. J. Clin. Microbiol. 46, 22–30.
Kongsuwan, K., Cybinski, D.H., Cooper, J., Walker, P.J., 1998. Location of neutralizing epitopes on the G protein of bovine ephemeral fever rhabdovirus. J. Gen. Virol. 79, 2573–2581.
Kratsch, C., Klingen, T.R., Mumken, L., Steinbruck, L., McHardy, A.C., 2016. Determination of antigenicity-altering patches on the major surface protein of human influenza A/H3N2 viruses. Virus Evol. 2https://doi.org/10.1093/ve/vev025. vev025.
Kryazhimskiy, S., Dushoff, J., Bazykin, G.A., Plotkin, J.B., 2011. Prevalence of epistasis in the evolution of influenza A surface proteins. PLoS Genet. 7, e1001301. https://doi.org/10.1371/journal.pgen.1001301.
Kurath, G., 2012. Fish novirhabdoviruses. In: Dietzgen, R.G. (Ed.), Rhabdoviruses: Molecular Taxonomy, Evolution, Genomics, Ecology, Host-Vector Interactions, Cytopathology and Control. Caister Academic Press, Norfolk, UK, pp. 89–116.
Kurath, G., Garver, K.A., Troyer, R.M., Emmenegger, E.J., Einer-Jensen, K., Anderson, E.D., 2003. Phylogeography of infectious hematopoietic necrosis virus in North America. J. Gen. Virol. 84, 803–814.
Lanave, C., Preparata, G., Saccone, C., Serio, G., 1984. A new method for calculating evolutionary substitution rates. J. Mol. Evol. 20, 86–93.
LaPatra, S.E., Corbeil, S., Jones, G.R., Shewmaker, W.D., Lorenzen, N., Anderson, E.D., Kurath, G., 2001. Protection of rainbow trout against infectious hematopoietic necrosis virus four days after specific or semi-specific DNA vaccination. Vaccine 19, 4011–4019.
LaPatra, S.E., Evilla, C., Winston, V., 2008. Positively selected sites on the surface glycoprotein (G) of infectiouss hematopoietic necrosis virus. J. Gen. Virol. 89, 703–708.
Liao, Y.C., Lin, H.H., Lin, C.H., 2013. Monitoring the antigenic evolution of human influenza A viruses to understand how and when viruses escape from existing immunity. BMC Res. Notes 6, 227. https://doi.org/10.1186/1756-0500-6-227.
Lorito, G., Hatzopoulos, S., Laurell, G., Campbell, K.C.M., Petruccelli, J., Giordano, P., Kochanek, K., Sliwa, L., Martini, A., Skarzynski, H., 2011. Dose-dependent protection on cisplatin-induced ototoxicity - an electrophysiological study on the effect of three antioxidants in the Sprague-Dawley rat animal model. Med. Sci. Monit. 17, BR179–186.
Morzunov, S.P., Winton, J.R., Nichol, S.T., 1995. The complete genome structure and phylogenetic relationship of infectious hematopoietic necrosis virus. Virus Res. 38, 175–192.
Neill, J.D., Workman, A.M., Hesse, R., Bai, J., Porter, E.P., Meadors, B., Anderson, J., Bayles, D.O., Falkenberg, S.M., 2019. Identification of BVDV2b and 2c subgenotypes in the United States: genetic and antigenic characterization. Virology 158, 19–20.
Nichol, S.T., Rowe, J.E., Winton, J.R., 1995. Molecular epizootiology and evolution of the glycoprotein and non-virion protein genes of infectious hematopoietic necrosis virus, a fish rhabdovirus. Virus Res. 38, 159–173.
Novella, I.S., Gilbertson, D.L., Borrego, B., Domingo, E., Holland, J.J., 2005. Adaptability costs in immune escape variants of vesicular stomatitis virus. Virus Res. 107, 27–34.
OIE/World Organization for Animal Health, 2016. Manual of diagnostic tests for aquatic animals. Available online at: http://www.oie.int/international-standard-setting/aquatic-manual/access-online/.
Park, J.W., Moon, C.H., Wargo, A.R., Purcell, M.K., Kurath, G., 2010. Differential growth of U and M type infectious hematopoietic necrosis virus in a rainbow trout-derived cell line, RTG-2. J. Fish Dis. 33, 583–591.
Pauszek, S.J., Barrera, J.D.C., Goldberg, T., Allende, R., Rodriguez, L.L., 2011. Genetic and antigenic relationships of vesicular stomatitis viruses from South America. Arch. Virol. 156, 1961–1968.
Peñaranda, M.M.D., Wargo, A.R., Kurath, G., 2011. In vivo fitness correlates with host-specific virulence of Infectious hematopoietic necrosis virus (IHNV) in sockeye salmon and rainbow trout. Virology 417, 312–319.
Peng, Y., Wang, D., Wang, J., Li, K., Tan, Z., Shu, Y., Jiang, T., 2017. A universal computational model for predicting antigenic variants of influenza A virus based on conserved antigenic structures. Sci. Rep. 7, 42051. https://doi.org/10.1038/srep42051.
Ping, J., Li, C., Deng, G., Jiang, Y., Tian, G., Zhang, S., Bu, Z., Chen, H., 2008. Single-amino-acid mutation in the HA alters the recognition of H9N2 influenza virus by a monoclonal antibody. Biochem. Biophys. Res. Commun. 371, 168–171.
Pontremoli, C., Forni, D., Cagliani, R., Filippi, G., Gioia, L.De, Pozzoli, U., Clerici, M., Sironi, M., 2016. Positive selection drives evolution at the host-filovirus interaction surface. Mol. Biol. Evol. 33, 2836–2847.
Poon, A.F.Y., Lewis, F.I., Pond, S.L.K., Frost, S.D.W., 2007. An evolutionary-network model reveals stratified interactions in the V3 loop of the HIV-1 envelope. PLoS Comput. Biol. 3, 1–12.
Purcell, M.K., Garver, K.A., Conway, C., Elliott, D.G., Kurath, G., 2009. Infectious hematopoietic necrosis virus genogroup-specific virulence mechanisms in sockeye salmon, *Oncorhynchus nerka* (Walbaum), from Redfish Lake, Idaho. J. Fish Dis. 32, 619–631.
Rajao, D.S., Anderson, T.K., Kitikoon, P., Stratton, J., Lewis, N.S., Vincent, A.L., 2018. Antigenic and genetic evolution of contemporary swine H1 influenza viruses in the United States. Virology 518, 45–54.
Rasmussen, D.A., Magnus, C., Bouckaert, R., 2015. Substitution Model Selection in BEAST2 A Tutorial Using bModelTest Practical: Selecting a Substitution Model Installing the bModelTest Package. pp. 1–10.
Roche, S., Albertini, A.A.V., Lepault, J., Bressanelli, S., Gaudin, Y., 2008. Structures of vesicular stomatitis virus glycoprotein: membrane fusion revisited. Cell. Mol. Life Sci. 65, 1716–1728.
Roche, S., Rey, F.A., Gaudin, Y., Bressanelli, S., 2007. Structure of the prefusion form of the vesicular stomatitis virus glycoprotein G. Science 315, 843–848.
Sandbulte, M.R., Westgeest, K.B., Gao, J., Xu, X., Klimov, A.I., Russell, C.A., Burke, D.F., Smith, D.J., Fouchier, R.A.M., Eichelberger, M.C., 2011. Discordant antigenic drift of neuraminidase and hemagglutinin in H1N1 and H3N2 influenza viruses. Proc. Natl. Acad. Sci. U.S.A. 108, 20748–20753.
Simonetti, F.L., Teppa, E., Chernomoretz, A., Nielsen, M., Marino Buslje, C., 2013. MISTIC: mutual information server to infer coevolution. Nucleic Acids Res. 41, W8–W14.
Sironi, M., Cagliani, R., Forni, D., Clerici, M., 2015. Evolutionary insights into host-pathogen interactions from mammalian sequence data. Nat. Rev. Genet. 16, 224–236.
Sun, H., Yang, J., Zhang, T., Long, L.P., Jia, K., Yang, G., Webby, R.J., Wan, X.F., 2013. Using sequence data to infer the antigenicity of influenza virus. mBio 4https://doi.org/10.1128/mBio.00230-13. e00230-13.
Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123, 585–595.
Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731–2739.
Tsetsarkin, K.A., Chen, R., Leal, G., Forrester, N., Higgs, S., Huang, J., Weaver, S.C., 2011. Chikungunya virus emergence is constrained in Asia by lineage-specific adaptive landscapes. Proc. Natl. Acad. Sci. U.S.A. 108, 7872–7877.
Verjan, N., Ooi, E.L., Nochi, T., Kondo, H., Hirono, I., Aoki, T., Kiyono, H., Yuki, Y., 2008. A soluble nonglycosylated recombinant infectious hematopoietic necrosis virus (IHNV) G-protein induces IFNs in rainbow trout (*Oncorhynchus mykiss*). Fish Shellfish Immunol. 25, 170–180.
Walker, P.J., Kongsuwan, K., 1999. Deduced structural model for animal rhabdovirus glycoproteins. J. Gen. Virol. 80, 1211–1220.
Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R., Schwede, T., 2018. SWISS-MODEL: homology modeling of protein structures and complexes. Nucleic Acids Res. 46, W296–W303.
Xu, L., Mourich, D.V., Engelking, H.M., Ristow, S., Arnzen, J., Leong, J.C., 1991. Epitope mapping and characterization of the infectious hematopoietic necrosis virus glycoprotein, using fusion proteins synthesized in Escherichia coli. J. Virol. 65, 1611–1615.
Xu, L.M., Liu, M., Zhao, J.Z., Cao, Y.S., Yin, J.S., Liu, H.B., Lu, T., 2014. Epitope mapping of the infectious hematopoietic necrosis virus glycoprotein by flow cytometry. Biotechnol. Lett. 36, 2109–2116.
Xu, W., Zhang, Z., Nfon, C., Yang, M., 2018. Genetic and antigenic relationship of foot-and-mouth disease virus serotype O isolates with the vaccine strain O1/BFS. Vaccine 36, 3802–3808.
Yang, Z., 2000. Maximum likelihood estimation on large phylogenies and analysis of adaptive evolution in human influenza virus A. J. Mol. Evol. 51, 423–432.