Journal of General Virology
September 2017, Volume 98, Pages 2438-2446
http://dx.doi.org/10.1099/jgv.0.000894
http://archimer.ifremer.fr/doc/00399/51015/

Molecular evolution and phylogeography of infectious hematopoietic necrosis virus with a focus on its presence in France over the last 30 years

Bellec Laure [1, 2], Louboutin Lénaïg [3], Cabon Joëlle [3], Castric Jeanne [3], Cozien Joelle [4], Thiéry Richard [5], Morin Thierry [3, *]

[1] IFREMER, Centre Brest, REM/EEP/LEP, ZI de la Pointe du Diable, CS10070, 29280 Plouzané, France
[2] IFREMER, Centre Brest, REM/EEP/LMEE, UMR6197, ZI de la Pointe du Diable, CS10070, 29280 Plouzané, France
[3] French Agency for Food, Environmental and Occupational Health and Safety (ANSES), Ploufragan-Plouzané Laboratory, Viral Fish Pathology Unit, National Reference Laboratory for Regulated Fish Diseases, Bretagne Loire University, Technopôle Brest-Iroise, BP 70, 29280 Plouzané, France
[4] IFREMER, Laboratoire Santé Environnement et Microbiologie (PDG-RBE-SG2M-LSEM), Technopôle Brest-Iroise, 29280 Plouzané, France
[5] French Agency for Food, Environmental and Occupational Health and Safety (ANSES), Sophia Antipolis Laboratory, 06902 Sophia-Antipolis, France

* Corresponding author : Thierry Morin, email address : Thierry.MORIN@anses.fr

Abstract :

Infectious hematopoietic necrosis virus (IHNV) is among the most important pathogens affecting the salmonid industry. Here, we investigated the molecular evolution and circulation of isolates from 11 countries or regions all over the world, with a special focus on the epidemiological situation in France. The phylogeography, time to the most recent common ancestor (TMRCA) and nucleotide substitution rate were studied using 118 full-length glycoprotein gene sequences isolated from 9 countries (5 genogroups) over a period of 47 years. The TMRCA dates back to 1943, with the L genogroup identified as the likely root (67%), which is consistent with the first report of this pathogen in the USA. A Bayesian inference approach was applied to the partial glycoprotein gene sequences of 88 representative strains isolated in France over the period 1987–2015. The genetic diversity of these 88 sequences showed mean nucleotide and amino-acid identities of 97.1 and 97.8%, respectively, and a d N/d S ratio (non-synonymous to synonymous mutations) of 0.25, indicating purifying selection. The French viral populations are divided into eight sub-clades and four individual isolates, with a clear spatial differentiation, suggesting the predominant role of local reservoirs in contamination. The atypical ‘signatures’ of some isolates underlined the usefulness of molecular phylogeny for epidemiological investigations that track the spread of IHNV.

Keywords : phylogeography, dates of divergence, molecular epidemiology, evolutionary dynamics, infectious hematopoietic necrosis virus

Introduction

Belonging to the *Salmonid novirhabdovirus* species of the *Novirhabdovirus* genus, *infectious hematopoietic necrosis virus* (IHNV) is the causative agent of a severe aquatic disease affecting wild and farmed salmonid species [1,2,3,4,5]. Structurally, the typically bullet-shaped virion encapsidates a non-segmented, negative-sense, single-stranded RNA of about 11 000 nucleotides [6]. The linear genome encodes six proteins in the following order (3′-5′): a nucleoprotein (N), a phosphoprotein (P), a matrix protein (M), a glycoprotein (G), a non-virion protein (NV), and a polymerase (L) [7,8,9,10].

The first outbreak of IHNV was detected in hatcheries in western North America in the 1950s [11]. Since then, the etiological agent has spread through North America, Asia and Europe — mostly due to the international trade of juvenile fish or eggs — and has become a major threat to the aquaculture industry. Detection of IHNV in Japan dates from 1971 and the virus was probably imported with a shipment of contaminated fish eggs from Alaska [12]. In Europe, the virus was reported for the first time in 1987 from two independent cases, one in France and one in Italy, and later in Germany (1992) [13,14,15]. Many outbreaks have since been reported around the world [16,17] and phylogenetic analyses based on the complete or partial sequence of the G gene, which shows relatively high genetic diversity compared with the other genes, have led to the definition of five major genogroups: U (upper), M (middle), L (lower), E (Europe) and J (Japanese rainbow trout) [10,18]. U, M and L correspond to the observed geographical range in North America [19,20,21]. Several diverse genotypes have been identified in Europe and seem closely related to the M genogroup [22]. The Japanese isolates show high divergence, but share a common origin with genogroup U [23,24].

Due to its high infectivity and wide distribution, IHNV represents a serious economic impact in aquaculture species such as rainbow trout (*Oncorhynchus mykiss*) and Atlantic salmon (*Salmo salar*). The associated disease is therefore listed as one of the regulated non-exotic

fish diseases in the European Union [25] and is notifiable to the World Organization for Animal Health (OIE) [26]. Historically based on virus isolation in susceptible cell cultures followed by confirmation steps, direct official methods for surveillance and diagnosis now also include RT-qPCR detection [27,28]. The extensive use of these methods for health monitoring and veterinary controls of international fish trade has strongly reduced the number of incidents associated with IHNV in Europe in the past few years. Recent surveillance data indicate that about 0.6% of the listed European fish farms are considered infected [29], but this number is probably underestimated, particularly because it is difficult to detect the virus in fish which are not in the most susceptible stage (juvenile) and/or in permissive conditions [30].

Here, we applied a Bayesian coalescent method to the five genogroups of IHNV to better understand the evolutionary history and phylogeography of this pathogen. In addition, genetic diversity and geographic distribution of 88 strains isolated in France over a period of 30 years were analysed for the first time and used to develop hypotheses for the different possible routes of viral spread among trout farms.

Results

Worldwide phylogeography of IHNV

A discrete phylogeographic analysis of IHNV was conducted on 118 complete G gene sequences from 11 locations (China, Croatia, France, Germany, Italy, Japan, South Korea, Switzerland, USA lower, USA middle and USA upper) and covering all recognised genogroups: J, E, U, M and L (Figure 1). The distribution clearly suggests that the USA Lower specimens constitute the root of the tree with over 67% probability. The topology defined three clear genogroups (L, E and M) and two others (U and J) that were more ambiguous. For example, U showed a basal position compared with J, but they clustered together. Genogroup J was composed of isolates from Japan, South Korea and China. Isolates from South Korea formed two groups with one Japanese isolate at a basal position each time. Recent (2012-2013) Chinese isolates constituted one group and appeared to emerge from Japan. The European group derived from the M genogroup, with a probable German origin which gave rise to separate infections in Italy, Switzerland, Croatia and France. Several introduction events involving French-related viruses in Germany or Switzerland were also observed.

The nucleotide substitution rates (per site and per year) for the G gene was estimated to be 7.14×10^{-4} substitutions (subs) $site^{-1}$ $year^{-1}$ and the emergence of IHNV was estimated to date from 1943 (Table 1). For a more accurate analysis, the most recent common ancestor (TMRCA) and mean genetic diversity were also calculated for the five genogroups (Table 2). The oldest genogroup was the L genogroup and the youngest genogroup appeared to be the E genogroup whose divergence dates back to 1981.

Phylogenetic analysis of French isolates

We sequenced a region of the G gene from 88 representative French IHNV isolates collected by our National Reference Laboratory (NRL) and covering a 30 year period (1987-2015) (Figure 2). A dataset of a partial G-gene sequence of 570 bp was created, which allowed to group the 88 French isolates plus 5 outgroups (from the M genogroup), and used for Bayesian phylogenetic reconstruction (Figure 3). The structure of the French isolates in the phylogeny show 8 sub-clades (6 with posterior support > 0.75) and 4 individual isolates. Two main monophyletic group including both 23 French viruses (sub-clades A and E) and defined by posterior support of 0.75 and 1 respectively were observed. The other genetic sub-clades (B, C, D, F, G, H) were smaller and composed of 3 to 15 sequences. Interestingly, isolates corresponding to the first French outbreak (71_87_1987) and the most recent one (SA15.1442_2015) didn't cluster with other sequences. Temporal link within sub-clades is not obvious, for example the two most abundant sub-clades A and E have large range of isolation date with 1999-2014 and 1990-2005, respectively. Though, we notice that viruses were isolated from three distinct areas in France: North (orange), East (purple) and Centre and South (Blue) (Figure 3). All viruses from the East area belong to the sub-clade A except one strain (isolate 016650_2007) suggesting that all viruses from East are derived and spread from one common infection. IHNV isolates from Centre and South form three different sub-clades (B, C and H). Isolates from North seem to be the more divergent with 4 sub-clades (D, E, F and G) plus 4 individual isolates representing multiple independent introduction or parallel evolution of separate viral lineage. However, some strains had genetic profiles that were not related to the geographical distribution of strains isolated in the respective area time period. For example, the isolate 503_2000 from the North showed only 97.8% nucleotide identity with two other strains from the same area, but had 100% identity with three isolates (Y6_2000; J13325_1999 and K2165-2000) from the East group. In another example, isolate

016650_2007, identified in the East area, near the French-Swiss border, clustered with the North group, sharing 100% nucleotide identity with an isolate (LDOn1_2007) of this group, despite a distance of more than 600 km between the groups. Similarly, isolate M13316_2002, isolated from fish in the area of the South and Centre, was 100% identical to a strain (L5889_2001) isolated one year before in the East area.

The nucleotide sequence identity ranged from 95.4 to 100%, with a mean value of 98.1%, for all isolates considered (Table 3).

Genetic diversity

Worldwide sequences of *infectious hematopoietic necrosis virus* displayed low genetic diversity for the G gene, with mean nucleotide diversity of 3.4%; within the five genogroups, diversity reached a maximum of 3.5% and a minimum of 1.3% (Table 2). Nevertheless, nucleotide diversity between the five genogroups showed that the Japanese group was the most distant, with diversity ranging from 4.2 to 5.6% with respect to each of the four other groups. Genetic diversity, the d_N/d_S ratio (non-synonymous to synonymous mutations) and Tajima's *D* (test based on polymorphism frequencies) were calculated for our 88 French isolates (Table 3). The partial G region displayed more synonymous than non-synonymous changes (ratio <1), suggesting purifying selection on the viral population. Tajima test was negative, indicating negative selection on the viral population.

Discussion

Infectious haematopoietic necrosis (IHN) is one of the most economically important viral diseases in farmed rainbow trout in Europe, and in most parts of the Northern hemisphere. Here, we performed a discrete phylogeographic analysis on 118 sequences from the five genogroups of IHNV using BEAST2. Further, we conducted, for the first time, an extensive study on the genetic diversity and geographic distribution of 88 strains collected by our laboratory, the French NRL. These strains are representative of all outbreaks reported since the first description of the disease in rainbow trout farms in 1987 up until 2015. Our results (i) confirm the origin and the history of spread of this worldwide virus; (ii) demonstrate a correlation between molecular phylogeny and geographical distribution in France; and (iii) illustrate how phylogeny can contribute to epidemiological investigations.

During the last six decades, extensive phylogenetic studies on IHNV have led to the identification of five genogroups that clearly correlate with specific geographical origins [10]. Our analysis of the evolutionary history (phylogeography and date of divergence) of these genogroups confirms that the L genogroup is located at the putative root (origin) of the tree and the U and M diverged from this group early on. The finding that the L genogroup is at the origin of the phylogenetic evolution is consistent with the first reported outbreaks in hatcheries in the state of Washington in the early 1950s [11] and with previous studies on IHNV [18,20,22]. A second significant divergence event involves the derivation of the J genogroup from the U genogroup presumably in Japan via an import from Alaska in the mid-1970s. Since then, Japanese strains have probably been introduced in Korea [31,32] and China [33]. Infections with isolates from the U genogroup in Japan were also observed. The youngest genogroup, the E genogroup (1981), derived from genogroup M and have spread in different European countries [21,22,34]. It was suggested that all European isolates were derived from the first introductions of the virus in France and Italy in 1987 [22]. Our analysis

198 suggests that Germany could also be a potential initial source of introduction despite a latter
199 detection (Figure 1). The first representative outbreaks in France are clearly associated to an
200 individual sequence (71_87_1987 in Figure 3), as it was previously shown in a phylogenetic
201 tree of *Salmonid novirhabdovirus* in Europe [21]. The first sequence in Italy has a basal
202 position belonging to a large clade (IO-87_FJ711518, Figure 1). This finding was observed
203 before by Enzmann and co-authors [21] and confirmed by a recent study on Italian IHNV
204 isolates [34]. Although the origin of genogroup E remains to be identified through further
205 investigations, IHNV circulation within the European Union appears very clearly and
206 demonstrates the difficulty of controlling the spread of this virus.
207 Significant efforts for health monitoring and veterinary control of the spread of IHN and Viral
208 Haemorrhagic Septicemia (VHS) have been undertaken over the past years in Europe
209 [25,28,35]. Overall, they have reduced the number of outbreaks and have helped improve the
210 availability of molecular characterization data for these viruses. A high capacity of IHNV to
211 circulate in continental Europe can be assumed from the phylogeny and topology of the E
212 genogroup, whereby there is some evidence for a spread starting from German and Italian
213 isolates. The virus has been demonstrated in previously IHN-free countries such as Croatia or
214 the Netherlands but has never been detected so far in UK, Ireland and Scandinavia [21]. It still
215 represents a serious threat to the large salmon-producing countries of North Europe [17]. This
216 latent dissemination may be due to the difficulty of observing clinical signs during veterinary
217 controls, particularly if the more susceptible juvenile stages are protected in farms by careful
218 containment from sub-adult and adult stages in which infection is generally latent and silent
219 [30,36,37]. Subclinical infections are also associated with lower virus levels in affected fish
220 compared to fish undergoing clinical infection and can lead to false negative diagnosis results
221 [38]. This ability of the disease to go unnoticed (compared with VHS for example) probably

leads to underestimation of the number of infected farms in Europe. Violations of European regulations can also not be excluded.

In France, phylogenetic analysis of partial G gene sequences of representative isolates collected over the past 30 years is in line with the presumed date of introduction [13]. Observed nucleotide and amino-acid diversities are low and correspond to a purifying selection process, stabilizing the viral population [34]. The most striking finding was the clear spatial distribution of genetically related isolates in three areas. This pattern of outbreaks, corresponding to the first diffusive patterns for IHNV spread in Italy described by Abbadi and colleagues, suggests the existence of local reservoirs around which recontaminations regularly occur through passive diffusion via water, fomites (landing nets, boots, vehicles) or piscivorous birds [34,39,40]. In this context of relative genetic homogeneity, the detection of isolates with a divergent genetic "signature" strongly suggests a transfer of isolates from other area with diffusion mainly associated with fish movement and trade practices [34]. In all cases, genetic data analysis provides information of interest to guide epidemiological investigations and to take the most appropriate health measures.

Our estimate of the evolutionary dynamics of IHNV is in the range of previously published data (7.14×10^{-4} subs site^{-1} year^{-1} in our study, 8.1×10^{-4} subs site^{-1} year^{-1}, [18]; 12×10^{-4} subs site^{-1} year^{-1}, [41] and 11×10^{-4} subs site^{-1} year^{-1}, [34]) and in agreement with RNA viruses that evolve quickly with rate between 10^{-2} and 10^{-5} subs site^{-1} year^{-1} [42,43]. Our results on the G gene are also consistent with findings in other fish viruses, such as the *Anguillid rhabdovirus* with a rate of 4.23×10^{-4} subs site^{-1} year^{-1} [44], VHSV with 5.91×10^{-4} subs site^{-1} year^{-1} [45], or the *Spring viremia of carp virus* with 5.47×10^{-4} subs site^{-1} year^{-1} [46].

Host switch is an important parameter for IHNV evolutionary dynamics. The J and M genogroups of IHNV showed higher genetic diversity compared with the three other genogroups (L, U, E). The lower viral genetic diversity observed for strains from genogroups

U and L, mostly found in Pacific salmon, is indicative of an evolutionary equilibrium. For the J genogroup, we confirm hypothesises that the host switch (from Pacific salmon (genogroup U) to rainbow trout) is associated with virus adaptation through an increase in genetic diversity. Evolution in the G gene of isolates from farmed rainbow trout has been reported to be up to six-fold higher than that from salmonid fish having an ocean migration phase [20,47,48].

This study brings new elements to the understanding of evolutionary history of IHNV. In France, the epidemiological situation is relatively stable due to the implementation of European legislation since the early 2000s. A national monitoring and eradication plan, also covering VHSV, is currently being set up. It will probably lead to the discovery of new isolates previously difficult to detect on the basis of clinical observations. This molecular evolutionary work on IHNV in France has provided new data on how it is spread and will help enhance the control strategy.

Materials and methods

Virus isolates and dataset compilation

A dataset of 118 complete G genes from the five IHNV genogroups was extracted from GenBank (all sequences available to date). Dates and country of isolation were established based on the available literature [19,21,23,31,49,50].

A dataset of 88 partial G genes was used in this study to represent IHNV isolates collected from France during the 1987-2015 period. Isolate, geographic area, isolation dates and GenBank accessions (nos. XX to XX) are reported in Table 4.

RNA extraction, PCR amplification and sequencing

RNA was extracted from various internal organs such as the anterior kidney, spleen, heart, brain or from cell culture supernatants using the Nucleospin RNA virus kit (Macherey-Nagel, France). The partial glycoprotein gene, sequence of 693 bp, was amplified using the following primers, 5'-AGAGATCCCTACACCAGAGAC-3' and antisense primers 5'-GGTGGTGTTGTTTCCGTGCAA-3' [51]. Reverse-transcription (RT) and amplification were performed with the SuperScript III One-Step RT-PCR System (Invitrogen, France) using the following mix: approximately 1 µg of RNA was added to 20 µM of each primer, 1 µl of high-fidelity Platinum Taq, 25 µl of reaction mix (2X) and water in a final volume of 50 µl. The G-gene RT-PCRs were conducted in a Mastercycler (Eppendorf, France) with an initial step of 52°C (30 min) followed by one step at 94°C (2 min) then 40 cycles of 94°C (15 s), 52°C (30 s) and 68°C (60 s), and a final extension at 68°C (8 min). After agarose gel electrophoresis, PCR products of interest were purified with a NucleoSpin gel and PCR clean-up kit (Macherey-Nagel, France) then cloned using the TOPO TA cloning kit (Invitrogen, France). For each PCR product, three clones were selected and sequenced in both orientations using

286 the Sanger method and a 3130 Genetic Analyser (Applied Biosystems). Sequence files were
287 analysed visually using VectorNTI v.11.5 software.
288
289 **Phylogenetic reconstruction**
290 An alignment of partial G DNA sequences was performed with Muscle using SeaView 4 [52].
291 Phylogenetic reconstruction was done using Bayesian inference (BI) from an alignment of 93
292 sequences of 570 bp. Partition schemes and evolutionary models were selected via Bayesian
293 Information Criterion calculated in PARTITION FINDER v1.1.1[53]: for 1^{st} and 2^{nd} codon
294 partitions the K81 + I + G model were used; for 3^{rd} codon partition a model K81 + G. BI was
295 carried out done with MrBayes 3.2.6 [54] on the CIPRES Science Gateway [55], with four
296 chains of 5 x 10^6 generations, trees sampled every 500 generations, and burn-in value set to
297 20 % of the sampled trees. We checked that standard deviation of the split frequencies fell
298 below 0.01 and confirmed convergence of the runs to ensure convergence in tree search using
299 the Tracer v1.6 software (http://tree.bio.ed.ac.uk/software/tracer/). The tree was visualized
300 using FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).
301
302 **Ancestral reconstruction and discrete phylogeography**
303 The BEAST software package v2.1.3 was used to estimate date divergence and discrete
304 phylogeography from 118 full-length IHNV G gene sequences [56]. We used discrete
305 phylogeography to perform ancestral reconstruction on a single character (i.e. location) and to
306 determine the geographic location for the root of the tree [57]. The dataset was analysed using
307 a HKY + G model under a relaxed uncorrelated exponential molecular clock. This model was
308 evaluated with the coefficient of variation (CoV) [58], where a CoV value > 0 was considered
309 as evidence of non-clocklike evolutionary behaviour. The "exponential coalescent tree prior"
310 was used to infer the complex population dynamics of IHNV. Three independent Bayesian

311 MCMC runs were carried out for 30 million generations (to obtain effective sample size
312 values of at least 200 for each parameter), to retain a sample of 10 000 trees. Convergences of
313 the runs were confirmed using Tracer v1.5, MCC tree was generated using TreeAnnotator
314 v1.5 and visualized using FigTree v1.4.
315
316 **Genetic diversity**
317 Basic population statistics and single nucleotide polymorphisms were calculated using two
318 programs: DnaSp v5 [59,60] and MEGA v6 [61].
319

Funding information

This work received no specific grant from any funding agency.

Acknowledgments

We are grateful to the following laboratories of the French surveillance network for providing IHNV isolates for characterization: Laboratoire des Pyrénées et des Landes (LPL), Laboratoire Départemental d'Analyses du Jura, Labocea, Laboratoire Départemental Vétérinaire de l'Hérault, Laboratoire Départemental de l'Orne, Laboratoire Départemental d'Analyses du Pas de Calais, Laboratoire Agro-Vétérinaire Départemental de Seine Maritime. We also thank the ANSES sequencing facility (Unit Viral genetic and biosecurity) located in Ploufragan and D. Pottratz, translator in the "External Communication Unit" at Anses, for editorial assistance.

Conflicts of Interest

The authors declare no conflicts of interest.

Ethical statement

Not applicable

References

1. **EFSA.** Scientific Opinion of the Panel on Animal Health and Welfare (Question No EFSA-Q-2007-044). Possible vector species and live stages of susceptible species not transmitting disease as regards certain fish diseases. *The EFSA Journal* 2007;584: 1-16.

2. **Bootland LM, Leong JAC.** Infectious haematopoietic necrosis virus. *Fish diseases and disorders* 2011;3:66-109.

3. **Hoffmann B, Beer M., Schütze H, Mettenleiter TC.** *Fish rhabdoviruses: molecular epidemiology and evolution.* In: Springer Berlin Heidelberg. *The World of Rhabdoviruses; 2005.* pp. 81-117.

4. **Stone DM, Kerr R, Hughes M, Radford AD, Darby AC.** Characterisation of the genomes of four putative vesiculoviruses: tench rhabdovirus, grass carp rhabdovirus, perch rhabdovirus and eel rhabdovirus European X. *Arch Virol* 2013;158:2371-2377.

5. **International Committee on Taxonomy of Viruses.** https://data.ictvonline.org/proposals/2016.018aM.A.v1.Novirhabdovirus_spren.pdf. Accessed on 29/06/2017.

6. **Pringle CR.** Virus taxonomy–1999. *Arch Virol* 1999;144:421-429.

7. **Kurath G, Ahern KG, Pearson GD, Leong JC.** Molecular cloning of the six mRNA species of infectious hematopoietic necrosis virus, a fish rhabdovirus, and gene order determination by R-loop mapping. *J Virol* 1985;53:469-476.

8. **Morzunov SP, Winton JR, Nichol ST.** The complete genome structure and phylogenetic relationship of infectious hematopoietic necrosis virus. *Virus Res* 1995;38:175-192.

9. **Schütze H, Enzmann PJ, Kuchling R, Mundt E, Niemann H, Mettenleiter TC.** Complete genomic sequence of the fish rhabdovirus infectious haematopoietic necrosis virus. *J Gen Virol* 1995;76:2519-2527.

10. **Kurath G.** *Fish novirhabdoviruses*. In: *Rhabdoviruses: molecular taxonomy, evolution, genomics, ecology, host-vector interactions*; 2012. pp. 89-116.

11. **Rucker RR, Whipple WJ, Parvin, JR, Evans CA.** A contagious disease of salmon, possibly of virus origin. *Fishery bulletin. United States Fish and Wildlife Service* 1953;54:35-46.

12. **Sano T, Nishimura T, Okamoto N, Yamazaki T, Hanada H, Watanabe Y.** Studies on viral diseases of Japanese fishes. VI. Infectious hematopoietic necrosis (IHN) of salmonids in the mainland of Japan. *Journal of the Tokyo University of Fisheries* 1977;63:81-85.

13. **Baudin-Laurencin F.** IHN in France. *Bull Eur Ass of Fish Pathol* 1987;7:104.

14. **Bovo G, Giorgetti G, Jørgensen PEV, Olesen NJ.** Infectious haematopoietic necrosis: first detection in Italy. *Bull Eur Ass of Fish Pathol* 1987;7:124.

15. **Enzmann PJ, Dangschat H, Feneis B, Schmitt D, Wizigmann G *et al.*** Demonstration of IHN virus in Germany. *Bull Eur Ass of Fish Pathol* 1992;12:185.

16. **Adel M, Amiri AB, Dadar M, Breyta R, Kurath G. *et al.*** Phylogenetic relationships of Iranian infectious hematopoietic necrosis virus of rainbow trout (*Oncorhynchus mykiss*) based on the glycoprotein gene. *Arch Virol* 2016;161:657-663.

17. **Haenen OL, Schuetze H, Cieslak M, Oldenburg S, Spierenburg MA *et al.*** First evidence of infectious hematopoietic necrosis virus (IHNV) in the Netherlands. *J Fish Dis* 2016;39:971-979.

18. **He M, Ding NZ, He CQ, Yan XC, Teng CB.** Dating the divergence of the infectious hematopoietic necrosis virus. *Infec Genet Evol* 2013;18:145-150.

19. **Nichol ST, Rowe JE, Winton JR.** Molecular epizootiology and evolution of the glycoprotein and non-virion protein genes of infectious hematopoietic necrosis virus, a fish rhabdovirus. *Virus Res* 1995;38:159-173.

20. **Kurath G, Garver KA, Troyer RM, Emmenegger EJ, Einer-Jensen K *et al.*** Phylogeography of infectious haematopoietic necrosis virus in North America. *J Gen Virol* 2003;84:803-814.

21. **Enzmann PJ, Castric J, Bovo G, Thiery R, Fichtner D *et al.*** Evolution of infectious hematopoietic necrosis virus (IHNV), a fish rhabdovirus, in Europe over 20 years : implications for control. *Dis Aquat Organ* 2010;89:9-15.

22. **Enzmann PJ, Kurath G, Fichtner D, Bergmann SM.** Infectious hematopoietic necrosis virus : monophyletic origin of European isolates from North American Genogroup M. *Dis Aquat Organ* 2005;66:187-195.

23. **Nishizawa T, Kinoshita S, Kim WS, Higashi S, Yoshimizu M.** Nucleotide diversity of Japanese isolates of infectious hematopoietic necrosis virus (IHNV) based on the glycoprotein gene. *Dis Aquat Organ* 2006;71:267-272.

24. **Wang C, Zhao LL, Li YJ, Tang LJ, Qiao XY *et al.*** Analysis of the genome sequence of infectious hematopoietic necrosis virus HLJ-09 in China. *Virus Genes* 2016;52:29-37..

25. **European Commission.** Council Directive 2006/88/EC of 24 October 2006 on animal health requirements for aquaculture animals and products thereof, and on the prevention and control of certain diseases in aquatic animals. *Official Journal of the European Union* 2006; L328:14-56.

26. **World Organization for Animal Health OIE**. http://www.oie.int/en/animal-health-in-the-world/oie-listed-diseases-2017/ Accessed on 17/03/2017.

27. **Purcell MK., Hart SA, Kurath G, Winton JR.** Strand-specific, real-time RT-PCR assays for quantification of genomic and positive-sense RNAs of the fish rhabdovirus, Infectious hematopoietic necrosis virus. *J Virol Methods* 2006;132:18-24.

28. **European Commission.** Commission Implementing Decision (EU) 2015/1554 of 11 September 2015 laying down rules for the application of Directive 2006/88/EC as regards

requirements for surveillance and diagnostic methods. *Official Journal of the European Union* 2015;L247:1-62.

29. **EURL.** Overview of the disease situation and surveillance in Europe in 2014. http://www.eurl-fish.eu/activities/survey_and_diagnosis. Accessed on 17/03/2017.

30. **Bergmann SM, Fichtner D, Skall HF, Schlotfeldt HJ, Olesen NJ.** Age- and weight-dependent susceptibility of rainbow trout *Oncorhynchus mykiss* to isolates of infectious haematopoietic necrosis virus (IHNV) of varying virulence. *Dis Aquat Organ* 2003,55:205-210.

31. **Kim WS, Oh MJ, Nishizawa T, Park JW, Kurath G *et al.*** Genotyping of Korean isolates of infectious hematopoietic necrosis virus (IHNV) based on the glycoprotein gene. *Arch Virol* 2007;152:2119-2124.

32. **Kim KI, Cha SJ, Lee C, Baek H, Hwang SD, Cho MY, Jee BY, Park MA.** Genetic relatedness of infectious hematopoietic necrosis virus (IHNV) from cultured salmonids in Korea. *Arch Virol* 2016;161:2305-2310.

33. **Niu L, Zhao Z.** The epidemiological IHN and IPN of rainbow trout in northeast China. *Journal of Fisheries of China* 1988;12:351-352.

34. **Abbadi M, Fusaro A, Ceolin C, Casarotto C, Quartesan R *et al.*** Molecular Evolution and Phylogeography of Co-circulating IHNV and VHSV in Italy. *Front Microbiol* 2016;7**:** 1306.

35. **European Commission.** Commission Decision of 22 February 2001 laying down the sampling plans and diagnostic methods for the detection and confirmation of certain fish diseases and repealing Decision 92/532/EEC. Official Journal of the European Union 2001; L067:65-76.

439 36. **LaPatra SE, Groberg WJ, Rohovec JS, Fruer JL.** Size-related susceptibility of
440 salmonids to 2 strains of infectious hematopoietic necrosis virus. *Transactions of the*
441 *American Fisheries Society* 1990;119:25-30.
442 37. **LaPatra SE.** Factors affecting pathogenicity of infectious hematopoietic necrosis virus in
443 seawater. *Dis Aquat Organ* 1998;16:111-114.
444 38. **Miller TA, Rapp J, Wastlhuber U, Hoffmann RW, Enzmann PJ.** Rapid and sensitive
445 reverse transcriptase-polymerase chain reaction based detection and differential diagnosis of
446 fish pathogenic rhabdoviruses in organ samples and cultured cells. *Dis Aquat Organ* 1998;34:
447 13-20.
448 39. **Lyngstad TM, Hjortaas MJ, Kristoffersen AB, Markussen T, Karlsen E *et al.*** Use of
449 molecular epidemiology to trace transmission pathways for infectious salmon anaemia virus
450 (ISAV) in Norwegian salmon farming. *Epidemics* 2011;3:1-11.
451 40. **Lyngstad TM, Kristoffersen AB, Hjortaas MJ, Devold M, Aspehaug V *et al.*** Low
452 virulent infectious salmon anaemia virus (ISAV-HPR0) is prevalent and geographically
453 structured in Norwegian salmon farming. *Dis Aquat Organ* 2012;101:197-206.
454 41. **Troyer RM, Kurath G.** Molecular epidemiology of infectious hematopoietic necrosis
455 virus reveals complex virus traffic and evolution within southern Idaho aquaculture. *Dis*
456 *Aquat Organ* 2003;55:175-185.
457 42. **Jenkins GM, Rambaut A, Pybus OG, Holmes EC.** Rates of molecular evolution in
458 RNA viruses: a quantitative phylogenetic analysis. *J Mol Evol* 2002;54:156-165.
459 43. **Duffy S, Shackelton LA, Holmes EC.** Rates of evolutionary change in viruses: patterns
460 and determinants. *Nat Rev Genet* 2008;9:267-276.
461 44. **Bellec L, Cabon J, Bergmann S, de Boisseson C, Engelsma M, Haenen O, Morin T,**
462 **Olesen NJ, Schuetze H, Toffan A, Way K, Bigarré, L.** Evolutionary dynamics and genetic
463 diversity from three genes of Anguillid rhabdovirus. *J Gen Virol* 2014;95:2390-2401.

464 45. **He M, Yan XC, Liang Y, Sun XW, Teng CB.** Evolution of the viral hemorrhagic
465 septicemia virus: divergence, selection and origin. *Mol Phylogenet Evol* 2014;**77:**34-40.
466 46. **Padhi A, Verghese B.** Molecular evolutionary and epidemiological dynamics of a highly
467 pathogenic fish rhabdovirus, the spring viremia of carp virus (SVCV). *Vet Microbiol*
468 2012;15:54-63.
469 47. **Troyer RM, LaPatra SE, Kurath G.** Genetic analyses reveal unusually high diversity of
470 infectious haematopoietic necrosis viru**s** in rainbow trout aquaculture. *J Gen Virol* 2000;81:
471 2823-2832.
472 48. **Garver KA, Troyer RM, Kurath G.** Two distinct phylogenetic clades of infectious
473 hematopoietic necrosis virus overlap within the Columbia River basin. *Dis Aquat Organ*
474 2003;55:187-203.
475 49. **Anderson ED, Engelking HM, Emmenegger EJ, Kurath G.** Molecular epidemiology
476 reveals emergence of a virulent infectious hematopoietic necrosis (IHN) virus strain in wild
477 salmon and its transmission to hatchery fish. *Journal of Aquatic Animal Health* 2000;12:85-
478 99.
479 50. **Garver KA, Batts WN, Kurath G.** Virulence comparisons of infectious hematopoietic
480 necrosis virus U and M genogroups in sockeye salmon and rainbow trout. *Journal of aquatic*
481 *animal health* 2006;18:232-243.
482 51. **Emmenegger EJ, Meyers TR, Burton TO, Kurath G.** Genetic diversity and
483 epidemiology of infectious hematopoietic necrosis virus in Alaska. *Dis Aquat Organ* 2000;40:
484 163-176.
485 52. **Gouy M, Guindon S, Gascuel O.** SeaView version 4: a multiplatform graphical user
486 interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* 2010;27:221-
487 224.

488 53. **Lanfear R, Calcott B, Ho SYW, Guindon S.** Partitionfinder: combined selection of
489 partitioning schemes and substitution models for phylogenetic analyses. *Mol Biol Evol* 2012;
490 29:1695-1701.
491 54. **Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A *et al.*** 3.2: efficient
492 Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*
493 2012;61:539-542.
494 55. **Miller MA, Pfeiiffer W, Schwartz T.** Creating the CIPRES Science Gateway for
495 inference of large phylogenetic trees. In: *Proceedings of the Gateway Computing*
496 *Environments Workshop (GCE);* New Orleans, LA; 2010. pp. 1-8.
497 56. **Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH *et al.*** BEAST 2: A Software
498 Platform for Bayesian Evolutionary Analysis. *PLoS Computational Biology* 2014;10.
499 57. **Drummond AJ, Bouckaert RR.** Bayesian evolutionary analysis with BEAST.
500 Cambridge University Press; 2015.
501 58. **Drummond AJ, Ho SY, Phillips MJ, Rambaut A.** Relaxed phylogenetics and dating
502 with confidence. *PLoS Biol* 2006;4:e88.
503 59. **Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R.** DnaSP, DNA polymorphism
504 analyses by the coalescent and other methods. *Bioinformatics* 2003;19:2496-2497.
505 60. **Librado P, Rozas J.** DnaSP v5: a software for comprehensive analysis of DNA
506 polymorphism data. *Bioinformatics* 2009;25:1451-1452.
507 61. **Tamura K, Stecher G, Peterson D, Filipski A, Kumar S.** MEGA6: molecular
508 evolutionary genetics analysis version 6.0. *Molecular biology and evolution* 2013;30:2725-
509 2729.
510
511
512

513 **TABLES**

514

515 **Table 1**. Details of dataset and estimates related to the IHNV glycoprotein (G) gene. HPD:
516 highest probability density; TMRCA: time to the most recent common ancestor.

517

Parameter	G
Sequence length (nt)	1518
No. of sequences	118
Time span	1966–2013
Mean substitution rate, $\times 10^{-4}$ site^{-1} year^{-1} (95% HPD)	7.14 (5.2–9.13)
TMRCA, years (95% HDP)	1943 (1918–1961)
Coefficient of variation (95% HPD)	0.94 (0.84–1.04)
Mean genetic diversity (%)	3.4

518

519

520 **Table 2**. Estimates of diversity and divergence for the glycoprotein (G) according to the five
521 genogroups of IHNV. HPD: highest probability density; TMRCA: time to the most recent
522 common ancestor.

523

Genogroup	No. of sequences	TMRCA, years (95% HPD)	Mean genetic diversity (%)
L	5	1954 (1941-1966)	1.4
U	11	1965 (1957-1970)	1.3
M	7	1970 (1962-1977)	2.5
J	21	1977 (1974-1978)	3.5
E	74	1981 (1976-1985)	1.4

524

525

526

527

528

529 **Table 3**. Molecular data for partial G sequences from IHNV. N, Number; Nt ID, mean
530 nucleotide identity; AA ID, mean amino-acid identity; $d_{N,}$ number of non-synonymous
531 mutations; $d_{S,}$ number of synonymous mutations.

532

Parameter	Partial G
No. of sequences	88
No. of nucleotides (amino acids)	570 (190)
% Nt ID, (min-max)	98.15 (95.4-100)
% AA ID, (min-max)	97.82 (93.1-100)
% variable nucleotides	20.17
d_N/d_S	0.247
Tajima's *D*	-1.85

533

534

535

536

537

538

539

540

541

542

543

544

545

546

547

548

549

550

551 **Table 4**. French IHNV isolates used for partial glycoprotein gene (partial G region) analysis.

Virus isolate	Date of isolation	Geographic area	GenBank accession	Virus isolate	Date of isolation	Geographic area	GenBank accession
SA15	06.2015	Seine-Maritime		1071	11.2000	Pas-De-Calais	
331901	02.2014	Doubs		1022	11.2000	Pas-De-Calais	
VP0614	03.2014	Manche		SA000	12.2000	Vienne	
048478	09.2010	Pas-De-Calais		3LOT1	12.2000	Vaucluse	
02707	02.2008	Haute Savoie		X3	01.1999	Manche	
7445	04.2008	Ain		X24	02.1999	Nièvre	
T00121	02.2007	Drôme		V445	08.1999	Pas-De-Calais	
250501	05.2007	Ain		X255	11.1999	Puy-De-Dôme	
FF56	05.2007	Doubs		J13325	11.1999	Savoie	
LDOn1	11.2007	Orne		V1777	11.1999	Pas-De-Calais	
016650	12.2007	Doubs		J12428	11.1999	Ardennes	
S360	01.2006	Doubs	EU676237	W61	06.1998	Somme	
317630	06.2005	Calvados		2198	07.1998	Marne	
R13849	12.2005	Puy-De-Dôme	EU676229	FV116	07.1998	Somme	
N15097	01.2004	Jura	EU331451	V87	07.1997	Somme	
P794	01.2004	Haute-Saône	EU331455	F549	04.1996	Puy-De-Dôme	
P738	02.2004	Alpes-Maritimes	EU331445	F975	04.1996	Puy-De-Dôme	
P7136	06.2004	Doubs	EU331453	F1288	04.1996	Puy-De-Dôme	
Sa04	06.2004	Seine-Maritime	EU331456	F1651	04.1996	Puy-De-Dôme	
P7875	06.2004	Doubs	EU331454	E8643	04.1996	Puy-De-Dôme	
003454	02.2003	Somme	EU331444	V912	04.1996	Pas-De-Calais	
N11064	10.2003	Isère	EU331449	50	04.1996	Dordogne	
043725	10.2003	Somme	EU331452	8814	06.1996	Dordogne	
N15209	12.2003	Var	EU331450	V265	08.1996	Pas-De-Calais	
M13316	02.2002	Puy-De-Dôme	EU331447	V266	08.1996	Pas-De-Calais	
82701	03.2002	Seine-Maritime	EU331446	V313	08.1996	Pas-De-Calais	
AA92	04.2002	Seine-Maritime	EU331443	V543	08.1996	Pas-De-Calais	
1326	06.2002	Seine-Maritime		V463	08.1996	Pas-De-Calais	
K14547	01.2001	Drôme		V265	08.1996	Pas-De-Calais	
41222	01.2001	Somme		1	03.1995	Gard	
L5889	02.2001	Drôme		3	03.1995	Meuse	
L13389	02.2001	Puy-De-Dôme		12	03.1995	Hérault	
1091	05.2001	Aisne		13	03.1995	Pas-De-Calais	
1153	05.2001	Somme		16	03.1995	Meurthe-et-Moselle	
AA2	12.2001	Seine-Maritime	EU331442	19	04.1995	Meuse	
Y6	01.2000	Savoie		30	05.1995	Seine-Maritime	
K678	01.2000	Nièvre		40	05.1995	Bouches-du-Rhône	
J13667	02.2000	Meuse		V421	06.1995	Pas-De-Calais	
K2252	02.2000	Puy-De-Dôme		23	1994	Côte-d'Or	
K2165	02.2000	Savoie		27	1994	Pas-De-Calais	
J13992	02.2000	Ain		18	1990	Marne	
323	03.2000	Somme		186	1990	Haut-Rhin	
503	06.2000	Aisne		47	1988	Seine-Maritime	
K09117	07.2000	Savoie		71	1987	Pas-De-Calais	

552

553

554 **FIGURE LEGENDS**

555

556 **Figure 1.** Maximum clade credibility phylogeny of 118 full-length IHNV glycoprotein (G)
557 genes. The tree was scaled to time under the relaxed uncorrelated exponential molecular
558 clock. The thickness of the branches reflects the posterior probabilities (values below 50 are
559 thin). Isolate information (name, accession) and genogroup classification are shown on the
560 right. Isolate locations are indicated by colour (see colour key).

561

562 **Figure 2.** Temporal distribution of the 88 representative IHNV French isolates from 1987
563 (first outbreak) to 2015. The grey arrow indicates the year the virus was first detected.

564

565 **Figure 3.** Phylogenetic tree showing the relationship of the 88 French IHNV isolates. This
566 Bayesian analysis is based on the partial G sequence (570 bp). Sub-clades A to H are
567 identified using vertical lines. Sequences from 5 strains from M genogroup were included as
568 outgroup. The numbers are posterior probabilities reflecting clade support. Isolate information
569 (name, year of isolation) is shown. France was divided up in 3 geographical distributions
570 areas (North, East and Centre and South) coloured in orange, purple and blue respectively.
571 Spatial repartition of the 88 isolates is given by the red circles, whose sizes are correlated with
572 the number of strains (small: 1-5; medium: 6-9; large: 10-15). Percentages indicated on the
573 map correspond to the nucleotide sequence identities. Stars were used to localize strains with
574 atypical genetic profiles (in bold) and their potential origins (in normal character).

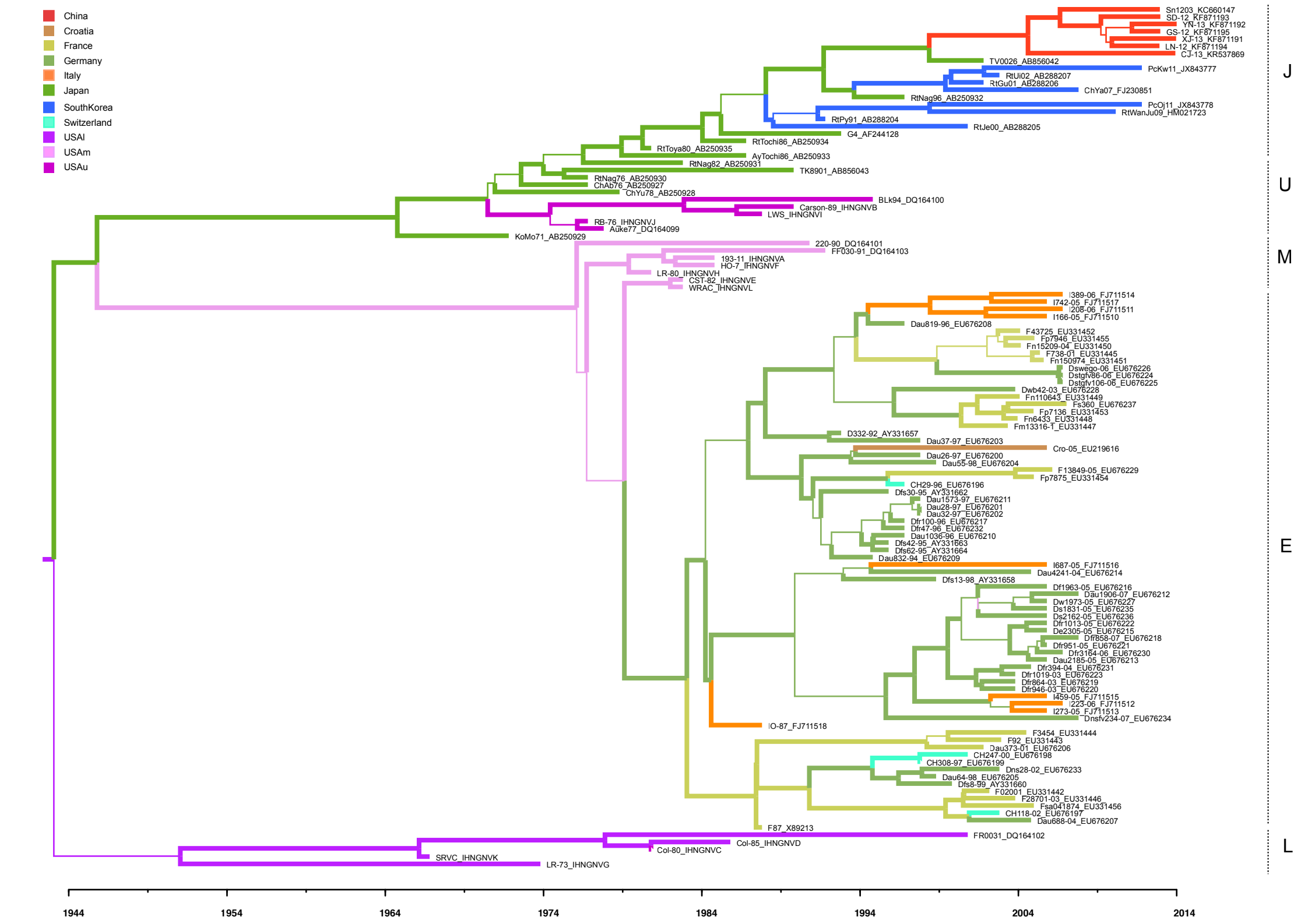
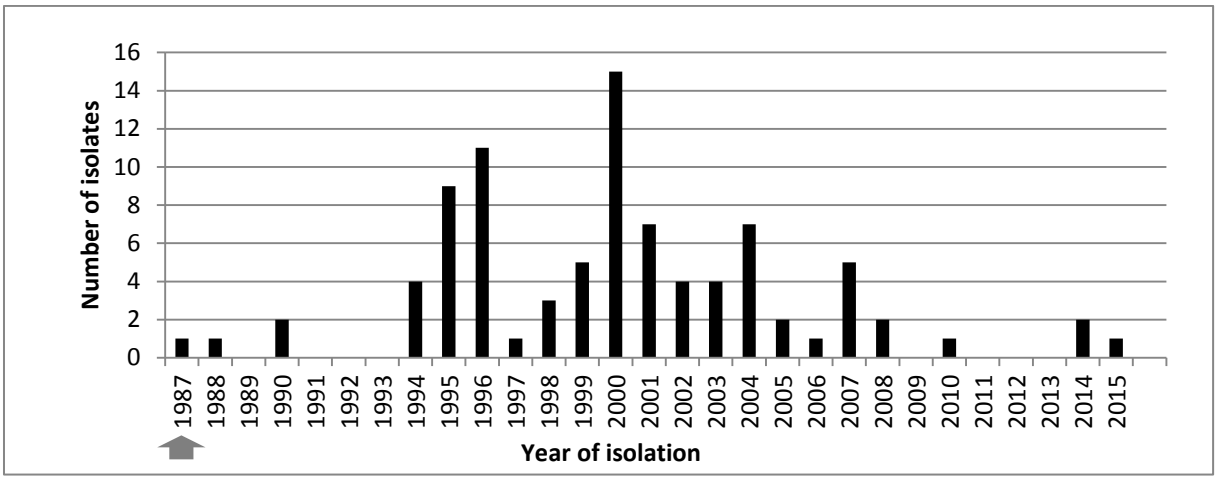
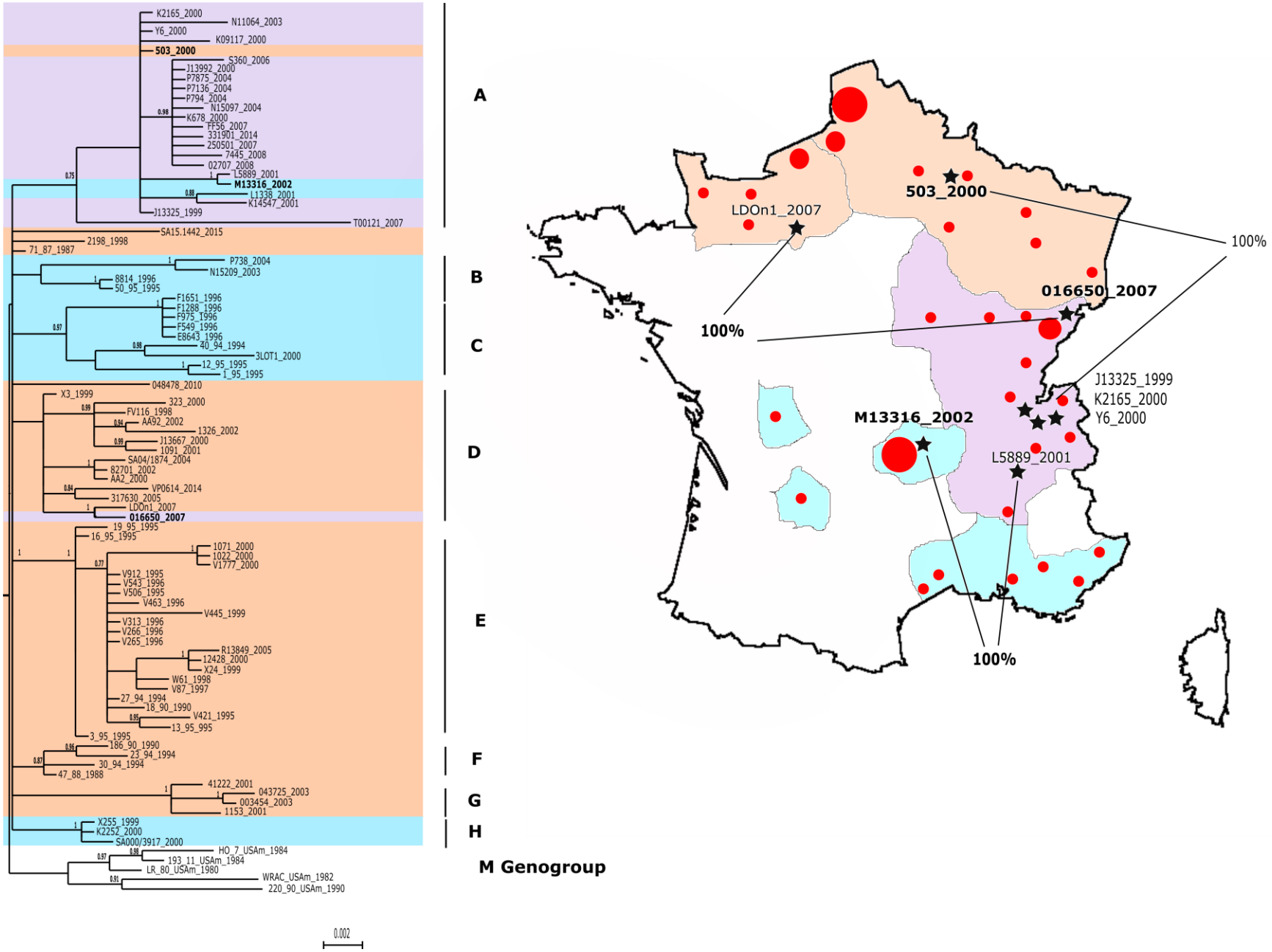

China
Croatia
France
Germany
Italy
Japan
SouthKorea
Switzerland
USAI
USAm
USAu
Sn1203_KC660147
SD-12_KF871193
YN-13_KF871192
GS-12_KF871195
XJ-13_KF871191
LN-12_KF871194
CJ-13_KR537869
TV0026_AB856042
PcKw11_JX843777
RtUi02_AB288207
RtGu01_AB288206
ChYa07_FJ230851
RtNag96_AB250932
PcOj11_JX843778
RtWanJu09_HM021723
RtPy91_AB288204
RtJe00_AB288205
G4_AF244128
RtTochi86_AB250934
RtToya80_AB250935
AyTochi86_AB250933
RtNag82_AB250931
TK8901_AB856043
RtNag76_AB250930
ChAb76_AB250927
ChYu78_AB250928
BLk94_DQ164100
Carson-89_IHNGNVB
LWS_IHNGNVI
RB-76_IHNGNVJ
Auke77_DQ164099
KoMo71_AB250929
220-90_DQ164101
FF030-91_DQ164103
193-11_IHNGNVA
HO-7_IHNGNVF
LR-80_IHNGNVH
CST-82_IHNGNVE
WRAC_IHNGNVL
I389-06_FJ711514
I742-05_FJ711517
I208-06_FJ711511
I166-05_FJ711510
Dau819-96_EU676208
F43725_EU331452
Fp7946_EU331455
Fn15209-04_EU331450
F738-01_EU331445
Fn150974_EU331451
Dswego-06_EU676226
Dstgfv86-06_EU676224
Dstgfv106-06_EU676225
Dwb42-03_EU676228
Fn110643_EU331449
Fs360_EU676237
Fp7136_EU331453
Fn6433_EU331448
Fm13316-1_EU331447
D332-92_AY331657
Dau37-97_EU676203
Cro-05_EU219616
Dau26-97_EU676200
Dau55-98_EU676204
F13849-05_EU676229
Fp7875_EU331454
CH29-96_EU676196
Dfs30-95_AY331662
Dau1573-97_EU676211
Dau28-97_EU676201
Dau32-97_EU676202
Dfr100-96_EU676217
Dfr47-96_EU676232
Dau1036-96_EU676210
Dfs42-95_AY331663
Dfs62-95_AY331664
Dau832-94_EU676209
I687-05_FJ711516
Dau4241-04_EU676214
Dfs13-98_AY331658
Df1963-05_EU676216
Dau1906-07_EU676212
Dw1973-05_EU676227
Ds1831-05_EU676235
Ds2162-05_EU676236
Dfr1013-05_EU676222
De2305-05_EU676215
Dfr858-07_EU676218
Dfr951-05_EU676221
Dfr3164-06_EU676230
Dau2185-05_EU676213
Df394-04_EU676231
Dfr1019-03_EU676223
Dfr864-03_EU676219
Dfr946-03_EU676220
I459-05_FJ711515
I223-06_FJ711512
I273-05_FJ711513
Dnsfv234-07_EU676234
IO-87_FJ711518
F3454_EU331444
F92_EU331443
Dau373-01_EU676206
CH247-00_EU676198
CH308-97_EU676199
Dns28-02_EU676233
Dau64-98_EU676205
Dfs8-99_AY331660
F02001_EU331442
F28701-03_EU331446
Fsa041874_EU331456
CH118-02_EU676197
Dau688-04_EU676207
F87_X89213
FR0031_DQ164102
Col-85_IHNGNVD
Col-80_IHNGNVC
SRVC_IHNGNVK
LR-73_IHNGNVG
J
U
M
E
L
1944
1954
1964
1974
1984
1994
2004
2014

Fig. 2


Number of isolates
16
14
12
10
8
6
4
2
0
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
Year of isolation


K2165_2000
N11064_2003
Y6_2000
K09117_2000
503_2000
S360_2006
J13992_2000
P7875_2004
P7136_2004
P794_2004
N15097_2004
K678_2000
FF56_2007
331901_2014
250501_2007
7445_2008
02707_2008
L5889_2001
M13316_2002
L1338_2001
K14547_2001
J13325_1999
T00121_2007
SA15.1442_2015
2198_1998
71_87_1987
P738_2004
N15209_2003
8814_1996
50_95_1995
F1651_1996
F1288_1996
F975_1996
F549_1996
E8643_1996
40_94_1994
3LOT1_2000
12_95_1995
I_95_1995
048478_2010
X3_1999
323_2000
FV116_1998
AA92_2002
1326_2002
J13667_2000
1091_2001
SA04/1874_2004
82701_2002
AA2_2000
VP0614_2014
317630_2005
LDOn1_2007
016650_2007
19_95_1995
16_95_1995
1071_2000
1022_2000
V1777_2000
V912_1995
V543_1996
V506_1995
V463_1996
V445_1999
V313_1996
V266_1996
V265_1996
R13849_2005
12428_2000
X24_1999
W61_1998
V87_1997
27_94_1994
18_90_1990
V421_1995
13_95_995
3_95_1995
186_90_1990
23_94_1994
30_94_1994
47_88_1988
41222_2001
043725_2003
003454_2003
1153_2001
X255_1999
K2252_2000
SA000/3917_2000
HO_7_USAm_1984
193_11_USAm_1984
LR_80_USAm_1980
WRAC_USAm_1982
220_90_USAm_1990
0.002
A
B
C
D
E
F
G
H
M Genogroup
LDOn1_2007
503_2000
100%
016650_2007
100%
J13325_1999
K2165_2000
Y6_2000
M13316_2002
L5889_2001
100%