FLAVIVIRIDAE — ADDITION TO THE FAMILY

I. V. Kruglov

Federal State Budgetary Scientific Institution
«I. I. Mechnikov Scientific Research Institute of Vaccines and Serums RAMS»
Moscow, Russian Federation

This review article analyzes and summarizes the history of the replenishment of the family Flaviviridae with new members over the last several decades on the example of the youngest genera of this family — Hepacivirus and Pegivirus. It all started in 1966 when surgeon George Barker, who had hepatitis, had blood serum taken in containing an unknown virus. This virus was named GBV, by patient initials. Samples of the serum were frozen. A nucleic acid recognized as corresponding to the genomes of 2 separate virus species was isolated from the tested material in 1995. These viruses were named GBV-A and GBV-B. By this time, the hepatitis C virus had already been discovered, which was assigned to the Flaviviridae family, where a separate, third genus of Hepacivirus was allocated for it.

In 2010, a more distantly related virus (named GBV-D) was found in bats (Indian flying fox — lat. Pteropus giganteus). GBV-B, which causes acute hepatitis in experimentally infected tamarisks, became the second species in the genus Hepacivirus to company with hepatitis C virus. The remaining GB viruses based on phylogenetic relationships, genome organization, and pathogenetic properties were proposed in 2011 to be classified as members of the fourth genus in the Flaviviridae family. This genus was named Pegivirus (pe — persistence, g — GB).

11 species of viruses have now been identified in the genus Pegivirus. They are indicated by letters in the order of the Latin alphabet — from Pegivirus A to Pegivirus K. And 14 species of viruses have now been identified in the genus Hepacivirus. So the story of the investigation, which began in 1966 with the discovery of the previously unknown GBV virus, has so far concluded with the discovery of two new genera of the family Flaviviridae. Numerous members of these two genera infect and also persist among a wide range of species belonging to different orders of the mammalian class, including Homo Sapiens.

Key words: Flaviviridae, Hepacivirus, Pegivirus, viral ecology.

В данной статье обзорного характера анализируется и обобщается история пополнения семейства Flaviviridae новыми членами за последние несколько десятилетий на примере наиболее молодых родов этого семейства — Hepacivirus и Pegivirus.

Все началось в 1966 году, когда у хирурга Джорджа Баркера, заболевшего гепатитом, была взята сыворотка крови, содержащая неизвестный вирус, который был назван GBV, по инициалам пациента. Образцы этой сыворотки крови были заморожены. В 1995 году из исследованного материала была выделена нуклеиноввая кислота, признанная соответствующей геномам 2-х отдельных видов вирусов, получивших наименование GBV-A и GBV-B. К этому времени уже был открыт вирус гепатита C, который был отнесен в семейство Flaviviridae, где для него был выделен отдельный, третий род Hepacivirus.

В 2010 году более отдаленным родственным вирус (названный GBV-D) был обнаружен у летучих мышей (индийская летучая лисица — lat. Pteropus giganteus). GBV-B, вызывающий острый гепатит у экспериментально инфицированных тармарионов, стал вторым видом в роде Hepacivirus, где составлял компанию вирусу гепатита C. Остальные GB вирусы на основании филогенетических взаимоотношений, организации генома и патогенетических свойств в 2011 году было предложено классифицировать как членов четвертого рода в семействе Flaviviridae, названного Pegivirus (pe — персистенция, g — GB).

В настоящее время идентифицировано 11 видов вирусов в составе рода Pegivirus. Они обозначены буквами по порядку латинского алфавита — от Pegivirus A до Pegivirus K. Идентифицировано также 14 гепацинусов, имеющих буквенные обозначения от A до N. Так история исследования, начавшаяся в 1966 году с выявления ранее неизвестного вируса GBV, завершилась к настоящему времени открытием двух новых родов семейства Flaviviridae, многочисленные представители которых инфицируют, а также персистируют среди широкого спектра видов, принадлежащих к различным отрядам класса млекопитающих, включая Homo Sapiens.

Ключевые слова: Flaviviridae, Hepacivirus, Pegivirus, экология вирусов.

I. V. Kruglov
Flaviviridae — Addition to the Family

The Flaviviridae family was separated as a separate family from the Togaviridae family in 1985, and included at that time two genera of viruses — Flavivirus and Pestivirus. After the discovery of hepatitis C virus in 1989, which has an external envelope, a linear single-stranded positive
RNA genome, based on these features as well as structural features of RNA, it was also included in the Flaviviridae family, where a separate, third genus Hepacivirus was established for it. In which this virus was initially in proud lone, unlike the hepatitis B virus, for example, which by that time had become a prototype virus for the whole new family of Hepadnaviridae.

But the hepatitis C virus wasn’t alone for long. Yet in 1966, the surgeon George Barker, who had suffered hepatitis without HBsAgemia, came to the attention of virologists. The patient's blood, taken on the 3rd day of the icteric period, was injected to tamarins, who first developed acute hepatitis, and then — postinfection immunity, protecting them from reinfection. After that, the supposed new hepatitis virus was named GBV, according to the patient's initials, and the patient's serum samples were frozen for further investigation by the future virology methods [15].

This future came in 1995, when the PCR-based method, known as representative differential analysis, was used to investigate this material. As a result, a nucleic acid recognized as corresponding to the genomes of 2 separate viral species, was isolated from the tested material. These viruses were named GBV-A and GBV-B [39]. When tamarins were infected, acute hepatitis and the subsequent development of immunity caused only GBV-B virus. This allowed this virus to be initially seen as an etiological agent of hepatitis in this monkey species. However, the GBV-B virus was not detected in wildlife and its natural host is unknown [41]. Although there is a suggestion that these are the New World’s wild monkeys [40]. Tamarins were resistant to GBV-A, and this virus was initially recognized as anthropoconic. However, a number of variants of GBV-A were subsequently identified in the New World’s wild monkeys. The chimpanzees are also the natural host for GBV-A.

Also in 1995, another virus was isolated from the blood of patients with hepatitis no A — no E. The nucleotide sequence of the viral genome was quite close to the above-mentioned viruses. Therefore, this new virus has been designated as GBV-C [38]. GBV-C was especially close to GBV-A, which with it, as was initially thought, had 59% identical nucleotides and 64% identical amino acids, while with GBV-B it had only 30% identical amino acids. As a result, GBV-A and GBV-C were recognized as different isolates of the GBV-A/C virus.

Among the many hepatitis C virus isolates, a virus was also identified, which was subsequently found to be frequent in intravenous drug abusers and people with repeated blood transfusions and in the absence of hepatitis C. Since 1995, this virus has been named the hepatitis G virus — HGV. However, it turned out that the nucleotide and amino acid sequences of this virus were very close to GBV-C — 95% of the nucleotide and 85% of the amino acid sequences coincided. As a result, HGV and GBV-C were recognized as different isolates of a single virus, which at that stage received the unifying name of HGV.

In 2010, a more distantly related virus (named GBV-D) was found in bats (Indian flying fox — lat. Pteropus giganteus) [17]. In the result of the analysis of the nucleotide composition and genome structure, all these viruses were also classified to the same family of Flaviviridae, in which the hepatitis C virus was at that time. Initially, only GBV-B causing acute hepatitis in experimentally infected tamarins, was classified. It became the second species in the genus Hepacivirus to make up the hepatitis C virus company. The other GB viruses were not assigned to any genus in the Flaviviridae family.

Therefore, Jack Stapleton, Peter Simmonds et al. proposed in 2011 to classify GBV-A-like viruses, GBV-C and GBV-D on the basis of phylogenetic relationships, genome organization and pathogenetic properties as members of the fourth genus in the Flaviviridae family, named Pegivirus (pe — persistence, g — GB). And it was done [41].

Two species of viruses belonging to the genus Pegivirus were initially identified. These were pegivirus A, combining viruses formerly known as GBV-A and GBV-C/HGV, and pegivirus B, formerly known as GBV-D. It has been found that humans and chimpanzees are the natural hosts for the GBV-C/HGV subspecies of Pegivirus A, and chimpanzees and some New World’s monkeys for the GBV-A subspecies of Pegivirus A. The first natural hosts identified for Pegivirus B were bats, as mentioned above.

It should be noted that, although the terminology based on the host species is not approved by the International Committee on Virus Taxonomy, pegiviruses A have received additional names in the classification of viruses: HPgV — for isolates obtained from humans; SPgV — for isolates obtained from New World’s monkeys; SPgVcpz — for isolates obtained from chimpanzees.

Further research has identified several other species as candidates for the Pegivirus genus. It was a rodent pegivirus identified in white-throated woodrats (lat. Neotoma albigula) in 2013 [23]. Horse pegivirus was detected in these animals, respectively [22]. It was found that Theiler’s disease (horse hepatitis with parenteral transmission) is caused by pegivirus too. This virus was initially named Theiler’s disease virus [11]. Both horse viruses were identified in 2013. As well as the new human hepegivirus HHPgV or HPgV-2, detected in people with history of blood transfusions, in 2015 [21]. A closely related virus was identified in the graceful catshark (lat. Proscyllium habereri). However, the question of whether this virus be-
longs to pegiviruses or hepaviruses has not yet been resolved [35].

At present, 11 virus species have been identified in the genus *Pegivirus* on the basis of deep sequencing technologies. These viruses differ more than 50% in the nucleotide sequence and more than 55% in the amino acid sequence [40]. They are indicated by letters in the order of the Latin alphabet — from *Pegivirus A* to *Pegivirus K*. For pegiviruses persisting with high frequency in the human population and initially described as GB virus C (*GBV-C*) and Hepatitis G virus (*HGV*) [26, 38], the name of the human pegivirus HPGV was subsequently proposed. Since at present there is no evidence that this virus is associated with hepatitis [41], and the surgeon GB was not infected with it. HPGV is transmitted sexually, through contact with infected blood and from mother to child. This virus has now been identified as *Pegivirus C* [40]. Another subspecies of *Pegivirus C* (*GBV-Ctro*) infects the chimpanzees [3, 9]. Initial data on the high degree of nucleotide and amino acid sequence homology of *GBV-A* and *GBV-C/HGV* viruses have not been confirmed. Thus, these viruses, initially combined into one species of *Pegivirus A*, were subsequently still divided into two species — *Pegivirus A* and *Pegivirus C*, respectively. *Pegivirus C* is not the lone human pegivirus. There is also the parenterally transmitted *HHPGV* or *HPgV-2* pegivirus [8, 21], now identified as *Pegivirus H* [40]. *Pegiviruses A, B, D, E, F, G, I, K* infect Old World’s monkeys [7, 36]. At the same time, as already mentioned, *Pegivirus A* (*GBV-A*) also infects several species of New World’s monkeys [10, 30], and *Pegivirus B* infects bats. *Pegivirus E* (equine pegivirus) [22] and *Pegivirus D* (Theiler’s disease virus) [11] also infect horses. *Pegivirus K* also infects pigs [6]. *Pegiviruses F, G, I, J* also infect a wide range of rodent and bat species [23]. Pegiviruses, as well as hepaviruses, differ from other flaviviruses — members of the genera *Flavivirus* and *Pestivirus*, limited ability to replicate in cell cultures. Although this barrier has been overcome for the hepatitis C virus [1, 2] as well as for the *Pegivirus C* [13].

*Pegivirus C*, formerly known as *GBV-C*, persists with high frequency in human society. But, except for the association with non-Hodgkin’s lymphoma, *Pegivirus C* is not associated with the development of any identifiable disease [12, 24]. According to current information, the infection of other mammalian species with pegiviruses is persistent and nonpathogenic, except for infection of horses by *Pegivirus D* [11]. In general, this is quite typical for viruses of this genus, the name of which came from the word «persistence».

Pegiviruses have both similarities and some differences in the genome organization from viruses of the genus *Hepacivirus* and other genera of the Flaviviridae family (Fig. 1, 2, 3). All pegiviruses have a long 5'-untranslated region with predicted internal ribosome entry site (IRES) function. The most part of pegiviruses have a picornavirus-like IRES type I, that has no structural analogues in other flaviviruses, while some others have IRES type IV structurally similar to hepaviruses and pestiviruses. Besides that, pegiviruses do not encode a protein similar to the nucleocapsid protein of other flaviviruses [28, 31, 41]. IRES is a structural part of the RNA, that allows to initiate cap-independent translation as part of the overall process of protein synthesis. IRES sequences were first discovered in 1988 in polyomavirus (PV) [29] and encephalomyocarditis virus (EMCV) [19].

As with other flaviviruses, the pegivirus genome contains one open reading frame (ORF). The genome encodes a polyprotein that is co- and posttranslationally cuts into individual viral proteins (Fig. 1). The structural proteins common to all pegiviruses are the E1 and E2 envelope glycoproteins. The non-structural proteins NS2 — NS5B are also common to all pegiviruses. Cutting of structural proteins by cellular signal peptidases, NS2/NS3 by NS2-NS3 autoprotease and other NS proteins by NS3-NS4A protease complex is carried out in the same order as in hepaviruses.

The size of the pegivirus genome ranges from 8900 to 11300 nucleotide bases.

Pegiviruses with a longer genome are thought to encode the additionally predicted structural proteins X and Y. Thus, although pegiviruses lack an analogue of the nucleocapsid protein characteristic for other flaviviruses, some of them probably have the predicted Y protein with an unknown function preceding E1. Several pegiviruses also probably have the predicted additional glycoprotein X, following E2 (Fig. 1).

And what is the situation with the hepaviruses? Initially, we focused on the fact that the second species in the genus *Hepacivirus*, which accompanied the hepatitis C virus, was *GBV-B*, causing acute hepatitis in experimentally infected tamarins (see above). Then hepaviruses were detected in bats [16], rodents [16, 23], horses [16] and dogs, and African primates [25]. Based on the results of such methods as immunoblot of the sera of infected animals, PCR and sequencing of viral genomes, J.F.Drexler et al. obtained the following initial picture of the phylogenetic relationships of potential members of the genus *Hepacivirus* with each other and with other viruses of the Flaviviridae family (Fig. 4) [16].

14 species of viruses have now been identified in the genus *Hepacivirus* [40]. *Hepacivirus A* is the dog’s and horse hepavirus. This virus is evolutionarily closest to the hepatitis C virus [43].
Hepacivirus B is a virus formerly known as GBV-B. Hepacivirus C is the hepatitis C virus. Hepacivirus D is the guereza hepacivirus (the eastern black-and-white colobus or the guereza (lat. Colobus guereza) — a species of monkey inhabiting central Africa and Ethiopia). Hepacivirus E is the virus formerly known as rodent hepacivirus 339. Hepacivirus F is the virus formerly known as rodent hepacivirus NLR07-oct70. Hepacivirus G corresponds to hepacivirus 1 of brown rats (lat. Rattus norvegicus) and Hepacivirus H corresponds to hepacivirus 2 of brown rats. Hepacivirus I corresponds to rodent hepacivirus SAR-3/RSA/2008 and Hepacivirus J corresponds to rodent hepacivirus RMU10-3382/GER/2010. Hepaciviruses K, L, M are hepaciviruses of bats formerly known as PDB-829, PDB-112 and PDB-491.1, respectively. Hepacivirus N is the cattle hepacivirus [5, 14]. And it is questionable whether hepacivirus is a new virus isolated from hoary bamboo rats (lat. Rhizomys pruinosus) [44]. As already mentioned above, the question of whether a virus identified in the graceful catshark (lat. Proscyllium habereri) belongs to Hepaciviruses or Pegiviruses has not yet been resolved [35].

Hepaciviruses have both similarities and some differences in the genome organization from viruses of other genera of the Flaviviridae family (Fig. 3). The most part of hepaciviruses contain IRES type IV in the 5'-untranslated region, however the IRES of such rodent hepaciviruses as F and J, has some homology with the corresponding sequence of pegiviruses [16, 18, 28, 40]. The hepacivirus genome sizes range from 8900 to 10500 nucleotide bases. The hepacivirus genome contains one ORF encoding a polyprotein of about 3000 amino acids (Fig. 2). The HCV virion contains at least three proteins: the nucleocapsid C core protein and two envelope glycoproteins, E1 and E2. An additional protein, p7 (which is supposed to have the properties of an ion channel protein important for the virus assembly), is not completely cleaved from the E2 precursor, forming E2-p7 and p7 [34]. However, it is unknown whether they are structural elements of the virion. The GBV-B corresponding protein, p13, is cut into p7 and p6 [42]. The genomes of other hepaciviruses have the similar organization with genomes of HCV and GBV-B, with the predicted splitting sites in the coding region, potentially producing the proteins C, E1, E2, p7/p13, NS2, NS3, NS4A, NS4B, NS5A and NS5B, comparable in the size to those of HCV and GBV-B. The only exception is the additional amino acid sequence of the NS5A guereza hepacivirus protein (Hepacivirus D) [25].

The cutting of viral proteins from the polyprotein in hepaciviruses occurs generally in a similar order to pegiviruses. Structural proteins are also cut by cellular signal peptidases. The peculiarity of hepaciviruses is that they have a signal peptide peptidase involved in this process. This enzyme cuts a part of the protein already sliced from the polyprotein, unlike signal peptidases in general. Concretely, it cuts off the C-end of the nucleocapsid protein in hepaciviruses. Since pegiviruses have not the analogous nucleocapsid protein, the analogous peptide peptidase is not involved in the processing of their proteins. Cutting of non-structural proteins in hepaciviruses is the same as in pegiviruses — the NS2/NS3 sequence is cut by NS2-NS3 autoprotease and the other NS proteins are cut by the NS3-NS4A protease complex.

Hepatitis C, the infection caused by Hepacivirus C, is well known, and here is no need to describe it. The natural host of Hepacivirus B, formerly named GBV-B, is unknown [41], although there is a suggestion that these are wild New World’s monkeys [40]. Horse hepacivirus infection is accompanied by hepatitis with a slight increase of enzyme levels [27, 30]. Quantitative PCR and histopathology of hepacivirus-infected rodents confirmed tropism to liver tissue with hepatitis C-like inflammation [16]. The pathogenicity of other hepaciviruses has not yet been sufficiently studied. However, the presence of miR-122 microRNA binding sites in the majority of hepaciviruses suggests widespread hepatotropism among species of this genus [40]. MicroRNA 122 (miR-122) is a small RNA molecule that is specifically expressed in the liver. According to the information obtained, miR-122 promotes both hepacivirus RNA replication [20, 32] and its translation [32, 33] as a result of interaction between miR-122 and the 5'-non-coding region of the viral genome [4]. This is quite typical for viruses of this genus, the name of which came from the word «hepatotropism».

However, some of the characteristics that were initially used in the classification in the long run are no strictly different for all the viruses that form these two genera. It concerns to the IRES type, the presence of miR-122 binding sites in the 5'-non-coding region of the viral genome and the virus persistence. At present, the demarcation between these genera is based on the phylogenetic difference and, above all, the presence or absence of the nucleocapsid protein.

So the story of the investigation, which began in 1966 with the discovery of the previously unknown GBV virus, has so far concluded with the discovery of two new genera of the family Flaviviridae. Numerous members of these two genera infect and also persist among a wide range of species belonging to different orders of the mammalian class, including Homo Sapiens.
Fig. 1 — Pegivirus genome organization and viral protein processing [37]

Fig. 2 — Hepacivirus genome organization and viral protein processing [37]

Fig. 3 — Genome organization of typical representatives of the Flaviviridae family by the example of viruses of the genus Flavivirus and viral protein processing [37]
REFERENCES


Fig. 4 — Phylogenetic relationships of potential members of the genus Hepacivirus with each other and with other viruses of the Flaviviridae family according to the results of the investigation J. F. Drexler et al. [16]


