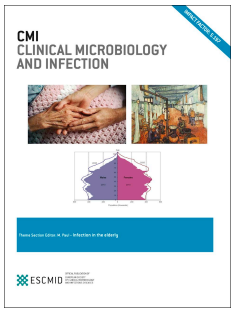
Accepted Manuscript



A Recent epidemiological-cluster of acute hepatitis B genotype F1b infection in a restricted Geographic area of Italy

Michela Pollicita, Claudia Alteri, Maria Concetta Bellocchi, Daniele Armenia, Luca Carioti, Romina Salpini, Luna Colagrossi, Arianna Battisti, Marianna Aragri, Lavinia Fabeni, Rinalda Mariani, Margherita Dalessandro, Armando Ranelli, Maurizio Paoloni, Giustino Parruti, Carlo Federico Perno, Valentina Svicher

PII: S1198-743X(15)00783-1

DOI: 10.1016/j.cmi.2015.07.022

Reference: CMI 350

To appear in: *Clinical Microbiology and Infection*

Received Date: 23 February 2015

Revised Date: 19 June 2015

Accepted Date: 24 July 2015

Please cite this article as: Pollicita M, Alteri C, Bellocchi MC, Armenia D, Carioti L, Salpini R, Colagrossi L, Battisti A, Aragri M, Fabeni L, Mariani R, Dalessandro M, Ranelli A, Paoloni M, Parruti G, Perno CF, Svicher V, A Recent epidemiological-cluster of acute hepatitis B genotype F1b infection in a restricted Geographic area of Italy, *Clinical Microbiology and Infection* (2015), doi: 10.1016/j.cmi.2015.07.022.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

A Recent Epidemiological-cluster of Acute Hepatitis B Genotype F1b Infection in a Restricted Geographic Area of Italy

Running title: Cluster of Acute HBV F1b Infection

Michela Pollicita[1], Claudia Alteri[1], Maria Concetta Bellocchi[1], Daniele Armenia[1], Luca Carioti[1], Romina Salpini[1], Luna Colagrossi[1], Arianna Battisti[1], Marianna Aragri[1], Lavinia Fabeni[2], Rinalda Mariani[3], Margherita Dalessandro[3], Armando Ranelli[3], Maurizio Paoloni[3], Giustino Parruti[4], Carlo Federico Perno[1], Valentina Svicher[1]

[1]*Department of Experimental Medicine and Surgery, University of Rome Tor Vergata, Rome, Italy*

[2]*INMI "L. Spallanzani", Rome, Italy*

[3]*''S.S. Filippo e Nicola'' Hospital, Infectious Disease Unit, Avezzano, Italy*

[4]*''Spirito Santo'' Hospital, Infectious Disease Unit, Pescara, Italy*

Valentina Svicher: valentina.svicher@uniroma2.it, Department of Experimental Medicine and Surgery, University of Rome Tor Vergata, Rome, Italy, Via Montpellier 1 CAP 00133, Rome, Italy, tel 0039 0672596564 fax 00390672596039

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Abstract

In this study, by phylogenetic-analysis, we identified an epidemiological-cluster involving 8 individuals diagnosed for acute HBV-infection related to unprotected sexual-intercourses in a restricted area of Central Italy (time-period: 2011-2014). Notably, these patients (6/8 Italians) were infected by sub-genotype F1b, not commonly found in Western-Countries. Ultra-deep pyrosequencing confirms a superimposable composition of HBV-quasispecies in these patients. Despite the availability of effective vaccination, this study highlights the importance not to underestimate the risk of HBV-infection, to continue to set-up surveillance programs for HBV-infection, and to investigate the pathogenetic potential of these atypical genotypes.

Keywords: HBV, Genotype, Vaccine, Acute infection, Epidemiological-cluster.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Background

Ten HBV-genotypes (A-H plus tentative genotypes I and J) and sub-genotypes were so far identified based on intergroup divergence of 8% or 4% in complete HBV-genome, respectively [1]. Among them, genotype F was divided into four sub-genotypes (HBV/F1-F4). HBV/F1 is found in Central America (HBV/F1a), Alaska and South America (HBV/F1b). HBV/F3 is found in Central America and in northern South America, whereas HBV/F2a, HBV/F2b and HBV/F4 are found in South America [2].

In this study, we described a local epidemiological-cluster of acute HBV-infection, sexually transmitted in a short time period, involving 8 individuals infected with sub-genotype F1b.

ACCEPTED MANUSCRIPT

Methods

Protocols used for population and ultra-deep HBV-sequencing, and for phylogenetic analysis were extensively reported in Supplementary Materials (SM_I-III).

Results

Between November 2011 and May 2014, 8 individuals (7 females and 1 male) were diagnosed for acute HBV-infection in two Hospitals of Abruzzo region in Central Italy (defined as A-H in Fig_1, Table_1). The definition of acute HBV-infection was based on: HBsAg and IgM anti-HBc positivity, and ALT levels ≥10 times the normal value. All patients had a compatible clinical history for acute HBV-infection.

All individuals referred unprotected sexual intercourses within weeks before the appearance of symptoms. Specific risk factors for blood-related transmission were excluded (no tattoos, no transfusion of blood or blood derivatives, no drug use, no unsafe healthcare-related injection practices). No patients worked in a health care setting or live with HBsAg-positive individuals.

Six individuals (not-vaccinated for HBV) were Italian (age from 37 to 59 years), and two were from Philippine and Argentina (age of 26 and 44 years, respectively) (Table_1). The 7 women (patients: B-H) spontaneously resolved HBV-infection (HBsAg-negative, anti-HBc-positive +/-anti-HBs). The only man (patient_A) showing a HBsAg-persistence for 6 months, resolved the infection as consequence of anti-HBV treatment.

For diagnostic purposes, a genotypic test based on the population-based sequencing of HBV Reverse-Transcriptase (RT)/HBsAg (SM_I), was performed at the University Hospital "Tor Vergata", Rome, Italy. This test revealed no drug-resistance RT-mutations, and some amino acids (wild-type in F1b genotype) at specific HBsAg-positions (L110_T114_K122_L127_F134_S140) known to affect HBsAg-recognition by antibodies (http://hbv.bioinf.mpi-inf.mpg.de/index.php).

Phylogenetic-analysis by NJ method (SM_II) showed that the RT-sequences from the 8 individuals belonged to HBV F1b subgenotype, and formed a well-defined cluster (bootstrap-value >90%) (Fig_S1). Once HBV-genotype was assigned, the statistical robustness of cluster was confirmed by

ML tree (SM_II), containing only HBV/F1b RT-sequences and HBV/F1a RT-sequences used as outgroup (Fig_1A). Again, RT-sequences from the 8 individuals formed a cluster distinct from the other 70 F1b RT-sequences included in the analysis (bootstrap >85%) (Fig_1A).

This cluster was confirmed also by analyzing concatenated HBV/F1b BCP/Pre-Core/Core/RT-sequences (bootstrap >85%) (Fig_1B).

Finally, the mean genetic distance in both RT and BCP/pre-Core/Core-sequences involved in the cluster was extremely low (RT:0.00006, standard error [SE]:±0.00001; BCP/pre-Core/Core:0.00001 [SE:±0.0000001]), confirming a high homology among sequences.

Furthermore, the composition of HBV-quasispecies in the 8 individuals involved in the cluster was investigated by UDPS, performed on RT/HBsAg and BCP/Pre-Core/Core regions (SM_III).

The patterns of mutations detected in these genomic regions, with an intra-patient prevalence ranging from 10 to 100%, were superimposable in all the 8 individuals (Table_S1).

Only in patient H, additional mutations were detected in BCP, all with prevalence of 100% (Table_S1).

Conversely, the patterns of mutations detected in RT/HBsAg and BCP/Precore/Core, with an intra-patient prevalence <10%, varied among the patients (Table_S1), suggesting a genetic diversification only at the level of minority species.

Discussion

This study provides molecular evidence of a local epidemiological-cluster involving 8 individuals, diagnosed for acute HBV-infection related to unprotected sexual intercourses in a restricted area of Central Italy (all resident in two cities no more than 115 Km apart) in a short period of time.

Previous European studies reported epidemiological-clusters mainly in the setting of nosocomial infections [3, 4]. To our knowledge, this is the first study showing an epidemiological-cluster not sustained by an iatrogenic route of infection.

We emphasize that this is an epidemiological-cluster and not a transmission-chain. Thus, we cannot exclude the involvement of other individuals (not-identified) contributing to HBV/F1b-transmission.

In the cluster identified, the 8 individuals were not-vaccinated: 6 Italians and 2 from Countries where the vaccination coverage is middle/low. In Italy, the nationwide HBV-vaccination program has targeted all infants born since 1992, and adolescents born between 1980-1991. Thus, there is a reservoir of Italian population (sexually-active) born before 1980, susceptible to HBV-infection. This suggests the importance to encourage a larger portion of adult individuals to HBV-vaccination and to offer HBV-vaccination to foreign individuals negative for HBV-infection.

This epidemiological-cluster was sustained by sub-genotype F1b, not commonly found in Europe. This raises the issue on the circulation of non-D and non-A genotypes progressively increasing in Western-Countries. Epidemiological models have shown that, within next decades, HBV-epidemiology will be completely modified by waves of immigration, affecting the local prevalence of HBV-infection, routes of transmission, and impacting prophylactic, diagnostic and therapeutic measures [5, 6]. The imported HBV-infections can be also transmitted to native population (as described in this cluster), further contributing to the spread of atypical genotypes.

Furthermore, HBV-genotypes are known to profoundly affect the natural history of HBV-infection. In particular, it is known the faster progression of genotype C and E to liver-cancer [7]. This was recently described also for genotype F [2]. Thus, the increased prevalence of these genotypes in Western-Countries may have implications for the burden and clinical management of liver-related morbidity and mortality.

In the epidemiological-cluster identified, all individuals resolved acute HBV-infection. It is known that the resolution of acute HBV-infection does not imply HBV eradication, but its persistence in an occult form in the vast majority of patients [8-10]. Occult HBV-infection is recognized as a co-factor promoting liver-cancer [11, 12], and can reactivate spontaneously or under immune-suppression conditions. These observations suggest to set-up interventions aimed at promoting

screening and (when necessary) treatment at the migrant groups coming from Countries with high HBV-prevalence.

In conclusions, despite the availability of effective vaccination, this study highlights the need not to underestimate the risk of HBV-infection, to continue to set-up surveillance programs of HBV-infection, and to investigate the pathogenetic potential of these atypical genotypes.

Acknowledgments

This work was supported by FIRB project [RBAP11YS7K_001]; by the Italian Ministry of Instruction, University and Research [Progetto_Bandiera PB05]; and Aviralia Foundation. We thank Massimiliano Bruni and Andrea Biddittu for data management.

Transparency declaration

The authors declare no conflicts of interest

ACCEPTED MANUSCRIPT

References

1) Pourkarim MR, Amini-Bavil-Olyaee S, Kurbanov F, Van Ranst M, Tacke F. Molecular identification of hepatitis B virus genotypes/subgenotypes: revised classification hurdles and updated resolutions. World J Gastroenterol 2014; 20: 7152-7168.
2) Svicher V, Gori C, Trignetti M, et al. The profile of mutational clusters associated with lamivudine resistance can be constrained by HBV-genotypes. J Hepatol 2009; 50: 461-470.
3) Götz HM, Schutten M, Borsboom GJ, Hendriks B, van Doornum G, de Zwart O. A cluster of hepatitis B infections associated with incorrect use of a capillary blood sampling device in a nursing home in the Netherlands, 2007. Euro Surveill 2008; 13. pii: 18918.
4) Pourkarim MR, Amini-Bavil-Olyaee S, Kurbanov F, Van Ranst M, Tacke F. Molecular identification of hepatitis B virus genotypes/subgenotypes: revised classification hurdles and updated resolutions. World J Gastroenterol 2014; 20: 7152-7168.
5) Mitchell T, Armstrong GL, Hu DJ, Wasley A, Painter JA. The increasing burden of imported chronic hepatitis B-United States, 1974-2008. PLoS One 2011; 6: e27717
6) Lin CL, Kao JH. The clinical implications of hepatitis B virus genotype: Recent advances. J Gastroenterol Hepatol 2011; 26 Suppl 1:123-130.
7) Marciano S, Galdame OA, Gadano AC. HBV-genotype F: natural history and treatment. Antivir Ther 2013; 18:485-488.
8) Rehermann B, Ferrari C, Pasquinelli C, Chisari FV. The hepatitis B virus persists for decades after patients' recovery from acute viral hepatitis despite active maintenance of a cytotoxic T-lymphocyte response. Nat Med 1996; 2:1104-1108.
9) Raimondo G, Navarra G, Mondello S, et al. Occult hepatitis B virus in liver tissue of individuals without hepatic disease. J Hepatol 2008; 48: 743-746.
10) Pollicino T, Raimondo G. Occult hepatitis B infection. J Hepatol 2014**;** 61: 688-689.
11) Raimondo G, Caccamo G, Filomia R, Pollicino T. Occult HBV-infection. Semin Immunopathol 2013; 35: 39-52.

12) Nishikawa H, Osaki Y. Clinical significance of occult hepatitis B infection in progression of liver disease and carcinogenesis. J Cancer 2013; 4: 473-480.

ACCEPTED MANUSCRIPT

Legend to Figures

Figure_1. A) Maximum likelihood phylogenetic tree was constructed by using HBV F1b RT-sequences from: 70 isolates downloaded from GenBank, 8 individuals involved in the cluster, and 2 individuals collected from routine laboratory testing at the University Hospital "Tor Vergata" (identified with arabic numbers). Two HBV F1a RT-sequences were used as outgroup. **B)** Maximum likelihood phylogenetic tree was constructed by using concatenated HBV F1b BCP/preCore/Core/RT-sequences from: 58 isolates downloaded from GenBank, and 5 individuals involved in the cluster. Two HBV F1a BCP/preCore/Core/RT-sequences were used as outgroup. Branch lengths were estimated with the best fitting nucleotide substitution model (GTR+G+I) according to a hierarchical likelihood ratio test. The bar at the bottom indicating 0.001 nucleotide substitution per site. The asterisks (*) along a branch represent bootstrap support >85%. The clusters involving the 8 HBV F1b RT-sequences and the 5 concatenated RTS/BCP/preCore/Core-sequences are in the grey box.

Figure_S1: Neighbour joining phylogenetic tree constructed on the polymerase gene sequences of 659 isolates and additional 155 HBV F genotypes references. The bar at the bottom of the tree indicates 0.01 nucleotide substitution per site. Isolates of genotype F1b are shown in red, and isolates of genotype F1a are in blue. The cluster involving the 8 HBV F1b sequences is in the grey box. Bootstrap values ≥80% were showed.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table_1. Characteristics of patients.

	PATIENTS							
	A	**B**	**C**	**D**	**E**	**G**	**F**	**H**
SEX	M	F	F	F	F	F	F	F
NATIONALITY	ITALIAN	ITALIAN	PHILIPPINE	ITALIAN	ITALIAN	ITALIAN	ITALIAN	ARGENTINIAN
AGE (YEARS)	59	37	26	40	43	58	39	44
DATE OF BLOOD SAMPLE COLLECTION[a]	March 2012	February 2014	February 2014	November 2013	November 2013	September 2013	September 2013	May 2014
SERUM HBV DNA (IU/ml)[b]	>170,000,000	462	6,490	5,233	54	346,176	1578	44,346
AST[b]	55	341	1,510	1,776	565	1,379	61	1,421
ALT[b]	157	1,305	2,227	2,663	2,324	2,237	122	1,865
HBSAg (IU/ml)[b]	115,261	7,765	10,292	153	1,360	23	1,481	positive [c]
ANTI HBS[b]	-	-	-	-	-	-	-	-
HBeAg[b]	+	+	+	+	-	+	-	n.a.
ANTI Hbe[b]	-	+	-	+	+	-	+	n.a.
HBcAg IgG/IgM[b]	+	+	+	+	+	+	+	+

[a] This date is referred to the blood sample withdrawal for genotypic testing

[b] Values are those observed at the time of genotypic testing

[c] For this patient only qualitative HBsAg was available

n.a. Not available data

ACCEPTED MANUSCRIPT

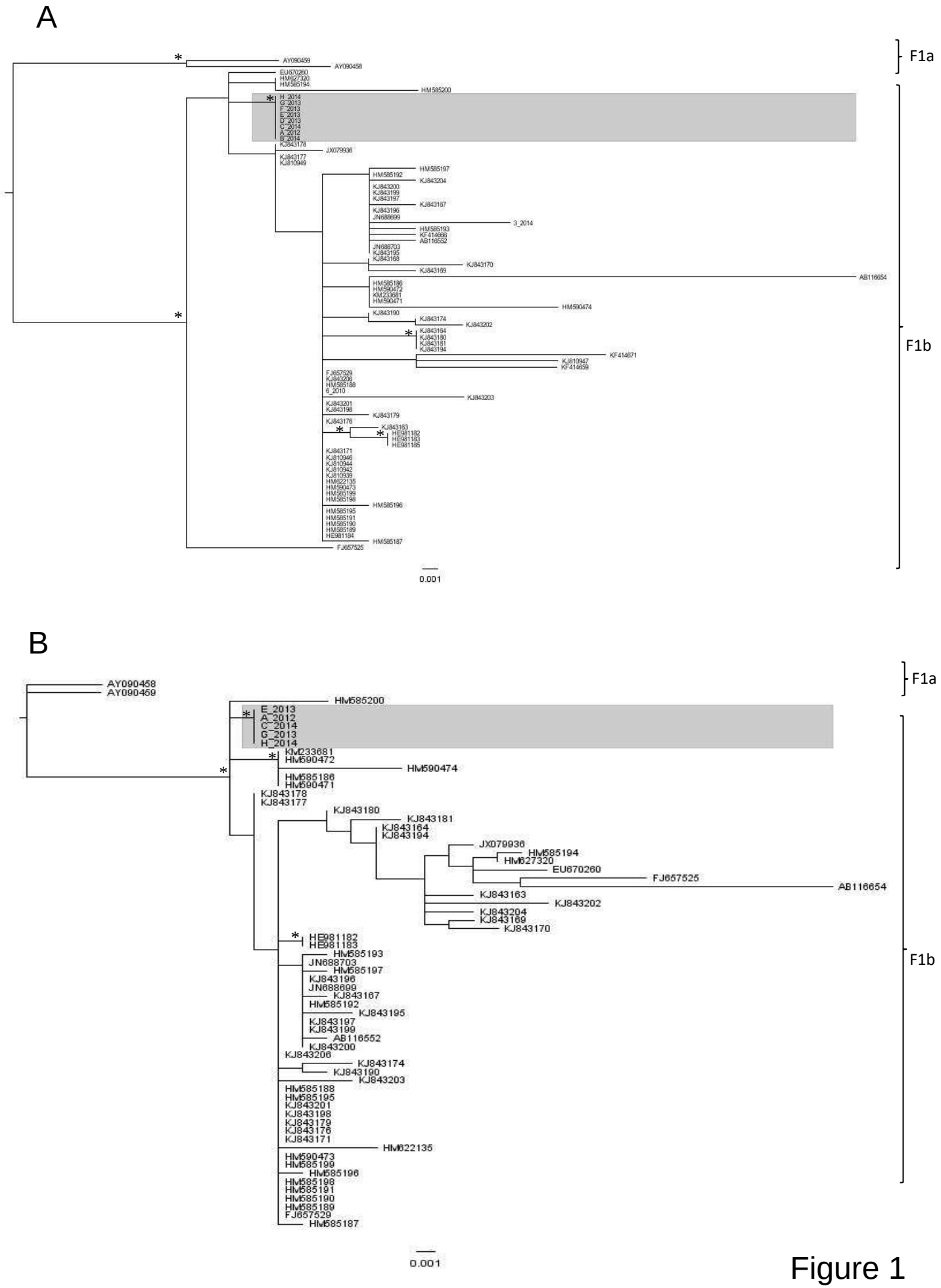

A
F1a
F1b
AY090459
AY090458
EU670260
HM627320
HM585194
HM585200
H_2014
G_2013
F_2013
E_2013
D_2013
C_2014
A_2012
B_2014
KJ843178
JX079936
KJ843177
KJ810949
HM585197
HM585192
KJ843204
KJ843200
KJ843199
KJ843197
KJ843167
KJ843196
JN688699
3_2014
HM585193
KF414666
AB116552
JN688703
KJ843195
KJ843168
KJ843170
KJ843169
AB116654
HM585186
HM590472
KM233681
HM590471
HM590474
KJ843190
KJ843174
KJ843202
KJ843164
KJ843180
KJ843181
KJ843194
KF414671
KJ810947
KF414659
FJ657529
KJ843206
HM585188
6_2010
KJ843203
KJ843201
KJ843198
KJ843179
KJ843176
KJ843163
HE981182
HE981183
HE981185
KJ843171
KJ810946
KJ810944
KJ810942
KJ810939
HM622135
HM590473
HM585199
HM585198
HM585196
HM585195
HM585191
HM585190
HM585189
HE981184
HM585187
FJ657525
0.001
B
F1a
F1b
AY090458
AY090459
HM585200
E_2013
A_2012
C_2014
G_2013
H_2014
KM233681
HM590472
HM590474
HM585186
HM590471
KJ843178
KJ843177
KJ843180
KJ843181
KJ843164
KJ843194
JX079936
HM585194
HM627320
EU670260
FJ657525
AB116654
KJ843163
KJ843202
KJ843204
KJ843169
KJ843170
HE981182
HE981183
HM585193
JN688703
HM585197
KJ843196
JN688699
KJ843167
HM585192
KJ843195
KJ843197
KJ843199
AB116552
KJ843200
KJ843206
KJ843174
KJ843190
KJ843203
HM585188
HM585195
KJ843201
KJ843198
KJ843179
KJ843176
KJ843171
HM622135
HM590473
HM585199
HM585196
HM585198
HM585191
HM585190
HM585189
FJ657529
HM585187
0.001

Figure 1

ACCEPTED MANUSCRIPT

Supplementary Methods I

Population-based sequencing of HBV Reverse Transcriptase, HBsAg, Basal Core Promoter/Pre-Core/Core.

The population-based sequencing of HBV Reverse-Transcriptase (RT), Basal Core Promoter [BCP]/Pre-Core/Core was performed on plasma samples, following home-made protocols as described below [6,7]. Due to the overlapping between the RT and HBsAg genes, RT-sequences were also used to analyze the HBsAg region. RT/HBsAg population-based sequencing was available for 8/8 patients, and BCP/Pre-Core/Core population-based sequencing for 5/8 patients. Mutations were defined as amino acid (aa) variation from the HBV F1b reference sequence.

Details of the protocols for the Population-based sequencing of HBV Reverse Transcriptase, HBsAg, Basal Core Promoter/Pre-Core/Core were reported below.

HBV DNA Extraction.

HBV-DNA was extracted using a commercially available kit (QIAmp DNA blood mini-kit, Qiagen Inc., USA).

Protocol for population-based sequencing of HBV Reverse Transcriptase/HBsAg.

Extracted HBV DNA was amplified with Amplitaq-Gold polymerase using the following primer pairs: F1-5'GGTCACCATATTCTTGGGAA and R1-5'GTGGGGGTTGCGTCAGCAAA. Polimerase chain reaction (PCR) conditions were: one cycle at 93 °C for 12 min, 40 cycles (94 °C 50 s, 57 °C 50 s, 72 °C 1 min and 30 s), and a final cycle at 72 °C for 10 min. For samples with serum low HBV-DNA, 2 additional heminested-PCR were performed, starting from the same first amplicon: eminested_1 used the following primer pairs (F1-5'GGTCACCATATTCTTGGGAA and R2-GAGGACAAACGGGCAACATACCTT and eminested-2 used F2-GTTGACAAGAATCCTCACAATA and R1-5'GTGGGGGTTGCGTCAGCAAA. Both eminested-PCRs conditions were: one cycle at 93 °C for 12 min, 35 cycles (94 °C 50 s, 56 °C 50 s, 72 °C 1 min), and a final cycle at 72 °C for 10 min. PCR-products were purified and sequenced by

using eight different overlapping sequence-specific primers, a BigDye terminator v. 3.1 cycle sequencing kit (Applied-Biosystems FosterCity, CA) and an automated sequencer (Genetic Analyzer 3130XL). The sequences were analyzed using SeqScape-v.2.5 software. The quality endpoint for each individual gene was ensured by a coverage of the S gene sequence by at least two segments. Sequences having a mixture of wild-type and mutant residues at single positions were considered to have the mutant(s) at that position.

Protocol for population-based sequencing of HBV Basal Core Promoter, Pre-Core, Core region.

Extracted HBV DNA was amplified with Amplitaq-Gold polymerase using the following primer pairs: F1-5' CCCAAGGTCTTACATAAGAGGACT and R1-5' CCACCTTATGAGTCCAAGGAA. PCR and Nested-PCR conditions were: one cycle at 94 °C for 10 min, 40 cycles (94 °C 30 s, 54 °C 30 s, 72 °C 1 min), and a final cycle at 72 °C for 10 min. For samples with serum low HBV-DNA, one additional nested-PCR was performed, used the following primer pairs (F2-5' TTAAAGACTGGGAGGAGTTGGG and R1-5' CCACCTTATGAGTCCAAGGAA).

PCR-products were purified and sequenced by using six different overlapping sequence-specific primers, a BigDye terminator v. 3.1 cycle sequencing kit (Applied-Biosystems FosterCity, CA) and an automated sequencer (Genetic Analyzer 3130XL). The sequences were analyzed using SeqScape-v.2.5 software. The quality endpoint for each individual gene was ensured by a coverage of the S gene sequence by at least two segments. Sequences having a mixture of wild-type and mutant residues at single positions were considered to have the mutant(s) at that position.

The list of drug resistance mutations in RT and immune escape mutations in HBsAg was retrieved at the following website http://hbv.bioinf.mpi-inf.mpg.de/index.php updated on October 2013.

Supplementary Methods II

Phylogenetic analysis. HBV RT-sequences, obtained by population-based sequencing, were aligned using CLUSTAL X software [8], and manually edited with the Bioedit program [9].

To define the sequence inter-relationships among the 8 HBV isolates and their HBV sub-genotype, a neighbor joining (NJ) tree [10] was constructed using a first dataset containing HBV RT reference sequences of the major HBV-genotypes, and 659 full-length RT-sequences (1,342bp) obtained from routine laboratory testing at the Virology Unit Hospital "Tor Vergata", in Rome, Italy. The dataset of 659 RT-sequences included the RT-sequences of the 8 patients involved in the cluster. The reliability of the branching orders was assessed by bootstrap analysis of 1000 replicates. HBV reference sequences were retrieved from GenBank (please, cite SM_III for the accession numbers). The robustness of the epidemiological cluster was further tested using the maximum likelihood (ML) method. This was inferred by the PhyML program (http://www.atgc-montpellier.fr/phyml/) using the GTR+I+Γ nucleotide substitution model. The simplest model that adequately fitted the sequence data was selected according to the Akaike Information Criterion (AIC) [11] included in the MEGA package (version 6.0). Robustness of the phylogenetic clades was evaluated by bootstrap analysis (1,000 replicates). For the maximum likelihood phylogenetic tree, the dataset included 80 full-length HBV F1b RT-sequences from: 70 isolates downloaded from GenBank, the 8 individuals involved in the cluster, and 2 individuals collected from routine laboratory testing at the University Hospital "Tor Vergata". Two HBV F1a RT-sequences were used as outgroup.

For 5 out of 8 patients, population-based BCP/Pre-Core/Core-sequences were available. These BCP/Pre-Core/Core sequences were concatenated to the corresponding RT-sequences, and used to infer the phylogenetic tree by ML method (bootstrap analysis based on 1,000 replicates). In this ML phylogenetic tree, additional 58 F1b and 2 F1a BCP/Pre-Core/Core/RT-sequences (downloaded from GenBank) were included.

The accession number of the sequences downloaded from GenBank are the following: JX849628.1 (Genotype-A); AY128092.1 (Genotype-A); AB033554.1 (Genotype-B); AB014381.1 (Genotype-C); V01460.1 (Genotype D3), AM494706.1 (Genotype-E); AF405706.1 (Genotype-G);

AY090454.1 (Genotype-H); HE974366 (Genotype-F2b), HE974368 (Genotype-F4), CS409748 (Genotype-F3), CS388978 (Genotype-F3), AB365453 (Genotype-F4), AF223965 (Genotype-F4), AY311369 (Genotype-F2a), DQ899142 (Genotype-F2a), DQ899144 (Genotype-F2b), AB036920 (Genotype-F3), AB116549 (Genotype-F3), AY090459 (Genotype-F1a), AY090458 (Genotype-F1a).

The accession number of F1b/RT sequences are the following: AY090459, AY090458, EU670260, HM627320, HM585194, HM585200, KJ843178, JX079936, KJ843177, KJ810949, HM585197, HM585192, KJ843204, KJ843200, KJ843199, KJ843197, HM585193, KJ843196, JN688703, KJ843167, KF414666, AB116552, JN688699, KJ843195, KJ843168, KJ843170, KJ843169, AB116654, HM585186, HM590472, KM233681, HM590471, HM590474, KJ843190, KJ843174, KJ843202, KJ843194, KJ843180, KJ843181, KJ843164, KJ843203, KJ843201, KJ843198, KJ843179, KJ843176, KJ843163, HE981182, HE981183, HE981185, KJ843171, KJ810946, KJ810944, KJ810942, KJ810939, HM622135, HM590473, HM585195, HM585198, HM585196, HM585199.

Genetic distances were calculated in the set of RT/HBsAg-sequences and BCP/Pre-Core/Core-sequences using MEGA 6.0 based on Kimura-2 parameter (K2P) model [12,13].

To avoid potential contaminations identical sequences amplified in the same run were excluded.

Supplementary Methods III

Ultra-deep pyrosequencing (UDPS) of HBV Reverse Transcriptase, HBsAg, Basal Core Promoter/Pre-Core/Core.

The extent of genetic heterogeneity in RT, HBsAg, BCP/Pre-Core/Core was also investigated by UDPS (Roche 454-Junior). RT/HBsAg UDPS was available for 8/8 patients, and BCP/Pre-Core/Core UDPS for 4/8 patients.

The RT/HBsAg and BCP/Pre-Core/Core-sequences obtained by UDPS were used to define the intra-patient prevalence of mutations in each patient analyzed. Mutations were defined as aa

variation from the HBV F1b reference sequence. Only mutations detected in ≥ 4 reads in both forward and reverse primers were considered for the analysis.

Overall, the mean ± standard deviation coverage per aminoacid position was 6,809±2,967 reads for RT/HBsAg, 6,841±1,682 reads for Pre-core, and 4,569±1,738 reads for Core. For BPC the variability was evaluated per each nucleotide position with a mean ± standard deviation coverage of 6,238±1,207 reads.

UDPS protocols for RT/HBsAg and BCP/Pre-Core/Core are reported below.

Protocol for ultra-deep pyrosequencing (UDPS) of HBV Reverse Transcriptase/HBsAg.

PCR-barcoded primers were designed to amplify four partially overlapping amplicons covering the full-length RT region, and the overlapping complete region of the HBsAg. In particular, four fragments of 424, 332, 384 and 427 nucleotides, encompassing the RT and HBsAg genomic region, were generated by nested PCR using barcoded-modified primers after a first amplification. Nested-PCR was performed with the Fast Start HiFi PCR system (Roche Diagnostics, Mannheim, Germany) under the following conditions: 1 cycle of 94°C for 3 min, 30 cycles of 94°C for 30 s, 60°C for 30 s, 72°C for 35sec, followed by a final extention at 72°C for 7 min. The amplified products were purified using Agencourt AMPure XP PCR purification beads (Beckman Coulter, Brea, CA) and quantified with the Quant-iT PicoGreen dsDNA Assay Kit (Life Technologies, Eugene, Oregon USA) by GloMax Multidetection System (Promega, Wisconsin USA). Equimolar pooled PCR products were clonally amplified on captured beads in water-oil emulsion microreactors (1.2 DNA-copies/beads). A total of 500,000 DNA-enriched beads were deposited in the GS Junior Titanium PicoTiterPlate device and pyrosequenced in both forward and reverse directions. The 200 nucleotide cycles were performed in a 10h sequencing run.

Protocol for ultra-deep pyrosequencing (UDPS) of HBV Basal Core Promoter, Pre-Core, Core region.

PCR-barcoded primers were designed to amplify four partially overlapping amplicons covering the BCP/Pre-Core/Core genomic region. In particular, four fragments of 470, 503, 334 and 474 nucleotides encompassing BCP/Pre-Core/Core genomic region were generated by Nested PCR.

The 4 amplicons were performed with the Fast Start HiFi PCR system (Roche Diagnostics, Mannheim, Germany). For the first amplicon (470 nucleotides fragment) the following conditions were used: 1 cycle of 94°C for 3 min, 30 cycles of 94°C for 30 s, 56°C for 30 s, 72°C for 35sec, followed by a final extention at 72°C for 7 min. For the other three amplicons, the PCR conditions were superimposable, with the exception of the annealing temperature that was of 64°C, 55°C and 55°C, respectively.

The amplified products were purified using Agencourt AMPure XP PCR purification beads (Beckman Coulter, Brea, CA) and quantified with the Quant-iT PicoGreen dsDNA Assay Kit (Life Technologies, Eugene, Oregon USA) by GloMax Multidetection System (Promega, Wisconsin USA). Equimolar pooled PCR products were clonally amplified on captured beads in water-oil emulsion microreactors (1.2 DNA-copies/beads). A total of 500,000 DNA-enriched beads were deposited in the GS Junior Titanium PicoTiterPlate device and pyrosequenced in both forward and reverse directions. The 200 nucleotide cycles were performed in a 10h sequencing run.

Bioinformatic Analysis of RT/HBsAg and BCP/Pre-Core/Core UDPS Sequences

The UDPS sequences obtained after 454-pyrosequencing were de-multiplexed and quantified using the standard flowgram format (SFF) tool by Roche. By an home-made pearl script calling for Blast algorithm, all reads were checked for homology with reference strains (>80%) and the orientation was discerned [1]. After this filtering, SHORAH package 0.5.1 [2] were used to correct sequences for homolymeric region associated errors, using as reference the sequences obtained from population sequencing. Subsequently the sequences were clustered, taking into account for reads redundancy and forward/reverse directions, using Exonerate 2.2.0 package [3] and aligned against the HBV F1b reference sequences using Needleman–Wunsch algorithm (Emboss package)

ACCEPTED MANUSCRIPT

[4, 5]. The final alignments were manually checked for insertion or deletion in homopolymeric region that would results in a frame shift. At each position the coverage and the amino acid variants were evaluated by a home-made pearl script. Phylogenetic analyses were performed to avoid any possible contamination or mixing among the pooled samples analyzed.

Sensitivity of UDPS for detecting minor variants

The frequency of errors resulting from amplification and deep sequencing was assessed using a plasmid containing the HBsAg gene as control. Mismatch nucleotide error rate was very low after sequencing correction (<0.006%). Furthermore, considering only the variants covered in both orientations in more than 4 reads, we never found mismatch in the control plasmid, thus only variants with these characteristic were considered reliable for the analyses.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

References

1. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol, **1990**; 215: 403-410.

2. Zagordi O, Bhattacharya A, Eriksson N, Beerenwinkel N. ShoRAH: estimating the genetic diversity of a mixed sample from next-generation sequencing data. BMC Bioinformatics, **2011**; 12: 119.

3. Slater GS, Birney E. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics, **2005**; 6: 31.

4. Needleman SB, Wunsch CD. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol, **1970**; 48: 443-453.

5. Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet, **2000**; 16: 276-277.

ACCEPTED MANUSCRIPT

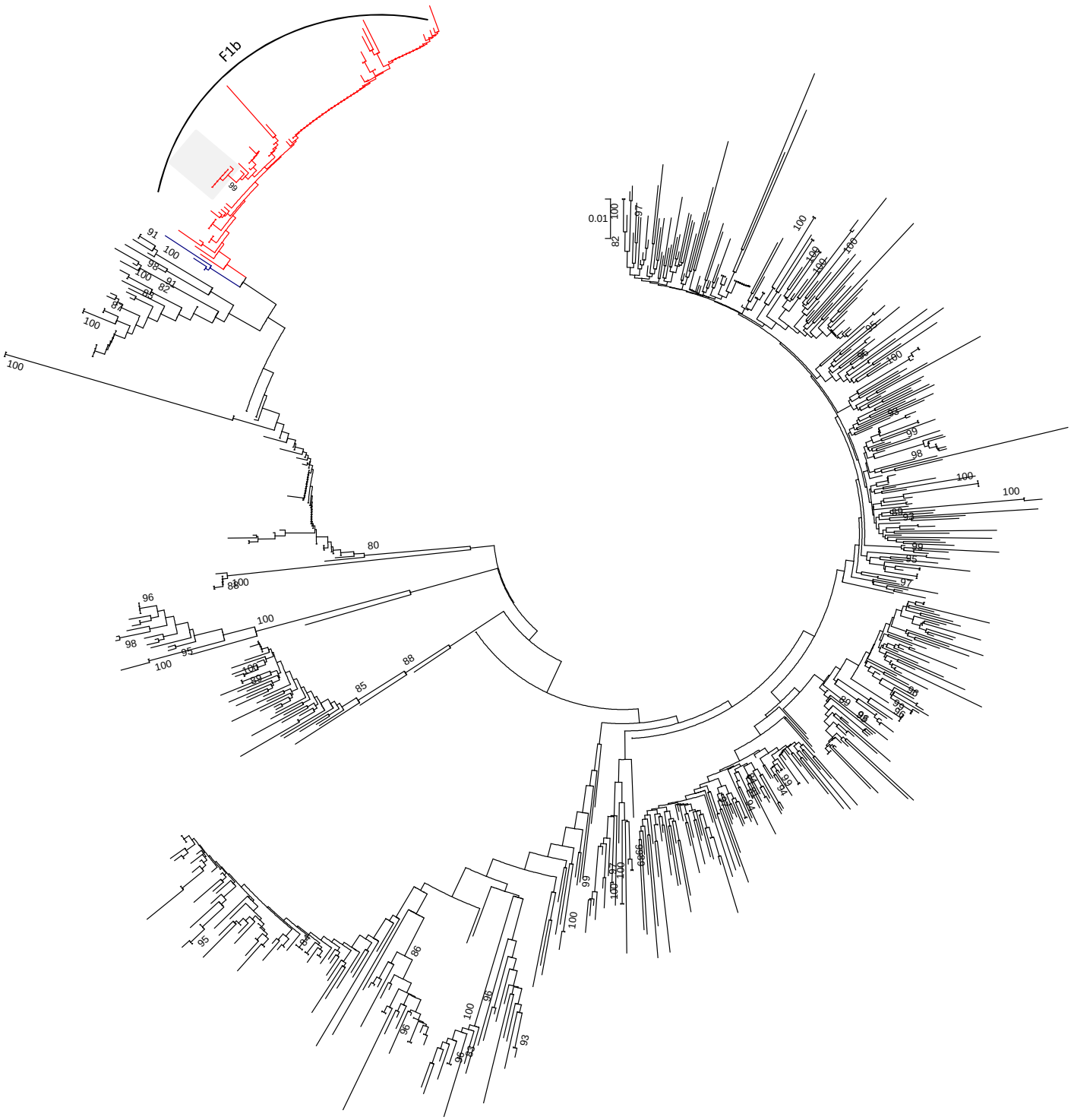


Figure S1