Origin and evolutionary dynamics of Hepatitis B virus (HBV) genotype E in Madagascar

Alessandra Lo Presti^{a†}, Soa Fy Andriamandimby^{b†}, Alessia Lai^c, Silvia Angeletti^d, Eleonora Cella^{a,e}, Giovanni Mottini^f, Michele Pier Luca Guarino^g, Claudia Balotta^c, Massimo Galli^c, Jean-Michel Heraud^b , Gianguglielmo Zehender^c and Massimo Ciccozzi^{a,f}

^aEpidemiology Unit, Department of Infectious, Parasitic and Immune-Mediated Diseases, Istituto Superiore di Sanità, Rome, Italy; ^bVirology Unit, Institut Pasteur of Madagascar, Antananarivo, Madagascar; ^cDepartment of Biomedical and Clinical Sciences “Luigi Sacco”, Infectious Diseases and Immunopathology Section, ‘L. Sacco’ Hospital, University of Milan, Milan, Italy; ^dClinical Pathology and Microbiology Laboratory, University Hospital Campus Bio-Medico of Rome, Rome, Italy; ^ePublic Health and Infectious Diseases, Sapienza University, Rome, Italy; ^fUniversity Hospital Campus Bio-Medico, Rome, Italy; ^gDigestive Disease Unit of Campus Bio Medico University of Rome, Rome, Italy

ABSTRACT

Africa is one of the endemic regions of HBV infection. In particular, genotype E is highly endemic in most of sub-Saharan Africa such as West African countries where it represents more than 90% of total infections. Madagascar, which is classified as a high endemic area for HBV and where the most prevalent genotype is E, might play a relevant role in the dispersion of this genotype due to its crucial position in the Indian Ocean. The aim of this study was to investigate the origin, population dynamics, and circulation of HBV-E genotype in Madagascar through high-resolution phylogenetic and phylodynamic approaches. The phylogenetic tree indicated that Malagasy isolates were intermixed and closely related with sequences mostly from West African countries. The Bayesian tree highlighted three statistically supported clusters of Malagasy strains which dated back to the years 1981 (95% HPD: 1971–1992), 1986 (95% HPD: 1974–1996), and 1989 (95% HPD: 1974–2001). Population dynamics analysis showed an exponential increase in the number of HBV-E infections approximately from the year 1975 until 2000s. The migration analysis was also performed and a dynamic pattern of gene flow was identified. In conclusion, this study confirms previous observation of HBV-E circulation in Africa and expands these findings at Madagascar demonstrating its recent introduction, and highlighting the role of the African countries in the spread of HBV-E genotype. Further studies on molecular epidemiology of HBV genotype E are needed to clarify the evolutionary history of this genotype.

KEYWORDS

Hepatitis B virus;
Madagascar; phylogeny;
phylodynamics; migration
analysis

Introduction

Hepatitis B virus (HBV) is a DNA virus belonging to *Hepadnaviridae* family and is one of the major causative agents of acute and chronic liver diseases in the world.

HBV is characterized by high genetic variability due to the use of an RNA intermediate and reverse transcriptase during replication. Phylogenetic analysis of HBV full-length genomes has led to the classification of HBV into 10 genotypes (A–J) that have been described so far further segregate into a number of subgenotypes which have distinct ethno-geographic distribution [1]. Recent study has hypothesized that HBV has been co-expanding and comigrating with human population, then this distribution of subgenotypes might be the result of the main modern human migration [2]. The different geographical manifestation of genotypes shows to affect the clinical manifestation as well as the course of HBV infection.

Africa is one of the highly endemic regions of HBV infection. Genotypes E, A, and D predominate in this

area depending on the region [3]. Genotype E is highly endemic in most of sub-Saharan Africa. Despite its high prevalence and wide geographic spread throughout large parts of Africa, HBV-E has a surprisingly low mean genetic diversity of only 1.75% over the full-length genome [4] supporting its recent emergence. Genotype E was by far the most prevalent genotype in Mali, Burkina Faso, Nigeria, Togo, Benin, Central African Republic, and The Democratic Republic of the Congo [5–8] with a prevalence ranging from 64 to 100%. Genotype E was also reported from other West African countries like Senegal [9], Ivory Coast [10], Ghana [11], the Gambia [12], Angola, and Namibia [13] where it represents more than 90% of total infections. As a result of the co-circulation of genotype A and E, recombination; a high number of mixed infections were detected in Cameroon where both genotypes are prevalent [5,14,15]. In addition, some strains were found further east like in Mozambique, Madagascar [15], and Tunisia [16]. Genotype E has not been found

outside of Africa, except for a few rare cases mostly in individuals with an African origin. In Europe, sporadic genotype E strains were reported in, e.g. France [16,17], Italy [18], Spain [19], Belgium [20], and Netherlands [21], and in the Americas, in Argentina [22], Brazil [23], and U.S.A [24].

Madagascar is classified as a high endemic area for HBV, and around 5% of the urban population and 30% of the rural population are chronic HBsAg carriers [25–27]. A recent study indicated that genotype E is the most prevalent in Madagascar followed by genotype A1 and D [28].

However, HBV is hyperendemic in Africa, there is a paucity of data on genotypes and their distribution in this continent. In this context, Madagascar might play a relevant role in the dispersion of genotype E due to its crucial position in the Indian Ocean contributing to the HBV-E spread. Only limited data are available on the date of emergence of this genotype in Africa and few data are present for Madagascan strains.

The aim of this study was to investigate in detail the origin, the population dynamics, and circulation of HBV-E genotype in Madagascar through high-resolution phylogenetic and phylodynamic approaches. Moreover migration analysis was also performed.

Material and methods

Phylogenetic datasets

For the purpose of phylogenetic analysis three data-sets were built. The first data-set included 445 HBV-E genotype P-PreS/S (Pre/S/S region with partial overlapping of P gene) sequences, 366 of which downloaded from NCBI (<http://www.ncbi.nlm.nih.gov/>) with associated information regarding the geographic location (CM, Cameroon, $n = 33$; GH, Ghana, $n = 22$; CD, Democratic Republic of the Congo, $n = 4$; NG, Nigeria, $n = 85$; HT, Haiti, $n = 10$; GN, Guinea, $n = 66$; AO, Angola, $n = 9$; SD, Sudan, $n = 23$; BN, Benin, $n = 12$; TG, Togo, $n = 22$; BF, Burkina Faso, $n = 10$; ML, Mali, $n = 18$; CF, Central African Republic, $n = 27$; ZA, South Africa, $n = 2$; CI, Cote d'Ivoire, $n = 7$; CO, Colombia, $n = 9$; SN, Senegal, $n = 3$; LR, Liberia, $n = 1$; GM, Gambia, $n = 3$), and 79 were isolates from Madagascar obtained in a previous work [28]. These 79 isolates were collected from different areas in Madagascar (urban, suburban, rural; north, center, and south) as previously described [28].

The first data-set was used to investigate both the phylogenetic relationships between Malagasy HBV-E isolates and HBV-E migration that flows from and to other African countries. The second data-set included 296 (of the 366 sequences of the first data-set) HBV-E genotype P-PreS/S sequences with known sampling date and location. This data-set was used to estimate the evolutionary rate.

The third data-set included only the 79 HBV-E genotype P-PreS/S isolates from Madagascar and was used to

perform the time-scaled tree and population dynamics analysis.

Phylogenetic analysis

The sequences of all data-sets were aligned and compared using Bioedit v7.2.5 (<http://www.mbio.ncsu.edu/bioedit/bioedit.html>) and the integrated Clustal W program followed by manual editing. For all the data-sets the hierarchical likelihood ratio test method, implemented in Model-Test version 3.07 [29] was used to select the evolutionary model that best fitted the sequences data. The General Time Reversible plus Gamma plus invariant sites (GTR + G + I) model was selected for each data-set.

Mega v. 7 was used to build the neighbor-joining phylogenetic tree of the first (<http://www.megasoftware.net/>) [30] with the LogDet method by inferring 1000 bootstrap replicates.

Evolutionary Rate estimation, time-scaled phylogeny and population dynamic analysis

The evolutionary rate was estimated on the second data-set using a Markov Chain Monte Carlo (MCMC) Bayesian approach implemented on the program BEAST v 1.8.2, under GTR + G + I model, using both a strict and an uncorrelated log-normal relaxed clock model. As BEAST uses the Bayesian Markov-Chain-Monte-Carlo approach to sample states from a probability distribution and samples the root position of the phylogeny along with the rest of the nodes, no outgroup strains are necessary for analysis. Furthermore, the inclusion of the sampling year of the individual strains allows a time-measured inference with BEAST. Four different coalescent priors were evaluated by enforcing either a strict or relaxed molecular clock: constant population size, exponential, and two non-parametric models of population changes (Bayesian Gaussian Markov Random Field skyride plot and Bayesian skyline plot). For each model, a Markov Chain Monte Carlo (MCMC) was run 50 million generations, sampling every 5000 steps. Models were compared pairwise by calculating the Bayes factor with marginal likelihoods estimated via path sampling and stepping-stone methods implemented in BEAST 1.8.2. [31–33] The selected molecular clock/demographic model (relaxed clock with Bayesian Skyline Plot coalescent prior) was then used for the further Bayesian analysis. The MCMC chains were run for at least 200 million generations, and sampled every 20,000 steps. Convergence was assessed on the basis of the effective sampling size. Only effective sampling size values of > 200 were accepted.

The dated tree was inferred on the third data-set using a MCMC Bayesian approach with BEAST v 1.8.2, under the GTR + G + I model, using the evolutionary rate, the clock and demographic model previously estimated. Uncertainty in the estimates was indicated by 95%

highest posterior density (95% HPD) intervals. Statistical support for specific clades was obtained by calculating the posterior probability of each monophyletic clade. The final trees were manipulated in FigTree version 1.4.2 for display purposes.

Population dynamics was also analyzed in the third data-set by performing the Bayesian skyline plot (BSP) with BEAST v1.8.2.

Migration analysis (Gene Flow)

The Mac Clade v 4 program (Sinauer Associates, Sunderland, MA) was used on the first data-set to test viral gene in/out flow among distinct HBV E genotype subpopulations within different geographic areas, using a modified version of the Slatkin and Maddison test as described below [34]. Specifically, a gene flow analysis was performed classifying the HBV genotype E sequences into different groups based on the sampling location of sequences. The analysis was conducted grouping the sequences into eight groups (Figure 4(a)) according to the geographical proximity of the countries from which the sequences were sampled (group: **1**, including sequences from Guinea, Mali, Cote d'Ivoire, Senegal, Liberia, and The Gambia; group: **2**, Ghana, Benin, Nigeria, Togo, and Burkina Faso; group **3**, Central Africa including sequences from Cameroon, Democratic Republic of the Congo, and Central African Republic; group **4**, East Africa including sequences from Sudan; group **5**, South Africa including sequences from South Africa and Angola; group **6**, including the isolates from Madagascar; group **7**, Central America including sequences from Haiti; group **8**, South America including sequences from Colombia). A one-character data matrix was obtained from the original data-set by assigning to each taxon in the tree a one-letter code indicating its group of origin. Then, the putative origin of each ancestral sequence (i.e. internal node) in the tree was inferred by finding the most parsimonious reconstruction (MPR) of the ancestral character. The final tree-length, i.e. the number of observed migrations in the genealogy, was compared to the tree-length distribution of 10,000 trees, after random joining-splitting. Observed genealogies significantly shorter than random trees indicate the presence of subdivided populations with restricted gene flow. Specific migrations among different areas (character states) were traced with the State changes and stasis tool (Mac Clade software).

Results

Phylogenetic analysis, evolutionary rate estimation, time-scaled phylogeny, and population dynamics analysis

The neighbor-joining phylogenetic tree of the first data-set (Figure 1) showed many statistically supported

clusters. Analyzing this tree, it was possible to notice that the Malagasy isolates were intermixed and closely related with sequences mostly from West African countries. A statistically supported cluster was also found among a Malagasy isolate and sequences from Colombia, one sequence from Nigeria, and two from Guinea.

On the second data-set, a Bayesian approach was implemented to estimate the evolutionary rate. The Bayes Factor comparison between strict versus relaxed clock showed that, the relaxed clock fitted the data significantly better than the strict clock ($\ln BF > 80$), in all the demographic models used, based on both path sampling and stepping-stone estimates of marginal likelihoods. Under the relaxed clock, the BF analysis showed that the Bayesian skyline plot (BSP) was better than other models ($\ln BF > 10$) in all the comparisons. The estimated mean value of the HBV Pre S/S gene evolutionary rate for the second data-set was 1.9×10^{-4} substitutions per site and year (s/s/y) (95% HPD: 0.67×10^{-4} – 3.84×10^{-4}).

Figure 2 showed the Bayesian maximum clade credibility tree and the tMRCA estimates on the third data-set. The root of the tree dated 159 years ago corresponding to the year 1854 (95% HPD: 1724–1976).

From this ancestor, three statistically supported (posterior probability, $pp > 0.90$) clusters were found. Cluster A dated back to the year 1981 (95% HPD: 1971–1992) and included six isolates from the rural zone of Madagascar (five collected from the northern part of Madagascar and one from the southern).

Cluster B included three isolates from the center of Madagascar, all from the rural zone; it dated back to the year 1989 (95% HPD: 1974–2001).

Finally, cluster C was composed of three sequences, two from the southern part and one from the center of Madagascar, all from the urban zone and dated back to 1986 (95% HPD: 1974–1996).

Population dynamics analysis performed on the third data-set showed that the effective number of infections remained constant until approximately the year 1975 (Figure 3), then the curve showed a rapid and exponential increase (of approximately 2 log) in the number of HBV-E infections until 2000s reaching a plateau.

Migration Analysis (Gene Flow)

The results of the migration analysis performed in this study were reported in Figure 4(b). The null hypothesis of panmixia (i.e. no population subdivision or complete intermixing of sequences from different geographic areas) was rejected by the randomization test ($P < 0.0001$).

In Figure 4(b), the gene flow analysis showed a great dynamic migration flows among the different countries considered i.e. group 1 vs. group 2 (7.6%), 3 (6.1%), 6, and 7 (3%); group 2 vs. group 1 (13.6%), group 3 (10.6%), group 5 (9.1%), group 7 (3%), and group 8 (1.5%).

Notes: The asterisk (*) along the branches represents significant statistical support for the clade subtending that branch (bootstrap values > 70%). The tree was mid-point rooted. The scale at the bottom of the tree represents 0.007 nucleotide substitution per site. A two-letter code flanked by the number of sequences present in each portion of the tree, was used; for the coding of the countries: CM, Cameroon; GH, Ghana; CD, Democratic Republic of the Congo; NG, Nigeria; HT, Haiti; GN, Guinea; AO, Angola; SD, Sudan; BN, Benin; TG, Togo; BF, Burkina Faso; ML, Mali; CF, Central African Republic; ZA, South Africa; CI, Cote d'Ivoire; CO, Colombia; SN, Senegal; LR, Liberia; GM, Gambia; MG, Madagascar. Malagasy isolates were highlighted in bold and underlined. Some clade were collapsed.

Figure 2. Bayesian maximum clade credibility tree of the third data-set including 79 HBV-E genotype P-PreS/S isolates from Madagascar.

Notes: The asterisk (*) along the branches represents significant statistical support for the clade subtending that branch (posterior probability > 0.90). The scale at the bottom of the tree represents time in years. Main statistically supported cluster were indicated with letters. Some cluster were condensed. The tMRCA and 95% HPD – tMRCA for the internal nodes (cluster A, B, and C) were reported at the top left of the figure.

Figure 3. Bayesian skyline plot (BSP) of the third data-set including 79 HBV-E genotype P-PreS/S isolates from Madagascar.

Notes: The effective number of infections is reported on the Y-axis. Time is reported on the X-axis. The colored area corresponds to the credibility interval based on 95% highest posterior density interval (HPD).

Figure 4. (a) and (b) Migration pattern of HBV-E circulation performed on the first data-set which included 445 HBV-E genotype P-Pre S/S sequences, 366 of which downloaded from NCBI and almost all originating from African countries, and 79 isolates from Madagascar.

Notes: (a) geographic map representing the different groups (labeled with numbers) used for migration analysis and highlighted with different symbols. (b) Migration pattern of HBV-E circulation. The bubblegram shows the frequency of gene flow (migrations) to/from different geographic areas. The surface of each circle is proportional to the percentage of observed migrations. Migrations were inferred with a modified version of the Slatkin and Maddison algorithm. **1:** Guinea, Mali, Cote d'Ivoire, Senegal, Liberia, The Gambia; **2:** Ghana, Benin, Nigeria, Togo, Burkina Faso; **3:** Cameroon, Democratic Republic of the Congo, Central African Republic; **4:** Sudan; **5:** South Africa, Angola; **6:** Madagascar; **7:** Haiti; **8:** Colombia.

Of our interest, from group 6 (Madagascar) three migration flows were identified, specifically a 10.6% of gene flow to group 1 (Guinea, Mali, Cote d'Ivoire, Senegal, Liberia, and The Gambia); a 12.1% of gene flow to group 2 (Ghana, Benin, Nigeria, Togo, and Burkina Faso) and only a 1.5% of gene flow to group 4. On the contrary, Madagascar seems to receive fluxes mostly from group 1 (3%).

Discussion

Madagascar is considered a country with a high level endemicity for HBV [25–27,35]; inside this country the genotype E is the most prevalent, followed by genotype A1 and D [28].

In the present study, the origin, the population dynamics, and circulation of HBV-E epidemic in Madagascar were investigated by phylogenetic analysis in addition to migration analysis.

Our data showed that Malagasy isolates were inter-mixed and closely related with foreign sequences, mostly from West African countries. This is in agreement with

Dupinay et al [35], that already reported a high sequence similarity of genotype E strains circulating in Madagascar with West African strains. In our analysis was also found a statistically supported cluster among a Malagasy isolate and the sequences from Colombia, one sequences from Nigeria, and two from Guinea.

Previous studies reported a recent origin of HBV E genotype, most likely within the last 130 years (mean tMRCA of 174 years; 95% HPD: 36–441 years) and only after that the transatlantic slave-trade had come to an end [3]. A recent study by Zehender et al [1], described the phylodynamic of genotype E with a relatively recent evolutionary history among humans and correlated their explosive spread in Africa continent to highly efficient route of transmission such as the unsafe usage of needles during vaccination campaigns between 1920s and 1960s.

The mean evolutionary rate estimated in the present study was 1.9×10^{-4} substitutions/site/year (s/s/y) (95% HPD: 0.67×10^{-4} – 3.84×10^{-4}), comparable to that found by Andernach et al [3]. Our tMRCA estimation is

in line with previous data, as we reported an origin for HBV-E genotype of 159 years (95% HPD: 36.7–289.5) corresponding to the year 1854 (95% HPD: 1724–1976).

The time-scaled phylogeny reconstruction suggested that HBV genotype E had three introductions in Madagascar: one around 1981 (95% HPD: 1971–1992) which corresponds to cluster A, one in 1989 (95% HPD: 1974–2001) corresponding to cluster B and the third one in 1986 (95% HPD: 1974–1996) which corresponds to cluster C. The recent introduction of HBV-E in Madagascar is also confirmed by population dynamics showing that the effective number of infections remained constant until approximately the year 1975, and then the curve showed a rapid and exponential increase in the number of HBV-E infections until 2000s, reaching a plateau. This rapid and relatively recent expansion can be explained by the low genetic diversity of the HBV-E strains in Madagascar and by the main reported transmission route of the analyzed samples that occurs during perinatal and childhood periods [26].

Probably HBV-E reached Madagascar before 1970s, remained confined to isolates Malagasy communities for a quite long period of time and was then reintroduced into the Malagasy population only recently, as already reported for the general expansion of this genotype in other African countries [15] and as indicated by population dynamics.

However, despite the quite recent introduction of HBV genotype E, the estimate of the evolutionary rate and tMRCA fluctuated depending on many factors, such as the specific data-set used, the genotype, the specific genetic portion, the sampling dates of the sequences, and others. For other HBV genotypes, some authors showed a long evolution. Given the presence of genotype F among the Amerindian population which suggests the long evolution of this strain, some authors [36] obtained a tMRCA of 4400 years for the entire tree and 700–800 years for the different F sub-genotypes/clades. Other studies [37] correlated the pattern of HBV dispersion with human migrations and suggested that the virus originated about 33600 years ago (95% HPD 22–47.1 KYA). Another example was the HBV Genotype C4 isolated from indigenous Australians. Archeological evidence suggests the presence of Indigenous Australians in Australia from at least 40,000 years [38], and possibly 60,000 years ago [39]. Paraskevis et al. suggested that genotype C may be the oldest of the HBV genotypes, diverging from the other genotypes 15.3–38.9 thousand years ago [37].

The gene flow analysis of the present study reported the majority of fluxes from West Africa to different African countries. Madagascar seems to receive gene flow only from group 1 (Guinea, Mali, Cote d'Ivoire, Senegal, Liberia, and The Gambia). An exchange of gene flow was also reported from Madagascar to West Africa (group 1 and 2) and to East Africa (group 4), highlighting an important role of this country in the spread of HBV-E genotype. Additionally to the large diffusion of genotype

E in West Africa, data of previous studies related to the circulation of HCV genotype 2 seem to support previous human migration from West Africa to Madagascar [40].

The pattern of viral migration could also be probably explained by the history of Madagascar. The French colonization has led to preferential exchanges and population mobility. Military conscripts from West African colonies, called *Tirailleurs Se'ne'galais*, may have played a role. Tens of thousands of both *Tirailleurs* and Malagasy soldiers fought in the French colonial army in the trenches of World War I [41,42]. During World war II and until independency, a lot of West African individuals belonging to the '*tirailleurs sénégalais*' stayed enough time in Madagascar to have probable intercourses with Malagasy population.

In conclusion, the results presented here confirmed previous observation of HBV-E circulation in Africa and expand these findings in Madagascar. In this study, Malagasy isolates were mainly related to West African strains and HBV genotype E showed three different introductions in Madagascar. Population dynamics highlighted an increase in the number of HBV-E infections in Madagascar from 1975 to 2000s.

Further studies on molecular epidemiology of HBV genotype E are needed to clarify the evolutionary history of this genotype and to investigate the potential spread of this viral strain.

Acknowledgments

The contents of the manuscript are solely the responsibility of the authors and do not necessarily represent the official views of the Centers for Disease Control and Prevention, the Department of Health and Human Services or the Department of Homeland Security.

Disclosure statement

The authors report no declaration of interest.

Funding

The human specimens (serums) were collected as part of an integrative global project leads by the Institut Pasteur of Madagascar (the ZORA project) funded by different partners according to thematic; the US Centers for Disease Control and Prevention (Cooperative Agreement Number, 5U51IP000812-02); the Department of Homeland Security Center of Excellence for Emerging and Zoonotic Animal Diseases (CEEZAD) [grant number 2010-ST061-AG0001].

ORCID

Jean-Michel Heraud <http://orcid.org/0000-0003-1107-0859>

References

- [1] Zehender G, Ebranati E, Gabanelli E, et al. Enigmatic origin of hepatitis B virus: an ancient travelling companion or a recent encounter? *World J Gastroenterol*. 2014;20:7622–34.

- [2] Paraskevis D, Magiorkinis G, Magiorkinis E, et al. Dating the origin and dispersal of hepatitis B virus infection in humans and primates. *Hepatology*. 2013;57:908–16.
- [3] Andernach IE, Hunewald OE, Muller CP. Bayesian inference of the evolution of HBV/E. *PLoS One*. 2013;8:e81690.
- [4] Andernach IE, Hübschen JM, Muller CP. Hepatitis B virus: the genotype E puzzle. *Rev Med Virol*. 2009;19:231–40.
- [5] Olinger CM, Venard V, Njayou M, et al. Phylogenetic analysis of the precore/core gene of hepatitis B virus genotypes E and A in West Africa: new subtypes, mixed infections and recombinations. *J Gen Virol*. 2006;87:1163–73.
- [6] Mulders MN, Venard V, Njayou M, et al. Low genetic diversity despite hyperendemicity of hepatitis B virus genotype E throughout West Africa. *J Infect Dis*. 2004;190:400–8.
- [7] Bekondi C, Olinger CM, Boua N, et al. Central African republic is part of the West-African hepatitis B virus genotype E crescent. *J Clin Virol*. 2007;40:31–7.
- [8] Odemuyiwa SO, Mulders MN, Oyedele OI, et al. Phylogenetic analysis of new hepatitis B virus isolates from Nigeria supports endemicity of genotype E in West Africa. *J Med Virol*. 2001;65:463–9.
- [9] Vray M, Debonne JM, Sire JM, et al. Molecular epidemiology of hepatitis B virus in Dakar, Sénégal. *J Med Virol*. 2006;78:329–34.
- [10] Suzuki S, Sugauchi F, Orito E, et al. Distribution of hepatitis B virus (HBV) genotypes among HBV carriers in the cote d'Ivoire: complete genome sequence and phylogenetic relatedness of HBV genotype E. *J Med Virol*. 2003;69:459–65.
- [11] Candotti D, Opare-Sem O, Rezvan H, et al. Molecular and serological characterization of hepatitis B virus in deferred Ghanaian blood donors with and without elevated alanine aminotransferase. *J Viral Hepat*. 2006;13:715–24.
- [12] Dumpis U, Holmes EC, Mendy M, et al. Transmission of hepatitis B virus infection in Gambian families revealed by phylogenetic analysis. *J Hepatol*. 2001;35:99–104.
- [13] Kramvis A, Kew MC. Epidemiology of hepatitis B virus in Africa, its genotypes and clinical associations of genotypes. *Hepatol Res*. 2007 Jul;37(s1):S9–S19.
- [14] Kurbanov F, Tanaka Y, Fujiwara K, et al. A new subtype (subgenotype) Ac (A3) of hepatitis B virus and recombination between genotypes A and E in Cameroon. *J Gen Virol*. 2005;86:2047–56.
- [15] Kramvis A, Restorp K, Norder H, et al. Full genome analysis of hepatitis B virus genotype E strains from South-Western Africa and Madagascar reveals low genetic variability. *J Med Virol*. 2005;77:47–52.
- [16] Bahri O, Cheikh I, Hajji N, et al. Hepatitis B genotypes, precore and core promoter mutants circulating in Tunisia. *J Med Virol*. 2006;78:353–7.
- [17] Halfon P, Bourliere M, Pol S, et al. Multicentre study of hepatitis B virus genotypes in France: correlation with liver fibrosis and hepatitis B e antigen status. *J Viral Hepat*. 2006;13:329–35.
- [18] Palumbo E. Immigration and hepatitis B virus in Italy. Current state and perspectives. *Recenti Prog Med*. 2007;98:548–52.
- [19] Toro C, Jiménez V, Rodríguez C, et al. Molecular and epidemiological characteristics of blood-borne virus infections among recent immigrants in Spain. *J Med Virol*. 2006;78:1599–608.
- [20] Liu HF, Sokal E, Goubau P. Wide variety of genotypes and geographic origins of hepatitis B virus in Belgian children. *J Pediatr Gastroenterol Nutr*. 2001;32:274–47.
- [21] van Steenberg JE, Niesters HG, Op de Coul EL, et al. Molecular epidemiology of hepatitis B virus in Amsterdam 1992–1997. *J Med Virol*. 2002;66:159–65.
- [22] Mathet VL, Cuestas ML, Ruiz V, et al. Detection of hepatitis B virus (HBV) genotype E carried—even in the presence of high titers of anti-HBs antibodies—by an Argentinean patient of African descent who had received vaccination against HBV. *J Clin Microbiol*. 2006;44:3435–9.
- [23] Sitnik R, Sette Jr H Jr, Santana RA, et al. Hepatitis B virus genotype E detected in Brazil in an African patient who is a frequent traveler. *Braz J Med Biol Res*. 2007;40:1689–92.
- [24] Kato H, Gish RG, Bzowej N, et al. Eight genotypes (A–H) of hepatitis B virus infecting patients from San Francisco and their demographic, clinical, and virological characteristics. *J Med Virol*. 2004;73:516–21.
- [25] Boiesier P, Rabarijaona L, Piollet M, et al. Hepatitis B virus infection in general population in Madagascar: evidence for different epidemiological patterns in urban and in rural areas. *Epidemiol Infect*. 1996;117:133–7.
- [26] Migliani R, Rousset D, Rakoto-Andrianarivelo M, et al. Hepatitis B virus infection: a public health problem in Madagascar. *Arch Inst Pasteur Madagascar*. 2000;66:50–4.
- [27] Migliani R, Rakoto Andrianarivelo M, Rousset D, et al. Seroprevalence of viral hepatitis B in the city of Mahajanga, Madagascar in 1999. *Med Trop*. 2000;60:146–50.
- [28] Andriamandimby SF, Lo Presti A, Lai A, et al. Genetic diversity of hepatitis B virus (HBV) in Madagascar. *J Med Virol*. 2016 Dec;88(12):2138–44.
- [29] Posada D, Crandall KA. MODELTEST: testing the model of DNA substitution. *Bioinformatics*. 1998;14:817–8.
- [30] Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol*. 2016;33:1870–4.
- [31] Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol*. 2007;7:214.
- [32] Baele G, Lemey P. Bayesian evolutionary model testing in the phylogenomics era: matching model complexity with computational efficiency. *Bioinformatics*. 2013 Aug 15;29(16):1970–9.
- [33] Baele G, Li WL, Drummond AJ, et al. Accurate model selection of relaxed molecular clocks in bayesian phylogenetics. *Mol Biol Evol*. 2013 Feb;30(2):239–43.
- [34] Slatkin M, Maddison WP. A clastic measure of gene flow inferred from the phylogenies of alleles. *Genetics*. 1989;123:603–13.
- [35] Dupinay T, Restorp K, Leutscher P, et al. High prevalence of hepatitis B virus genotype E in Northern Madagascar indicates a West-African lineage. *J Med Virol*. 2010;82:1515–26.
- [36] Torres C, Leone FG, Pezzano SC, et al. New perspectives on the evolutionary history of hepatitis B virus genotype F. *Mol Phylogenet Evol*. 2011;59:114–22.
- [37] Paraskevis D, Magiorkinis G, Magiorkinis E, et al. Dating the origin and dispersal of hepatitis B virus infection in humans and primates. *Hepatology*. 2013;57:908–16.
- [38] O'Connell James F, Allen Jim. Dating the colonization of Sahul (Pleistocene Australia–New Guinea): a review of recent research. *J. Archaeol. Sci*. 2004;31:835–53.
- [39] Roberts RG, Jones R, Smith MA. Thermoluminescence dating of a 50,000-year-old human occupation site in northern Australia. *Nature*. 1990;345:153–6.
- [40] Markov PV, Pepin J, Frost E, et al. Phylogeography and molecular epidemiology of hepatitis C virus genotype 2 in Africa. *J Gen Virol*. 2009;90:2086–96.
- [41] Clayton A. France, soldiers and Africa. London: Brassey's Defence Publishers; 1988.
- [42] Echenberg M. Colonial conscripts: the Tirailleurs Se'ne'galais in French West Africa, 1857–1960. Portsmouth, NH: Heinemann Educational Books; 1991.