



## Genetic diversity of Hepatitis B Virus (HBV) in Madagascar

Journal:	<i>Journal of Medical Virology</i>
Manuscript ID	JMV-16-5204.R1
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	<p>Andriamandimby, Soa Fy; Institut Pasteur of Madagascar, Virology Unit  Lo Presti, Alessandra; Istituto Superiore di Sanità, Department of  Infectious, Parasitic and Immunomediated Diseases, Epidemiology Unit,  Reference Centre on Phylogeny, Molecular Epidemiology and Microbial  Evolution (FEMEM)  Lai, A.; University of Milan, Department of Clinical Sciences 'L. Sacco'  Hospital, Section of Infectious Diseases and Immunopathology  Olive, Marie-Marie; Institut Pasteur of Madagascar, Virology Unit  Angeletti, Silvia; University Hospital Campus Bio-Medico, Rome, Italy,  Clinical Pathology and Microbiology Laboratory  De Florio, Lucia; University Hospital Campus Bio-medico, Clinical Pathology  and Microbiology Laboratory  Cella, Eleonora; Istituto Superiore di Sanità, Malattie Infettive  Razafindramparany, Minoharimbola; Institut Pasteur de Madagascar,  Virology Unit  Ravalohery, Jean-Piere; Institut Pasteur de Madagascar, Virology Unit  Andriamamonjy, Seta; Institut Pasteur de Madagascar, Virology Unit  Gioffrè, Sonia; L. Sacco, University of Milan  Zehender, Gianguglielmo; University of Milan, Department of Clinical  Sciences-Infectious Diseases Section  Mottini, Giovanni; University Hospital Campus Bio-medico  Ciccozzi, Massimo; Istituto superiore di sanità, dipartimento di malattie  infettive e immunomediate, reparto di epidemiologia;  Heraud, Jean-Michel; Institut Pasteur de Madagascar, Virology</p>
Keywords:	Hepatitis B virus < Virus classification, Epidemiology, Evolution

SCHOLARONE™  
Manuscripts

**Genetic diversity of Hepatitis B Virus (HBV) in Madagascar**

Soa Fy Andriamandimby<sup>1#</sup>, Alessandra Lo Presti<sup>2#</sup>, Alessia Lai<sup>3</sup>, Marie-Marie Olive<sup>1</sup>, Silvia Angeletti<sup>4</sup>, Lucia De Florio<sup>4</sup>, Eleonora Cella<sup>2,5</sup>, Minoharimbola Razafindramparany<sup>1</sup>, Jean-Piere Ravalohery<sup>1</sup>, Seta Andriamamonjy<sup>1</sup>, Sonia Giofrè<sup>3</sup>, Gianguglielmo Zehender<sup>3</sup>, Giovanni Mottini<sup>6</sup>, Massimo Ciccozzi<sup>2,6\*°</sup> and Jean-Michel Heraud<sup>1\*</sup>.

1 Virology Unit, Institut Pasteur of Madagascar, Antananarivo, Madagascar

2 Department of Infectious, Parasitic and Immunomediated Diseases, National Institute of Health, Rome, Italy;

3 Department of Biomedical and Clinical Sciences 'L. Sacco', University of Milan, Milan, Italy;

4 Clinical Pathology and Microbiology Laboratory, University Hospital Campus Bio-Medico of Rome, Rome, Italy;

5 Public Health and Infectious Diseases, Sapienza University, Rome, Italy

6 University Hospital Campus Bio-Medico, Rome, Italy.

# These authors contributed equally

\* These authors contributed equally (co-senior authors)

° Corresponding author: Prof. Massimo Ciccozzi, Department of Infectious, Parasitic and Immunomediated Diseases, National Institute of Health, Rome, Italy  
massimo.ciccozzi@iss.it

**Running head:** Hepatitis B Virus in Madagascar

## Abstract

Hepatitis B virus (HBV) is a DNA virus belonging to *Hepadnaviridae* family. Chronic infection with HBV is one major risk factor of hepatic disease. In Madagascar, former studies classified the country as part of high endemic area, as HBV prevalence can reach 23% in general population. However, this prevalence differs largely between urban and rural areas and is estimated to be respectively 5% and 26%. The aims of the present study were to describe the genetic diversity of HBV strains from different regions of Madagascar, and to describe the viral gene flow throughout the country by using phylogenetic analysis. This is the first large-scale molecular and phylogenetic study analyzing HBV sequences from 28 different Malagasy areas, never sampled in the past. In this study, the most prevalent genotype/sub-genotypes was E. Migration analysis showed a gene flow from zone 3 (rural) to zone 2 (suburban) and a greater gene flow from the middle part of Madagascar to the north than to the south. It is important to study the HBV infections in Madagascar and to monitor the potential spread of this viral strain inside this country.

**Key words:** Hepatitis B virus, Epidemiology, Evolution

INTRODUCTION

Hepatitis B virus (HBV) is a DNA virus belonging to *Hepadnaviridae* family. At least 2 billion people alive today have been infected with HBV and of these, it is estimated that 240 million individuals are chronically infected with this virus [Ott et al., 2012]. Chronic infection with HBV is one major risk factor of hepatic disease, and may bring to liver cancer [Lavanchy et al., 2005]. Three levels of chronic HBV infection, as measured by the prevalence of HBs antigenemia, exist in the world: high (>8%), intermediate (2–8%), and low (< 2%) [Kew, 1992].

HBV is characterised by high genetic variability. Indeed, during viral replication, the use of reverse transcriptase, characteristic of hepadnavirus, is responsible of this phenomenon as this polymerase lacks of reading proof activity. Therefore, ten genotypes were described and named A to J. Genotypes are segregated into subgenotypes which have distinct ethno-geographic distribution [Zehender et al., 2014]. Recent study have hypothesised that HBV has been coexpanding and co-migrating with human population, then this distribution of subgenotypes might be the result of the main modern human migration [Paraskevis et al., 2013].

Most of African countries have a high level of endemicity (chronical infection >8%). In these countries transmission occurs mainly during perinatal and childhood periods [Hou et al., 2005]. In East Africa, genotype A is the most prevalent genotype, while in other parts of Africa, genotype E is mostly observed [Croagh et al., 2015]. Some studies have suggested a relatively recent introduction of genotype E into the general African population [Andernach et al., 2009; Forbi et al., 2010].

In Madagascar, former studies on epidemiology of HBV classified the country as part of high endemic area as HBV prevalence can reach 23% in general population [Migliani et al., 2000].

1  
2  
3 However, this prevalence differs largely between urban and rural areas and is estimated to be  
4  
5 respectively 5% and 26% [Boisier et al., 1996; Arivelo et al., 2011; Randriamahazo et al.,  
6  
7 2015]. Like other African countries, transmission occurs mainly during perinatal and  
8  
9 childhood periods [Migliani et al., 2000]. Few studies report genotype diversity of HBV in  
10  
11 Madagascar. Genotype E is the most prevalent according to Dupinay et al. [Dupinay et al.,  
12  
13 2010], followed by genotype A1 and D. Phylogenetic analyses of the genotype E strains  
14  
15 showed a relationship with West African strains [Kramvis et al., 2005]. However, these  
16  
17 studies were performed using limited samples from northern and north-western parts of  
18  
19 Madagascar.  
20  
21

22  
23 The aims of the present study were (i) to describe the genetic diversity of HBV strains from  
24  
25 different regions of Madagascar, and (ii) to describe the viral gene flow throughout the  
26  
27 country.  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**MATERIALS AND METHODS**

**Human samples**

A national cross-sectional serological survey from November 2011 to April 2012 and from October 2012 to May 2013 was conducted. A total of 28 different zones (urban or rural) distributed over 34 administrative Malagasy districts were included (Fig. 1A). Individuals were randomly selected and a total of 100 individuals were finally included in the study. After reading of the informed consent letter, written and oral consent was obtained from volunteering individuals. The study protocol, including informed consent procedure, was approved by the Malagasy competent authorities: the Malagasy Ethic National Committee (authorization N° 066/MSAMP/CE, 26th July 2011). After consent, volunteers were blood sampled and sera were stored in Nitrogen liquid in the field, then at -80°C at the Virology Unit at the Institute Pasteur of Madagascar until use. The age of the volunteers ranged from 18 to 99 years with a median age of 35 (mean age of 37.6) with a sex-ratio of 0.97. Of the 1,778 sera tested, 141 sera (7.9% and IC95% [6.7-9.3-]) were positive for HBs antigen using in-house ELISA.

**HBV DNA Extraction and Amplification by Means of Nested PCR**

After a first serological screening searching HBs antigen, total DNA was extracted from positive sera using the NucleoSpin® Dx Virus (Macherey-Nagel), according to the manufacturer's instructions. Sequences of about 1,348 nucleotides were amplified using a previously described multiple nested PCR protocol in order to obtain the HBV polymerase (P) and surface antigen (PreS/S) sequences [Zehender et al., 2012]. Amplicons were then purified using QIAquick PCR Purification Kit® (QIAGEN) before performing DNA sequencing.

### HBV DNA Sequencing

The three DNA fragments were sequenced bidirectionally using a BigDye Terminator Kit version 3.1 (Applied Biosystems), according to the manufacture's instruction. The sequencing products were purified by precipitating 10  $\mu$ l in an ethanol/sodium acetate mixture. Finally, sequences were loaded into an automated DNA sequencer ABI PRISM\_ 3130 XL Genetic Analyser (Applied Biosystem). The sequence assembly was performed with Sequencer 4.1.4 program (Sequencher® sequence analysis software, Gene Codes Corporation, Ann Arbor, MI USA <http://www.genecodes.com>). The sequences will be submitted to GenBank after the acceptance of the manuscript.

### Phylogenetic Analysis

Each of the 100 HBV P-PreS/S genes, strains from Madagascar, were submitted to NCBI Genotyping Tool (<http://www.ncbi.nlm.nih.gov/projects/genotyping/formpage.cgi>) then to jumping profile Hidden Markov Model (jpHMM) ([http://jphmm.gobics.de/submission\\_hbv](http://jphmm.gobics.de/submission_hbv)) in order to quickly identify the genotype of the sequences. This preliminary analysis has allowed to identify 97 “pure” genotype isolates and three hypothetical recombinants.

For the purpose of phylogenetic analysis three data sets were built. The first data set included 187 HBV P-PreS/S genotype/subgenotype-specific reference sequences, downloaded from GenBank, and 97 HBV P-PreS/S strains from Madagascar, previously reported as “pure” genotype in both NCBI and jpHMM genotyping analysis. This data set was used to confirm the previous genotyping. The second data set included 79 HBV E genotype P-PreS/S strains and was used for gene flow analysis. The third data set included 187 HBV P-PreS/S genotype/subgenotype reference sequences and the three hypothetical recombinant sequences and was used to test the nature of recombinants in Malagasy strains. The sequences of the

three dataset were aligned by using Clustal W included in Bioedit [Hall, 1999; Ciccozzi et al., 2014], followed by manual editing.

The hierarchical likelihood ratio test (LRT) method implemented in Model-Test version 3.07, as already described [Posada and Crandall, 1998; Ciccozzi et al., 2013] was used to select the evolutionary model that best fitted the sequences data. The GTR+I+G model was selected for all the data set. The phylogenetic tree of the first dataset was reconstructed by using Maximum Likelihood analysis with FastTree 2.1 with the evolutionary model previously selected by ModelTest [Price et al., 2010]. The statistical robustness and reliability of the branching order within the phylogenetic tree was confirmed with the support values analysis (SH-like local support values >0.70). A Bayesian phylogenetic tree was reconstructed by means of Mr Bayes on the second data set [Huelsenbeck and Ronquist, 2001]. A Markov Chain Monte Carlo search was made for  $15 \times 10^6$  generations using tree sampling every 100<sup>th</sup> generation and a burn-in fraction of 25%. Statistical support for specific clades was obtained by calculating the posterior probability of each monophyletic clade, and a posterior consensus tree was generated after a 25 % burn-in. This tree was used as starting tree to perform the gene flow analysis. The possible presence of recombination among HBV strains was tested by using the Simplot (<http://sray.med.som.jhmi.edu/SCSoftware/simplot/>) and SplitsTree programs [Huson and Bryant, 2006].

**Migration analysis (gene flow)**

The Mac Clade version 4 program (Sinauer Associates, Sunderland, MA) was used on the second data set to test viral gene in/out flow among distinct HBV subpopulations within different area in Madagascar, using a modified version of the Slatkin and Maddison test as described below [Slatkin and Maddison, 1989; Lo Presti et al., 2013]. Two gene flow analyses were performed. The first one was conducted according to three groups defined as zone 1, 2



and 3 (nearly urban, suburban and rural) (Fig. 1B). Firstly, Zone 1 (urban,  $n= 8$ ) included the administrative communes of the former capitals of provinces, and administrative communes located at the edge of national routes which connect these main provinces. Zone 2 (suburban,  $n= 30$ ) included the communes next to the districts of the capitals of provinces and the communes of the administrative-centre of each districts. Zone 3 (rural,  $n= 40$ ) included the communes out of districts of capitals of provinces and out of administrative-centre of each district. The sequences were included in each zone according to their sampling location. The second gene flow analysis was conducted classifying the HBV genotype E strains into three groups (north,  $n= 31$ ; middle,  $n= 22$ ; south,  $n= 27$ ) based on the sampling location of the strains in Madagascar (Fig. 1A).

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), which counts the number of changes in a tree for each pair-wise character state. When multiple MPRs were present (as in our data set), the algorithm calculated the average migration count over all possible MPRs for each pair. The resulting pair-wise migration matrix was then normalized, and a randomization test with 10,000 matrices obtained from 10,000 random trees (by random joining-splitting of the original tree) was performed to assess the statistical significance of the observed migration counts.

**RESULTS**

**Genotype/subgenotype assignation**

The genotyping resulted by NCBI Genotyping Tool and jpHMM found 97 pure subtypes and three possible recombinant form.

Maximum likelihood phylogenetic tree (Fig. 2), showed different statistically supported clades, corresponding to different HBV genotype/subgenotypes. The ML tree showed that three Malagasy strains were classified as subgenotype D2 (3%), twelve strains as subgenotype D7 (12%), seventy-nine strains as subgenotype E (79 %) and three Malagasy strains as subgenotype A1 (3%). Simplot and Splits Recombination analysis found three “putative” recombinant strains (3%) (Fig. 3 and 4).

SplitsTree analysis showed that three sequences were unassigned, as reported in Figure 3. Bootscan analyses (Fig. 4, part A, B and C) classified the three strains labelled #17, #46 and #104 as recombinant D7/E, with a unique pattern of recombination and a single breakpoint.

**Migration fluxes**

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 ( $p<0.0001$ ).

Regarding the gene flow analysis performed by using the division in zone 1, 2 and 3 (urban, suburban and rural), most of the observed HBV-E genotype gene flow (80 %) was from zone 3 (rural) to zone 2 (suburban) (Fig. 5 a).

The second gene flow analysis classifying the HBV-E detected according to their origin (North, middle and south), showed a 57.1 % of gene flow from center to north and a 42.9 % of gene flow from center to south was also found (Fig. 5 b).

## DISCUSSION

HBV is characterized by a high degree of genetic heterogeneity and a worldwide distribution. Madagascar is an endemic area for HBV and previous studies indicated that genotype E is the most prevalent [Dupinay et al., 2010], followed by genotype A1 and D. In this study, the authors showed circulation of different HBV strains, some of them genetically related to strains isolated in other West African countries (the most divergent genotype E strain was derived from Madagascar). Nevertheless, in this study, only few sequences were obtained from blood collected from few villages and cities located in the northern and northwestern part of the country. Thus, those observations might not reflect the pattern of HBV genotypic/sub-genotypic circulation in the entire country.

The present study is the first large-scale molecular and phylogenetic study analyzing HBV sequences from 28 different Malagasy areas never sampled in the past. In addition, it represents an updated overview of the strains circulating in Madagascar. In this study, the most prevalent genotype/sub-genotypes was E in agreement with the previous articles [Dupinay et al., 2010]. Moreover most genotype D Malagasy strains were sub-genotype D7, as already found by Dupinay et al., 2010. Interestingly three D7/E recombinant forms has been described for the first time in Madagascar assuming a possible co-circulation of genotype E, the most prevalent genotype and genotype D7. However, as emergence of genotype E is relatively recent [Andernach et al., 2009], suggesting sexual intercourse as mode of transmission of this HBV genotype. Temporal and phylogeographic study of viruses will confirm or not this hypothesis.

Migration analysis showed a gene flow from zone 3 (rural) to zone 2 (suburban) and a greater gene flow from the middle part of Madagascar to the north than to the south. Seroprevalence studies showed firstly a of higher prevalence hepatitis B surface antigen (HBsAg) in rural area compared to urban area [Boisier et al., 1996]. It suggests a genomic flow from the more prevalent areas to the less prevalent ones reflecting the movement of the virus from chronic carriers in the general population. The largest gene flow from the highlands to the northern of the island may be explained: (i) by the invasion of this part during the Malagasy kingdom history, (ii) moreover, geographic accessibility of northern part of the island is easier than to the southern part, as the mountain chain at the south-west can play a key role as a natural barrier, (iii) the south part of the island is naturally dry with regular food shortages, therefore less attractive for people migration. Moreover the north of Madagascar (e.g. Nosy Be Island) is characteristic and visited by a lot of tourist during the summer season and a sentinel surveillance program can be useful to monitor the HBV flow in this area. A surveillance program could be important also because since 2004, the majority of European Union member states have introduced the vaccine, either as universal infant, universal newborn or universal adolescent. However, a number of countries from northern Europe had not introduced the vaccine into the routine program [Mereckiene et al., 2010]. At the end, a better characterization and monitoring of HBV infected individuals in Madagascar are crucial in order to understand the epidemiology and to assess the efficacy of prevention and therapy in controlling the epidemic. Hepatitis B vaccine is available and is part of recommended vaccine by Ministry of Health since 2002; nevertheless to date no study has evaluated the coverage of this vaccine amongst the community. Probably screening individuals at risk of acquiring hepatitis B would help in limiting the spread and public health repercussions of the virus in this country. Continuous suppression of HBV replication in individuals with advanced liver disease prolongs life, decreases the need for liver transplantation, and potentially reduces the

1  
2  
3 risk for hepatocarcinoma. To the light of these considerations, it is important to study the  
4  
5 HBV infections in Madagascar and to monitor the potential spread of this viral strain inside  
6  
7 this country.  
8  
9

## 10 11 12 **ACKNOWLEDGEMENTS**

13  
14 We would like to thank the different staff involved in the blood collection including the  
15  
16 leaders of the villages (chefs fokontany), the community health workers, our guides and the  
17  
18 health district officers.  
19

## 20 21 22 **FINANCIAL SUPPORTS**

23  
24 The human specimens (serums) were collected as part of an integrative global project leads by  
25  
26 the Institut Pasteur of Madagascar (the ZORA project) funded by different partners according  
27  
28 thematics; the US Centers for Disease Control and Prevention (Cooperative Agreement  
29  
30 Number, 5U51IP000812-02), the Department of Homeland Security Center of Excellence for  
31  
32 Emerging and Zoonotic Animal Diseases (CEEZAD) (Grant No. 2010-ST061-AG0001). The  
33  
34 contents of the manuscript are solely the responsibility of the authors and do not necessarily  
35  
36 represent the official views of the Centers for Disease Control and Prevention, the Department  
37  
38 of Health and Human Services or the Department of Homeland Security. The authors have no  
39  
40 conflicts of interest to declare.  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

REFERENCES

Andernach IE, Nolte C, Pape JW, Muller CP. 2009. Slave trade and hepatitis B virus genotypes and subgenotypes in Haiti and Africa. *Emerg Infect Dis* 15(8): p. 1222-8.

Arivelo RZ, Hendrison RD, Elie RF, Tantely RM, Ramamonjisoa A, Barnia RF, Paquerette HS, Raft HF, Aimée RA, Andry R. 2011. The seroprevalence of hepatitis B surface antigen among first time blood donors in Antananarivo (Madagascar) from 2003 to 2009. *Blood Transfus* 9(4): p. 475-7.

Boisier P, Rabarijaona L, Piollet M, Roux JF, Zeller HG. 1996. Hepatitis B virus infection in general population in Madagascar: evidence for different epidemiological patterns in urban and in rural areas. *Epidemiol Infect* 117(1): p. 133-7.

Croagh C.M., Desmond PV, Bell SJ. 2015. Genotypes and viral variants in chronic hepatitis B: A review of epidemiology and clinical relevance. *World J Hepatol* 7(3): p. 289-303.

Ciccozzi M, Ciccaglione AR, Lo Presti A, Equestre M, Cella E, Ebranati E, Gabanelli E, Villano U, Bruni R, Yalcinkaya T, Tanzi E, Zehender G. 2014. Evolutionary dynamics of HBV-D1 genotype epidemic in Turkey. *J Med Virol* 86(1): p. 109-16.

Ciccozzi M, Babakir-Mina M, Lo Presti A, Salpini R, Cella E, Gabanelli E, Teoharov P, Kevorkyan A, Perno CF, Zehender G, Ciotti M. 2013. Molecular analysis of hepatitis B virus in Bulgaria. *J Med Virol* 85(1): p. 49-54.

Dupinay T, Restorp K, Leutscher P, Rousset D, Chemin I, Migliani R, Magnius L, Norder H. 2010. High prevalence of hepatitis B virus genotype E in Northern Madagascar indicates a West-African lineage. *J Med Virol* 82(9): p. 1515-26.

Forbi JC, Vaughan G, Purdy MA, Campo DS, Xia GL, Ganova-Raeva LM, Ramachandran S, Thai H, Khudyakov YE. 2010. Epidemic history and evolutionary dynamics of hepatitis B virus infection in two remote communities in rural Nigeria. *PLoS One* 5(7): p. e11615.

Hall A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucleic Acids Symposium Series* 41: p. 95-98.

Hou J, Liu Z, Gu F. 2005. Epidemiology and Prevention of Hepatitis B Virus Infection. *Int J Med Sci* 2(1): p. 50-57.

Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17(8): p. 754-5.

- Huson D.H., Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* 23(2): p. 254-67.
- Kew MC. 1992. Tumours of the liver. *Scand J Gastroenterol Suppl* 192: p. 39-42.
- Kramvis A, Restorp K, Norder H, Botha JF, Magnius LO, Kew MC. 2005. Full genome analysis of hepatitis B virus genotype E strains from South-Western Africa and Madagascar reveals low genetic variability. *J Med Virol* 77(1): p. 47-52.
- Lavanchy D. 2005. Worldwide epidemiology of HBV infection, disease burden, and vaccine prevention. *J Clin Virol* 34 Suppl 1: p. S1-3.
- Lo Presti A, Ciccozzi M, Cella E, Giovanetti M, Zehender G, Valenti D, Iannotti N, Malfatto E, Bernardini C, Maggiolo F, Callegaro A. 2013. Migration patterns of HIV-1 subtype B virus in Northern Italy. *New Microbiol* 36(1): p. 75-9.
- Mereckiene J, Cotter S, Lopalco P, D'Ancona F, Levy-Bruhl D, Giambi C, Johansen K, Dematte L, Salmaso S, Stefanoff P, O'Flanagan D. 2010. Hepatitis B immunisation programmes in European Union, Norway and Iceland: where we were in 2009? *Vaccine* 28(28): p. 4470-7.
- Migliani R, Rousset D, Rakoto-Andrianarivelo M, Rabarijaona L, Ratsitorahina M, Rajaonarivony V, Maucière P. 2000. [Hepatitis B virus infection: a public health problem in Madagascar]. *Arch Inst Pasteur Madagascar* 66(1-2): p. 50-4.
- Ott JJ, Stevens GA, Wiersma ST. 2012. The risk of perinatal hepatitis B virus transmission: hepatitis B e antigen (HBeAg) prevalence estimates for all world regions. *BMC Infect Dis*, 2012. 12: p. 131.
- Paraskevis D, Magiorkinis G, Magiorkinis E, Ho SY, Belshaw R, Allain JP, Hatzakis A. 2013. Dating the origin and dispersal of hepatitis B virus infection in humans and primates. *Hepatology* 57(3): p. 908-16.
- Posada D., Crandall KA. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14(9): p. 817-8.
- Price MN, Dehal PS, Arkin AP. 2010. FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* 5(3): p. e9490.
- Randriamahazo TR, Raheinaivo AA, Rakotoarivelo ZH, Contamin B, Rakoto Alson OA, Andrianapalaninarivo HR, Rasamindrakotroka A. 2015. Prevalence of hepatitis B virus serologic markers in pregnant patients in Antananarivo, Madagascar. *Med Mal Infect* 45(1-2): p. 17-20.

Slatkin M., Maddison WP. 1989. A cladistic measure of gene flow inferred from the phylogenies of alleles. *Genetics* 123(3): p. 603-13.

Zehender G, Shkjezi R, Ebranati E, Gabanelli E, Abazaj Z, Tanzi E, Kraja D, Bino S, Ciccozzi M, Galli M. 2012. Reconstruction of the epidemic history of hepatitis B virus genotype D in Albania. *Infect Genet Evol* 12(2): p. 291-8.

Zehender G, Ebranati E, Gabanelli E, Sorrentino C, Lo Presti A, Tanzi E, Ciccozzi M, Galli M. 2014. Enigmatic origin of hepatitis B virus: an ancient travelling companion or a recent encounter? *World J Gastroenterol* 20(24): p. 7622-34.

For Peer Review



## Figure Legends

**Fig. 1A.** Map of Madagascar describing the different zone defined as zone 1 (in grey) representing urban area, zone 2 (in blue) representing suburban area and zone 3 (in white) that includes all districts outside of the two other zones. Zone1 include the administrative communes of the former capital of provinces, and administrative communes located at the edge of national routes which connect these main provinces. Zone 2 include the communes next to urban zones and the communes of the administrative-centre of each districts of Madagascar. Zone 3 include the communes out of district of capital of provinces and out of administrative-centre of each district. **B.** Madagascar map describing the geographic location of sampling sites divided in north, middle and south.

**Fig. 2.** Maximum Likelihood analysis of the first dataset. The tree was rooted by using the midpoint method. Branch lengths were estimated with the best fitting nucleotide substitution model according to a hierarchical likelihood ratio test, and were drawn to scale with the bar at the bottom indicating 0.03 nucleotide substitutions per site. One (\*) along a branch represents significant statistical support for the clade subtending that branch (SH-like local support values >0.70). The main supported clades are highlighted and reported in different colors.

**Fig. 3.** Network computed with Splits Tree by using the Neighbor-net method. The network of the three Malagasy unassigned sequences is shown. The scale bar indicates 1% of nucleotide divergence. The three Malagasy isolates are highlighted with arrows.

**Fig. 4.** Bootscanning plots implemented in Simplot software for the three Malagasy isolates labelled #17, #46 and #104 (Panel A, B and C).

**Fig. 5 a.** Maximum parsimony migration patterns of HBV E genotype Malagasy isolates to/from different zone (1, 2 and 3: urban, suburban and rural) of Madagascar. The bubblegram shows the frequency of gene flow (migrations) to/from different zone, as the percentage of the total observed migrations estimated from the tree with a modified version of the Slatkin and Maddison test. The surface of each circle is proportional to the percentage of observed migrations given within the circle. **b.** Maximum parsimony migration patterns of HBV E genotype Malagasy isolates to/from different groups/areas (north, center, south) of Madagascar. The bubblegram shows the frequency of gene flow (migrations) to/from different areas, as the percentage of the total observed migrations estimated from the tree with a modified version of the Slatkin and Maddison test. The surface of each circle is proportional to the percentage of observed migrations given within the circle.

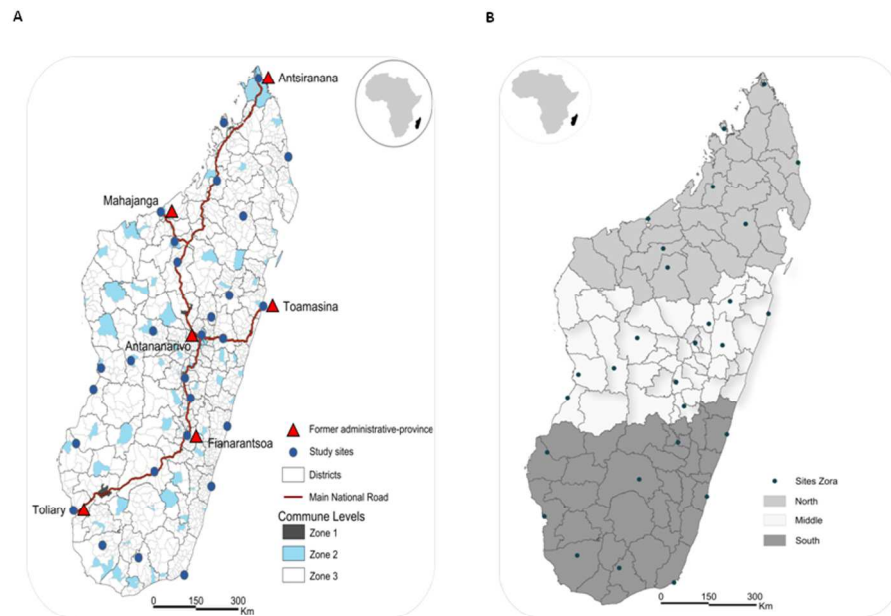


Fig. 1A. Map of Madagascar describing the different zone defined as zone 1 (in grey) representing urban area, zone 2 (in blue) representing suburban area and zone 3 (in white) that includes all districts outside of the two other zones. Zone1 include the administrative communes of the former capital of provinces, and administrative communes located at the edge of national routes which connect these main provinces. Zone 2 include the communes next to urban zones and the communes of the administrative-centre of each districts of Madagascar. Zone 3 include the communes out of district of capital of provinces and out of administrative-centre of each district. B. Madagascar map describing the geographic location of sampling sites divided in north, middle and south.

161x120mm (300 x 300 DPI)

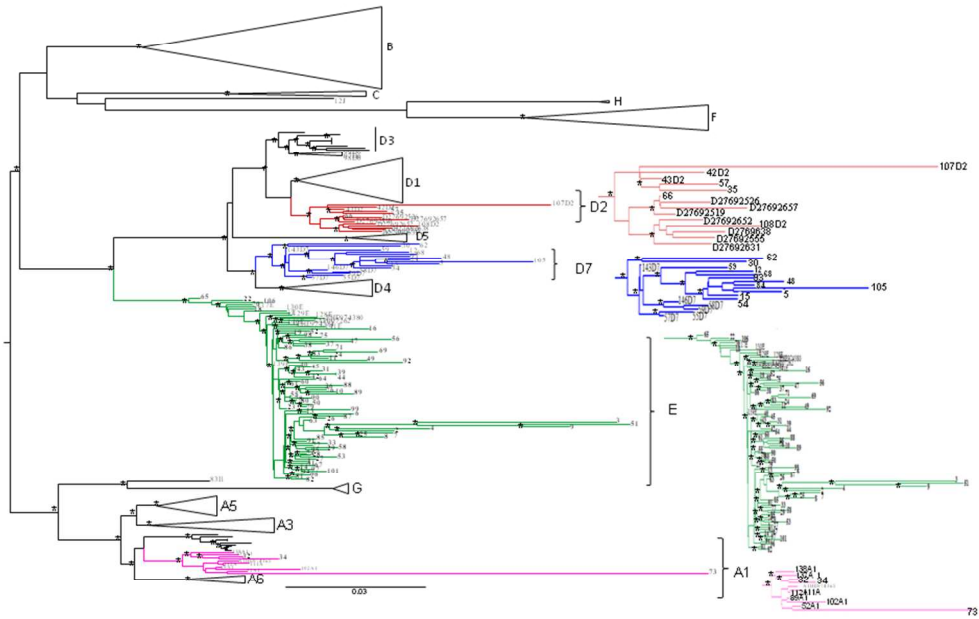


Fig. 2. Maximum Likelihood analysis of the first dataset. The tree was rooted by using the midpoint method. Branch lengths were estimated with the best fitting nucleotide substitution model according to a hierarchical likelihood ratio test, and were drawn to scale with the bar at the bottom indicating 0.03 nucleotide substitutions per site. One (\*) along a branch represents significant statistical support for the clade subtending that branch (SH-like local support values >0.70). The main supported clades are highlighted and reported in different colors.  
81x60mm (300 x 300 DPI)

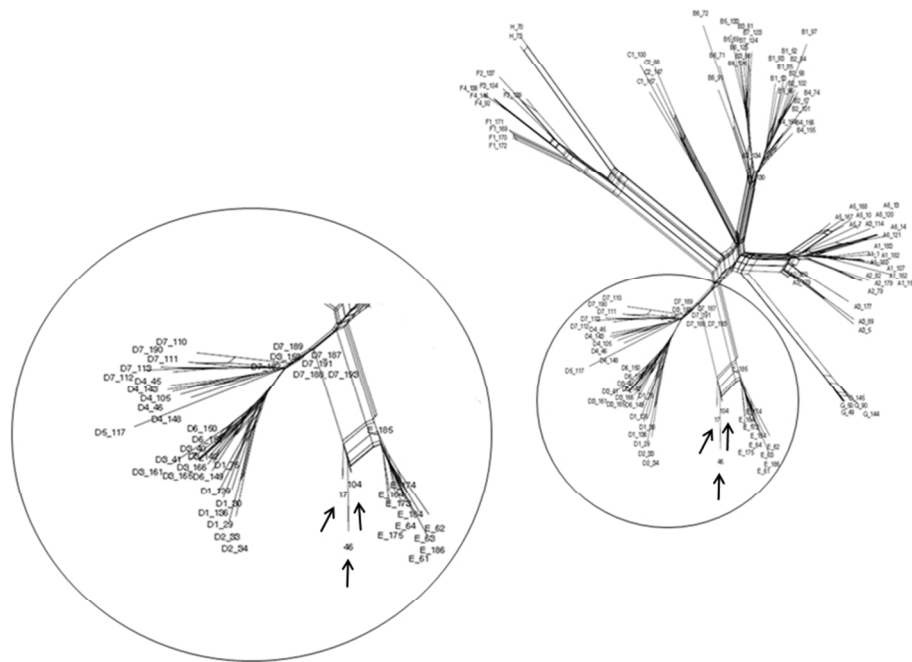


Fig. 3. Network computed with Splits Tree by using the Neighbor-net method. The network of the three Malagasy unassigned sequences is shown. The scale bar indicates 1% of nucleotide divergence. The three Malagasy isolates are highlighted with arrows.

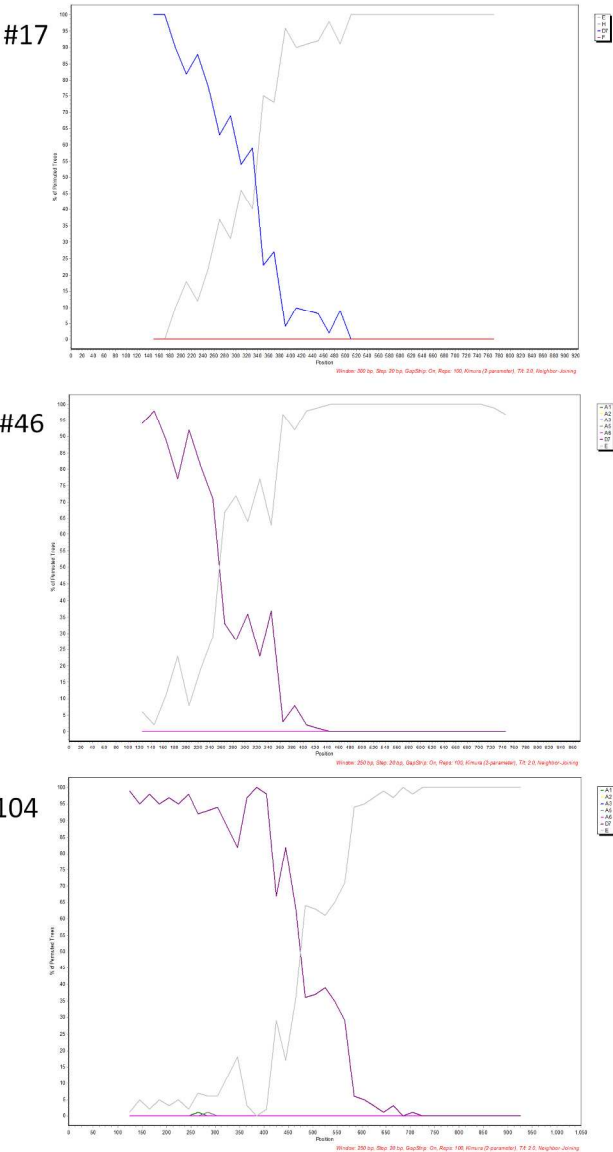


Fig. 4. Bootscanning plots implemented in Simplot software for the three Malagasy isolates labelled #17, #46 and #104 (Panel A, B and C).  
195x260mm (300 x 300 DPI)

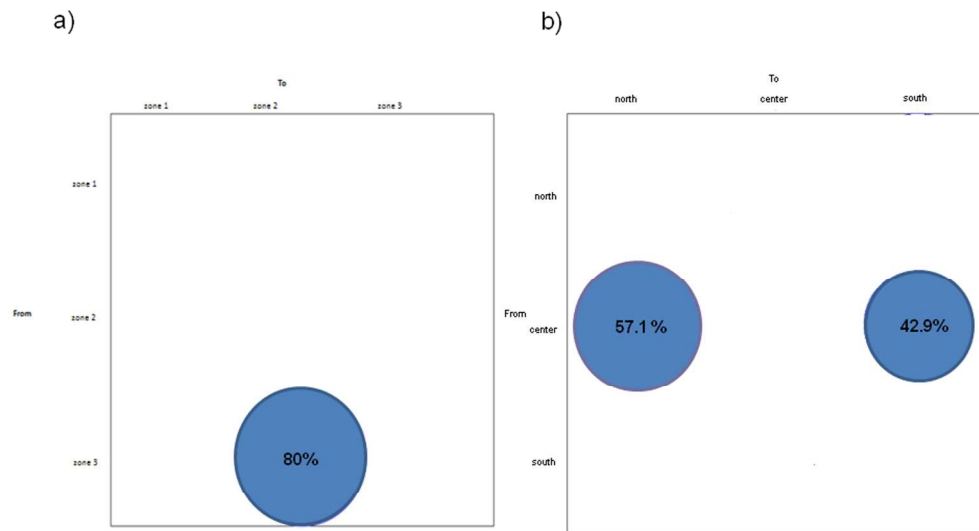


Fig. 5 a. Maximum parsimony migration patterns of HBV E genotype Malagasy isolates to/from different zone (1, 2 and 3: urban, suburban and rural) of Madagascar. The bubblegram shows the frequency of gene flow (migrations) to/from different zone, as the percentage of the total observed migrations estimated from the tree with a modified version of the Slatkin and Maddison test. The surface of each circle is proportional to the percentage of observed migrations given within the circle. b. Maximum parsimony migration patterns of HBV E genotype Malagasy isolates to/from different groups/areas (north, center, south) of Madagascar. The bubblegram shows the frequency of gene flow (migrations) to/from different areas, as the percentage of the total observed migrations estimated from the tree with a modified version of the Slatkin and Maddison test. The surface of each circle is proportional to the percentage of observed migrations given within the circle.

161x120mm (300 x 300 DPI)