Molecular Evolution of Hepatitis B Virus Genotype F in Latin America

Jonas Wolf （✉ jonasmwolf@gmail.com ）

  Lutheran University: Universidade Luterana do Brasil   https://orcid.org/0000-0001-7577-464X

Thiago Kastell Mazeto

  Universidade Luterana do Brasil

Vagner Reinaldo Zingalli Bueno Pereira

  Universidade Luterana do Brasil

Daniel Simon

  Universidade Luterana do Brasil

Vagner Ricardo Lunge

  Universidade Luterana do Brasil

Research Article

Abstract

Hepatitis B virus (HBV) genotype F evolution is not completely understood in Latin America. This study aims to evaluate the molecular evolution of HBV-F in Latin America by comparing 224 whole-genome sequences. Bayesian coalescent analysis was performed to estimate the time to the most recent common ancestor. Four main clades were formed dated back between 1245 and 1730. Also, four subclades were identified dated back between 1705 and 1801. HBV-F overall effective population size grew in the 18th century and showed an initial circulation of HBV-F from Venezuela to other countries from Latin America.

Background

Despite the accessibility of an effective immunization, more than 350 million individuals around the world are chronically infected with the hepatitis B virus (HBV), and numerous individuals have developed severe liver diseases, such as cirrhosis and hepatocellular carcinoma (HCC) [1]. HBV includes a partially double-stranded DNA genome composed of approximately 3.2 kb, with four overlapping open reading frames (ORF): preS/S, pre-core/core, X, and polymerase [2]. As a result of its unusual mechanism of replication by reverse transcription and the lack of proofreading activity of the reverse transcriptase, HBV shows exceedingly heterogeneous sequence diversity [3, 4].

Hepatitis B infection may be an especially ancient disease, with scientific evidence of its occurrence dating back to around 40,000 years ago. HBV presents an evolutionary history related to the human species since its scattering around the world was strongly associated with the history of migrations of the different populations [5–8]. HBV is also long-established in all America since this virus was detected in different Amerindian populations [9, 10–12].

Ten HBV genotypes (A to J) have been already described in the World.13 HBV-A is frequently detected in Africa, Europe, and Latin America [10, 11, 13]. HBV-B and HBV-C are common in Asia, while HBV-D is prevalent in the Mediterranean region and Americas. HBV-E, HBV-G, HBV-H, HBV-I, and HBV-J have a more restricted dissemination worldwide. Interestingly, HBV-F is more frequent in the Americas, from Alaska to Argentina [11, 13–15]. HBV-F seems to be originated in Native American populations. One HBV strain isolated from a wooly monkey (a New World monkey) presented a close similarity to this genotype, suggesting a possible spillover between species [15, 16].

HBV-F strains have a significant divergence when comparing strains from different countries in America, particularly among the Amerindian population. Four subgenotypes were detected until now with genetic divergence among 4.3–6.1% [10, 15, 16]. HBV-F evolutionary history is not completely understood [10, 16]. Therefore, the present study aimed to revise the molecular evolution of HBV-F in Latin America by comparing whole-genome sequences.

Non-clones and non-recombinant complete genome sequences of HBV-F in which the country and year of sampling were available were obtained from GenBank (www.ncbi.nlm.nih.gov/genbank/) (**Table S1**).

Putative recombination events were verified using the Recombination Detection Program version 4 (RDP4) software [17] with the default settings using the algorithms RDP, GENECONV, BootScan, MaxChi, Chimaera, SiScan, 3Seq, and LARD. The beginning and end breakpoints of the potential recombinant sequences were also defined by the RDP4 software. Putative recombinant events were considered significant when $p \leq 0.01$ was observed for the same event using four or more algorithms. Sequences that presented putative recombination events were excluded from the evolutionary analysis. Additionally, sequences with a common origin (i.e.: nosocomial outbreaks) were discarded. All HBV-F sequences were aligned separately by MAFFT v7 [18] and visually inspected with AliView v1.26.

Time-scaled phylogenetic tree estimation was performed using BEAST/BEAGLE v 2.5 software [19]. The best-fitting nucleotide substitution (GTR) model was selected using a hierarchical likelihood ratio, Akaike information criterion, and Bayesian information criterion tests with Model Finder in IQ-TRE web server (http://iqtree.cibiv.univie.ac.at/). For each run of 500 million of Monte Carlo Markov Chains (MCMC), the marginal likelihood was estimated via path sampling (PS) and stepping stone (SS) methods and the resulting Bayes Factors (BF) (ratio of marginal likelihoods) used to select the best-fitting clock/demographic model. The models can be compared to evaluate the strength of evidence against the null hypothesis (H0) defined in the following way: 2lnBF < 2 indicates no evidence against H0; 2–6, weak evidence; 6–10: strong evidence, and > 10 very strong evidence [20]. Both SS and PS estimators indicated the uncorrelated lognormal molecular clock (Bayes Factor = 15.8) as the best-fitted model to the datasets under analysis. MCMC analysis was performed and the MLE of the obtained trees were compared using a BF to select the best model and parameter values. BF analysis showed that the uncorrelated lognormal clock fitted the data significantly better than other clocks (2lnBF = 701.17). BF analysis showed that the Bayesian skyline plot was better than other models (2lnBF > 100). MCMC was run for 500 million generations to ensure stationary and adequate effective sample size (ESS) for all statistical parameters. Tracer v.1.6 software [21] was used to diagnose MCMC, adjust initial burn-in, and to perform the Skyline demographic reconstruction. Uncertainty in parameter estimates was evaluated in the 95% highest posterior density (HPD 95%) interval. TreeAnnotator v1.8.2 was used to summarize the maximum clade credibility (MCC) tree from the posterior distribution of trees and the MCC tree was visualized and edited in FigTree v.1.4.4 (available at http://tree.bio.ed.ac.uk/software/figtree/).

Phylogeographic analysis, incorporating both spatial and temporal information, was performed with BEAST v2.6.219 using a discrete trait, symmetric substitution model with Bayesian stochastic search variable selection (BSSVS). The reversible discrete Bayesian phylogeographic model with a continuous-time Markov chain (CTMC) rate reference prior was performed [22]. The number of viral migrations between locations was estimated using 'Markov Jump' counts of location-state transitions along with the posterior tree distribution. Migratory events across time were summarized using the SPREAD v.1.0.723. BF > 3 were considered as well supported diffusion rates constituting the migration graph [22].

The population dynamics and the root tMRCA for the HBV-F were analyzed in a Bayesian coalescent framework. HBV-F root tMRCA dated back to 1040 (HPD 95%: 643–1090). The substitution rate was 2.84E10-6 (HPD 95%: 1.34E10-7 – 1.36E10-6) substitutions per site per year (s/s/y). Four main clades

were formed, the clade F1 dated back to 1650 (HPD 95%: 1402–1758), F2 to 1245 (HPD 95%: 1150–1469), F3 to 1730 (HPD 95%: 1588–1854), and F4 to 1381 (HPD 95%: 1214–1458). Additionally, four subclades were identified: F1a dated back to 1714 (HPD 95%: 1515–1845), F1b to 1705 (HPD 95%: 1508–1856), F2a to 1728 (HPD 95%: 1450–1851), and F2b to 1801 (HPD 95%: 1627–1895) (Fig. 1).

The subgenotype F1a was disseminated in Central America (Costa Rica, El Salvador, and Nicaragua), while F1b between South American countries (Argentina, Brazil, Chile, Peru, and Uruguay) and North America (USA/Alaska and Mexico). The clade F4 disseminated from Argentina to Peru, Bolivia, and Brazil. The clade F3 disseminated from Venezuela to Colombia and Argentina. Finally, the clade F2 disseminated from Venezuela (F2b) to Nicaragua (single event), Brazil, and Argentina (F2a) (Fig. 1). Phylodynamic analysis of full-length sequences of HBV-F by using the relaxed Skyline model showed that the overall effective population size grew in the 18th century (Fig. 2).

Phylogeographic results suggest a circulation of HBV-F among South America (Argentina, Bolivia, Brazil, Chile, Colombia, Peru, Uruguay, and Venezuela), and Central America (Costa Rica, El Salvador, and Nicaragua). HBV-F seems to be disseminated first from Venezuela to Nicaragua (BF = 10.1), Brazil (BF = 20.1), Bolivia (BF = 22.3), Argentina (BF = 22.9), Colombia (BF = 19.8), El Salvador (BF = 9.8), Costa Rica (BF = 8.8), Peru (BF = 18.8), Chile (BF = 15.8), Uruguay (BF = 14.2), and Mexico (BF = 10.3) (Fig. 3).

Previous phylogenetic analyses have separated HBV-F into the four main clades: F1 (F1a and F1b), F2 (F2a and F2b), F3, and F4 [24, 25], but the origin of the geographic dissemination of HBV-F is still uncertain. Regardless, the determination of the common ancestors of each subgenotype is conceivable, since distinctive strains from several countries have been sequenced in previous studies, predominantly of Central and South America (where they are more prevalent) [14, 26]. Therefore, the present study aimed to revise the molecular evolution of HBV-F in Latin America.

In Venezuela, the subgenotypes F1, F2 (F2a and F2b), and F3 circulate among Amerindian tribes found in the East and West regions [24]. The HBV-F1a subgenotype is predominant in Central America, having been observed in Costa Rica and El Salvador [26], while HBV-F1b is predominant in South America mainly in the southern cone (Argentina, Chile, and Uruguay) [27, 28]. The HBV-F1b was also found in Colombia14 and it was detected among Alaskan native individuals with chronic HBV infection [29]. HBV-F3 has been observed in Central America and the northern region of South America [15, 24]. In Colombia, HBV-F3 was previously described in Bogota and Bucaramanga, and the Yucpa population of Venezuela [24]. In Brazil, in a nationwide study, HBV-F was detected in 11.3% of hepatitis B cases. The presence of this genotype varied among regions of the country: 23.5% in the Northeast, 10.9% in the North, 10.3% in the Central-West, 7.4% in the Southeast, and 1.0% in the South [30].

Bayesian analysis makes it possible to demonstrate inferences about the time of the common ancestor of HBV-F subgenotypes. In the present study, in the coalescent Bayesian analysis, four main clades were formed: the clade F1, F1a, F1b, F2a, and F2b. Also, four subclades were detected (F1, F1a, F1b, F2a, and F2b) with dissemination mainly in the 18th century. The root of HBV-F tMRCA dated back to the 11th century and the Skyline plot shown viral dissemination mainly in the 18th century. Our results, following a

previous study in Brazil, observed through Bayesian analysis that HBV-F was possibly disseminated throughout Latin America centuries ago, thus demonstrating a long evolutionary history [31]. In this sense, Monjsiejczuk et al. (2019) [10] reported that HBV-F4 could originate from Paraguay and spread in the early 18th century. However, the spread of HBV-F in Argentina was demonstrated to start more recently, in the late 20th century, in the province of Missiones and Buenos Aires [28, 32].

The phylodynamic analysis also showed that HBV-F infections increased in the 18th century. This genotype has been detected mainly in Native American groups [15, 33] and was possibly originated in Latin America in pre-Columbian times (between 14.5–16 thousand years ago) [34]. Notwithstanding, the entry of European conquers had a critical effect on local demography, as well as on the epidemiology of infectious diseases. During this period, the local population fell by 80%, which may have caused the extinction of some strains of HBV-F that were already endemic on the continent. Political and armed conflicts, such as the War of the Triple Alliance in the 19th century, and the establishment of Jesuit missions, gathering native populations into high-density communities, caused large-scale demographic changes. These changes in the post-colonization period, among others, would have given the conditions for the establishment of the surviving HBV-F lineages [15, 31, 33].

The overall analysis here presented demonstrated strong evidence about the period of HBV-F high spreading in Latin America. However, the sampling date ranges of the sequences used for evolutionary analysis were not wide enough and this could Bayesian inferences. To reduce bias due to these limitations, we carry out all analyses with care to ensure the best statistical supports of our inferences, such as the selection of the most statistically suitable molecular clocks for our data, the best temporal signal, the highest values of posterior probability and Bayes Factor to support the analyses of phylodynamics and phylogeography. Deep calibrations with HBV trees using divergence times of different indigenous human populations for genomic fossils and ancient human samples are necessary for more definitive results [5, 35, 36]. However, the present study demonstrated a historical correlation with important events (such as the War of the Triple Alliance) probably associated with the spread of HBV-F in Latin America. Noteworthy, all Bayesian analyses here performed were statistically supported to report the best molecular clock and genetic model of nucleotide substitution for the HBV-F dataset. Also, the nucleotide substitution rate observed for HBV-F is in agreement with previous studies that demonstrated that the evolutionary rate of HBV using contemporary samples can vary from 2.00E-4 to 2.20E-6 s/s/y [5, 10, 31, 32–34, 37].

The present study aimed at determining the spread of HBV-F in Latin America. More anthropological data would be necessary to fully understand the origin of HBV-F. We hypothesized that the dissemination of this genotype on a large scale occurred mainly after the arrival of Europeans in Latin America and due to the War of the Triple Alliance. Therefore, the findings presented here will be very useful to a better comprehension of the evolutionary history of HBV-F in Latin America.

In summary, the distribution of genotypes and the phylodynamic reconstruction showed the impact of both global and regional migrations in shaping the HBV molecular epidemiology in Latin America. The

new molecular data provide valuable information for characterizing the evolution of Native American HBV-F.

Declarations

Author contribuitions

JM Wolf, D Simon, and VR Lunge designed the study. JM Wolf and TK Mazeto performed the bioinformatics analyses. JM Wolf, TK Mazeto, D. Simon, and VR Lunge wrote the first draft of the manuscript and contributed to the literature review and discussion of results. All authors contributed to and have approved the final manuscript.

Conflicts of interest

The authors declare no conflicts of interest.

References

1. WHO, World Health Organization. Hepatitis B, 2020. Available from: [http://www.who.int/mediacentre/factsheets/fs204/en/]. Access: March 2021.
2. Tong S, Revill P. Overview of hepatitis B viral replication and genetic variability. J Hepatol. 2016;64(1 Suppl):S4-S16. Available from: https://doi.org/10.1016/j.jhep.2016.01.027
3. Cagliani R, Forni D, Sironi M. Mode and tempo of human hepatitis virus evolution. Comput Struct Biotechnol J. 2019;17:1384-1395. Available from: https://doi.org/10.1016/j.csbj.2019.09.007
4. McNaughton AL, D'Arienzo V, Ansari MA, Lumley SF, Littlejohn M, Revill P, et al. Insights From Deep Sequencing of the HBV Genome-Unique, Tiny, and Misunderstood. Gastroenterology. 2019;156(2):384-399. Available from: https://doi.org/10.1053/j.gastro.2018.07.058

5. Paraskevis D, Magiorkinis G, Magiorkinis E, Ho SY, Belshaw R, Allain JP, et al. Dating the origin and dispersal of hepatitis B virus infection in humans and primates. Hepatology. 2013;57(3):908-16. Available from: https://doi.org/1002/hep.26079

6. Mühlemann B, Jones TC, Damgaard PB, Allentoft ME, Shevnina I, Logvin A, et al. Ancient hepatitis B viruses from the Bronze Age to the Medieval period. Nature. 2018;557(7705):418-423. Available from: https://doi.org/1038/s41586-018-0097-z

7. Kahila Bar-Gal G, Kim MJ, Klein A, Shin DH, Oh CS, Kim JW, et al. Tracing hepatitis B virus to the 16th century in a Korean mummy. Hepatology. 2012;56(5):1671-80. Available from: https://doi.org/1002/hep.25852

8. Patterson Ross Z, Klunk J, Fornaciari G, Giuffra V, Duchêne S, Duggan AT, et al. The paradox of HBV evolution as revealed from a 16th century mummy. PLoS Pathog. 2018;14(1):e1006750. Available from: https://doi.org/1371/journal.ppat.1006750

9. Zehender G, Ebranati E, Gabanelli E, Sorrentino C, Lo Presti A, Tanzi E, et al. Enigmatic origin of hepatitis B virus: an ancient travelling companion or a recent encounter? World J Gastroenterol. 2014;20(24):7622-34. Available from: https://doi.org/3748/wjg.v20.i24.7622

10. Mojsiejczuk L, Elizalde MM, López G, Figueredo D, Marquez N, Campos RH, et al. Molecular epidemiology of hepatitis B virus in Paraguay. Infect Genet Evol. 2019;71:91-97. Available from: https://doi.org/1016/j.meegid.2019.03.020

11. 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(23):7152-68. Available from: https://doi.org/3748/wjg.v20.i23.7152

12. Croagh CM, Desmond PV, Bell SJ. Genotypes and viral variants in chronic hepatitis B: A review of epidemiology and clinical relevance. World J Hepatol. 2015;7(3):289-303. Available from: https://doi.org/4254/wjh.v7.i3.289

13. Velkov S, Ott JJ, Protzer U, Michler T. The Global Hepatitis B Virus Genotype Distribution Approximated from Available Genotyping Data. Genes (Basel). 2018;9(10):495. Available from: https://doi.org/3390/genes9100495

14. Roman S, Panduro A. HBV endemicity in Mexico is associated with HBV genotypes H and G. World J Gastroenterol. 2013;19(33):5446-53. Available from: https://doi.org/3748/wjg.v19.i33.5446

15. Roman S, Jose-Abrego A, Fierro NA, Escobedo-Melendez G, Ojeda-Granados C, Martinez-Lopez E, et al. Hepatitis B virus infection in Latin America: a genomic medicine approach. World J Gastroenterol. 2014;20(23):7181-96. Available from: https://doi.org/3748/wjg.v20.i23.7181

16. Godoy BA, Pinho JRR, Fagundes NJR. Hepatitis B Virus: Alternative phylogenetic hypotheses and its impact on molecular evolution inferences. Virus Res. 2020;276:197776. Available from: https://doi.org/1016/j.virusres.2019.197776

17. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. RDP4: Detection and analysis of recombination patterns in virus genomes. Virus Evol. 2015;1(1):vev003. Available from: https://doi.org/1093/ve/vev003

18. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772-80. Available from: https://doi.org/1093/molbev/mst010

19. Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, et al. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS Comput Biol. 2019;15(4):e1006650. Available from: https://doi.org/1371/journal.pcbi.1006650

20. Kass RE, Raftery AE. Bayes factors. J Am Stat Assoc. 1995;90(430):773-95.

21. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. Syst Biol. 2018;67(5):901-904. Available from: https://doi.org/1093/sysbio/syy032

22. Lemey P, Rambaut A, Drummond AJ, Suchard MA. Bayesian phylogeography finds its roots. PLoS Comput Biol. 2009;5(9):e1000520. Available from: https://doi.org/1371/journal.pcbi.1000520

23. Bielejec F, Rambaut A, Suchard MA, Lemey P. SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. Bioinformatics. 2011;27(20):2910-2. Available from: https://doi.org/1093/bioinformatics/btr481

24. Devesa M, Loureiro CL, Rivas Y, Monsalve F, Cardona N, Duarte MC, et al. Subgenotype diversity of hepatitis B virus American genotype F in Amerindians from Venezuela and the general population of Colombia. J Med Virol. 2008 Jan;80(1):20-6. Available from: https://doi.org/1002/jmv.21024

25. von Meltzer M, Vásquez S, Sun J, et al. A new clade of hepatitis B virus subgenotype F1 from Peru with unusual properties. Virus Genes. 2008;37(2):225-30. Available from: https://doi.org/1007/s11262-008-0261-x

26. Arauz-Ruiz P, Norder H, Robertson BH, Magnius LO. Genotype H: a new Amerindian genotype of hepatitis B virus revealed in Central America. J Gen Virol. 2002;83(Pt 8):2059-2073. Available from: https://doi.org/1099/0022-1317-83-8-2059

27. González López Ledesma MM, Mojsiejczuk LN, Rodrigo B, Sevic I, Mammana L, Galdame O, et al. Hepatitis B virus genotype distribution and genotype-specific BCP/preCore substitutions in acute and chronic infections in Argentina. PLoS One. 2015;10(3):e0121436. Available from: https://doi.org/1371/journal.pone.0121436

28. Rodrigo MB, Mojsiejczuk LN, Torres C, Sevic I, González López Ledesma MM, Perez PS, et al. Analysis of the Molecular Evolution of Hepatitis B Virus Genotypes in Symptomatic Acute Infections in Argentina. PLoS One. 2016;11(7):e0159509. Available from: https://doi.org/1371/journal.pone.0159509

29. Livingston SE, Simonetti JP, McMahon BJ, Bulkow LR, Hurlburt KJ, Homan CE, et al. Hepatitis B virus genotypes in Alaska Native people with hepatocellular carcinoma: preponderance of genotype F. J Infect Dis. 2007;195(1):5-11. Available from: https://doi.org/1086/509894

30. Lampe E, Mello FCA, do Espírito-Santo MP, Oliveira CMC, Bertolini DA, Gonçales NSL, et al. Nationwide overview of the distribution of hepatitis B virus genotypes in Brazil: a 1000-sample

multicentre study. J Gen Virol. 2017;98(6):1389-1398. Available from: https://doi.org/1099/jgv.0.000789

31. Mello FC, Araujo OC, Lago BV, Motta-Castro AR, Moraes MT, Gomes SA, et al. Phylogeography and evolutionary history of hepatitis B virus genotype F in Brazil. Virol J. 2013;10:236. Available from: https://doi.org/1186/1743-422X-10-236

32. Mojsiejczuk LN, Torres C, Sevic I, Badano I, Malan R, Flichman DM, et al. Molecular epidemiology of hepatitis B virus in Misiones, Argentina. Infect Genet Evol. 2016;44:34-42. Available from: https://doi.org/1016/j.meegid.2016.06.032

33. Torres C, Piñeiro y Leone FG, Pezzano SC, Mbayed VA, Campos RH. New perspectives on the evolutionary history of hepatitis B virus genotype F. Mol Phylogenet Evol. 2011;59(1):114-22. Available from: https://doi.org/1016/j.ympev.2011.01.010

34. Mojsiejczuk L, Torres C, Flichman D, Campos RH. Long-term evolution of hepatitis B virus genotype F: Strong association between viral diversification and the prehistoric settlement of Central and South America. J Viral Hepat. 2020; 27(6):620-630. Available from: https://doi.org/1111/jvh.13273

35. Gilbert C, Feschotte C. Genomic fossils calibrate the long-term evolution of hepadnaviruses. PLoS Biol. 2010;8(9):e1000495. Available from: https://doi.org/1371/journal.pbio.1000495

36. Simmonds P, Aiewsakun P, Katzourakis A. Prisoners of war - host adaptation and its constraints on virus evolution. Nat Rev Microbiol. 2019;17(5):321-328. Available from: https://doi.org/1038/s41579-018-0120-2

37. Alvarado Mora MV, Romano CM, Gomes-Gouvêa MS, Gutierrez MF, Botelho L, Carrilho FJ, et al. Molecular characterization of the Hepatitis B virus genotypes in Colombia: a Bayesian inference on the genotype F. Infect Genet Evol. 2011;11(1):103-8. Available from: https://doi.org/1016/j.meegid.2010.10.003
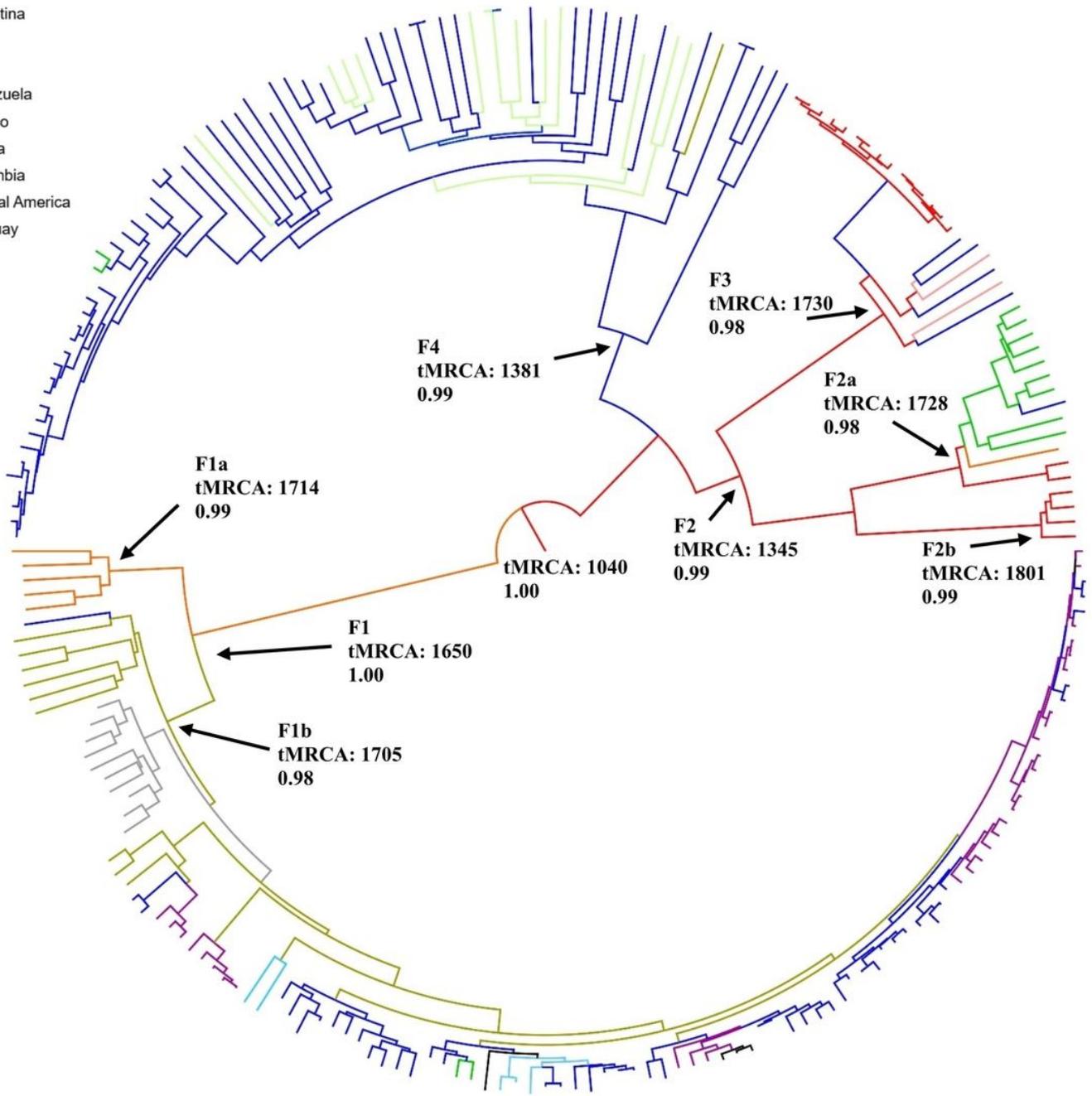
Figures

Figure 1

Bayesian tree of HBV-F complete genome sequences. Significant posterior probabilities and the time to the most recent common ancestor (tMRCA) are demonstrated in the nodes (posterior probabilities >0.95).
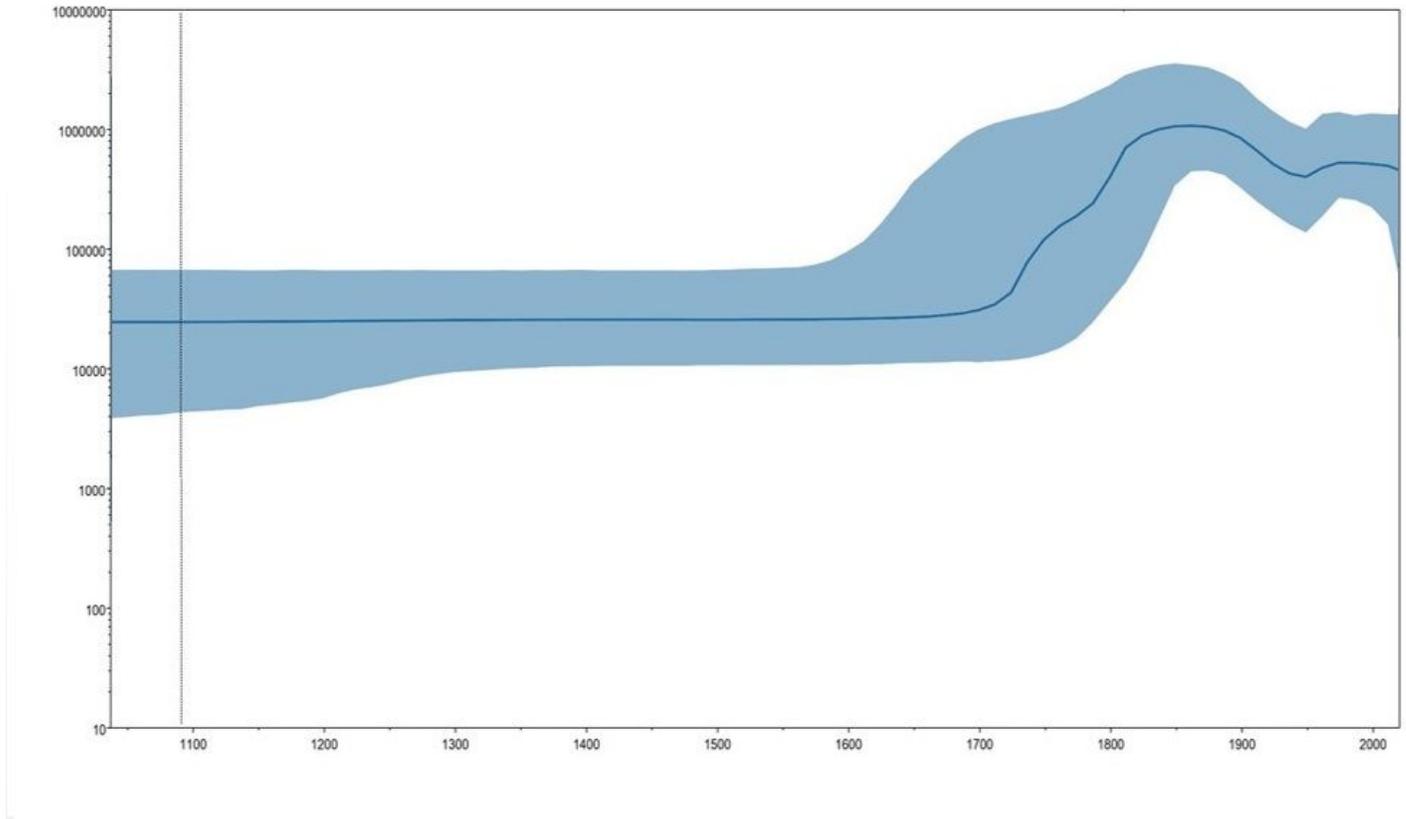
Figure 2

Bayesian skyline plot of HBV-F complete genome sequences. The effective number of infections is reported on the Y-axis. The timeline is reported on the X-axis. The colored area corresponds to the 95% credibility intervals of highest probability density (95% HPD). The vertical line indicates the 95% lower HPD (dotted) of the tree root.
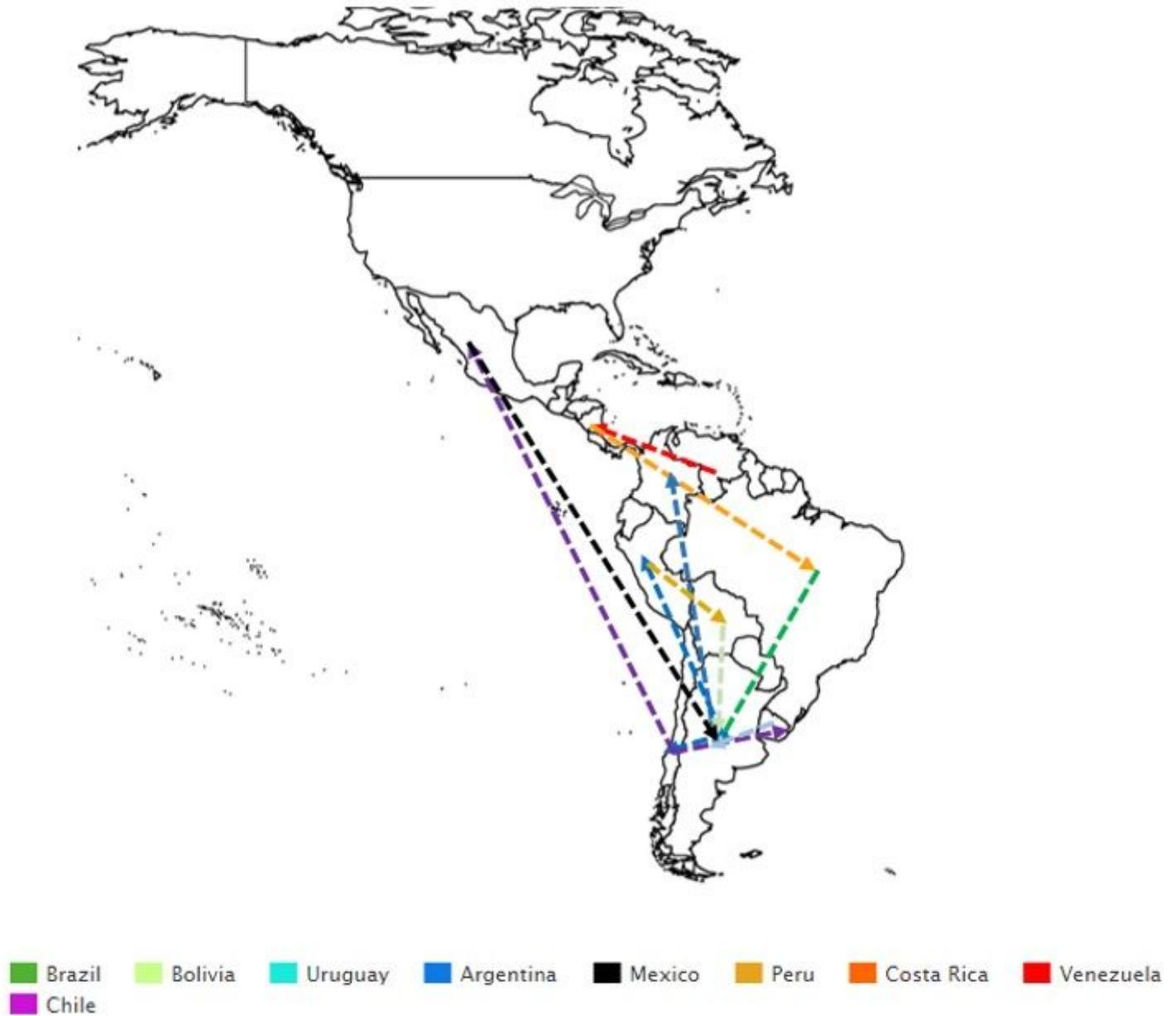
Figure 3

Spatial dissemination of HBV-F across the world. Lines between countries represent sequences in which transition location occurs and were supported by a significant non-zero rate (Bayes Factor >3).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- HBVFSupplementaryTable.docx