Molecular Epidemiological Characteristics and Risk Factors for Acquiring HBV Among Li Ethnic in Baisha County, Hainan Island-Subgenotype D3 Was First Discovered in China

Ru Xu\textsuperscript{1,2,†}, Dandan Song\textsuperscript{3,†}, Min Wang\textsuperscript{1,2}, Jieting Huang\textsuperscript{1,2}, Qiao Liao\textsuperscript{1,2}, Zhengang Shan\textsuperscript{1,2}, Xia Rong\textsuperscript{1,2,4,*} and Yongshui Fu\textsuperscript{1,2,4,5,*}

\textsuperscript{1} Institute of Clinical Blood Transfusion, Guangzhou Blood Center, Guangzhou, China; \textsuperscript{2} The Key Medical Laboratory of Guangzhou, Guangzhou, China; \textsuperscript{3} Department of Clinical Laboratory, The Fifth People’s Hospital of Zhuhai, Zhuhai, China; \textsuperscript{4} School of Laboratory Medicine and Biotechnology, Southern Medical University, Guangzhou, China; \textsuperscript{5} Zhujiang Hospital of Southern Medical University, Guangzhou, China

The residents of Baisha, a county of Hainan Island, mainly composed of Li ethnic population and relatively closed living environment with its unique geographical location. Our previous study showed that Li ethnic population of Baisha is an endemic center for hepatitis C virus, with significantly higher rates than in other parts of China. However, the epidemiology of HBV in this region remains unclear. Therefore, we conducted a comprehensive epidemiological survey of HBV in Baisha County, including 1,682 Li ethnic residents. The total seropositive rate for HBsAg was 10.2% and was higher than other parts of China. HBV-positive status was associated with the 20–40-year-old group (OR = 1.27, 95%CI 1.04–1.39, \( P < 0.01 \)) and alcohol consumption (OR = 2.17, 95%CI 1.58–2.99, \( P < 0.01 \)). Phylogenetic analysis showed that HBV subgenotype D3 was predominant in Baisha County which was first discovered in China, followed by C5, C1, B2, and undetermined subgenotypes which were significantly different from other geographical distribution of main genotypes in China. The most recent common ancestor (tMRCA) of the HBV genotype C in the Li ethnic of Baisha County was 1846 (95%CI: 1739–1932), and Baisha-C5 was earlier than Baisha-C1 and Baisha-C2. Most Baisha-D3 sequences were concentrated in one bundle and unrelated to those D3 genome sequences elsewhere in the world. According to the phylogenetic tree, D3 was introduced into Baisha County in 1884 (95%CI: 1816–1993) and became a local endemic virus. In conclusion, HBV infection in the Li ethnic group is characterized by a high prevalence rate in 20–40-year-old individuals and a unique genotype distribution which were significantly different from other geographical distribution of main genotypes in China, and subgenotype D3 was first discovered in China.

Keywords: hepatitis B virus, risk factors, epidemiology, Li ethnic, HBV genotype
INTRODUCTION

Hepatitis B virus (HBV) infection is a significant public health challenge, especially in developing countries with high prevalence. The epidemiology of HBV could be tracked by the prevalence of hepatitis B surface antigen (HBsAg), which declined from 9.8 to 7.2% in the general population aged 1–59 years in China (Xia et al., 1996; Liang et al., 2013). A modeling study showed that the prevalence of HBsAg was 6.1% in 2016 in mainland China, corresponding to 86 million infections (Polaris Observatory, 2018). Nevertheless, the prevalence of HBsAg was relatively low in the blood donors. For example, 1.16% was found from 1999 to 2009 in Xi’an blood donors (Zhao et al., 2014), 2.3% was found from 2005 to 2014 in Shenzhen blood donors (Wang et al., 2016); while in Guangdong, only 0.06% was found in repeat donors who were born after 1992 following nationwide implementation of universal HBV vaccination at birth (Tang et al., 2018). Baisha County is a mountainous rural area of Hainan Province that is located between Southeast Asia and East Asia close to Vietnam. Most native residents belong to the Li ethnic population (Li et al., 2008). We have reported the unique genetic background (HLA class I and II alleles) of the Baisha Li ethnic, which was distinct from Chinese Han ethnicity (Huang et al., 2019). Our recent study showed that Baisha County is an endemic center for HCV infection, where the rate of anti-HCV was significantly higher than other areas in China (Xu et al., 2019). However, the prevalence of HBV infection remains unclear in this county. Some reports have shown that the prevalence of HBV was higher in rural areas than in urban areas (Zhang et al., 2013; Liu et al., 2017; Wang et al., 2019). Two studies also reported that the seroprevalence of HBV infection was higher in Hainan Province than in other parts of China (Li et al., 2008; Sun et al., 2008). Therefore, conducting an epidemiological investigation of HBV in this special region is not only vital to understand the burden of HBV infection and to predict the future prevalence of HBV in the Hainan Li ethnic population, but also helpful to understand the cause of the distinctive HBV epidemiology in this area.

Hepatitis B virus is divided into 10 genotypes (A-J), and the geographical distribution of each genotype was distinct, even in the same country (Sunbul, 2014). The HBV genotype B was primarily distributed in the south and genotype C in the north of China (Lin et al., 2013; Li et al., 2017). Genotype D was endemic to northwestern China (Nie et al., 2012). C/D recombinant genotype was found in Qinghai-Tibet Plateau in Western China and predominantly in Tibet (Zhou et al., 2011). Genotypes E, F, G, H and J have not been reported in China (Su et al., 2020). A study showed that HBV vaccination did not interfere with HBV genotype distribution (Su et al., 2020). However, different genotypes are related to clinical progression, response to antiviral treatment, and prognosis (Yuen et al., 2018). HBV genotype can influence the clinical outcome of HBV infection. Mayerat et al. (1999) found that there is a clear relationship between genotype A and chronic infection when compared to genotype D. By contrast, studies have revealed that the clinical outcomes of chronic HBV infections are more serious in patients with genotypes C and D than in those with genotypes A and B. Furthermore, the patients with genotypes C and D were prone to cirrhosis and hepatocellular carcinoma (HCC), the risk of cirrhosis and HCC in patients with genotype F infection was similar to genotypes C and D (Shi, 2012). HBV genotypes can also influence the response to antiviral treatment of HBV infection. For example, genotypes A and B have higher response to interferon-based therapy than genotypes C, D and mixed genotypes. A study showed that genotype A was a 20-fold increase in the risk of nucleos(t)ide analogs resistance when compared to that of genotype D for a mean period of 12 months (Maria et al., 2002). In addition, the characterization of different HBV genotypes in a given population may have epidemiological importance, as the HBV genotype generally reflects its country of origin and can be used to track transmission patterns.

In this study, we focused on the molecular epidemiological characteristics and risk factors for HBV infection in Baisha County in southernmost China. We found that the HBV infection rate was high, and there were unique genotypes, especially D3, which was the first subgenotype discovered in China. We speculated that the closed geographical location and the unique living habits of Li ethnic played an important role in the local transmission of the virus. Such information could help improve HBV prevention and control strategies in this county and potentially avoid transmission to other regions of Hainan Island. Moreover, it could give insight into the evolution and origin of HBV in the Li ethnic population.

MATERIALS AND METHODS

Sample Collection

From July 2014 to October 2015, 1,682 11–95-year-old volunteers from four communities and seven townships in Baisha County were recruited in this study by a random sampling method, the same cohort as our previous study (Xu et al., 2019). The participants’ socio-demographic features and history of blood transfusion, alcohol consumption, surgery, acupuncture, tattoos, drug abuse, body piercings, dental operations, and HBV infection among family members were recorded. Then, blood samples were collected to assess HBV markers. The physicians ensured that individuals were personally interviewed to assure their complete understanding of the study, and the participants provided written informed consent before enrollment. The Institutional Review Board approved this study at the Guangzhou Blood Center, and the guidelines set by this board were strictly followed. All the study protocols conformed to the 1975 Declaration of Helsinki’s ethical guidelines and were approved by the Medical Ethics Committee of the Guangzhou Blood Center.

Hepatitis B Surface Antigen and Hepatitis B Virus DNA Detection

According to the manufacturers’ instructions, the 1,682 plasma specimens were tested for HBsAg using two independent HBV ELISA assays (Beijing Wantai, China and Monolisa HBsAg ULTRA, Bio-rad, United States). The HBV DNA levels were assessed using an in-house combination of qPCR, with the lowest detection limit being 5 IU/ml, as described previously (Zheng et al., 2015).
PCR Amplification and Sequencing
According to the manufacturers’ instructions, HBV DNA was extracted from 250 µL plasma samples using the MagNA Pure LC Nucleic Acid Isolation Kit-Large Volume (Roche Diagnostics). A nested PCR performed amplification with a partial P region primer from the protocol available at the GOV.UK website1, (primers were shown in Supplementary Table 1). The final amplicon was approximately 1 kb in length and covered the entire HBsAg region (Position: nucleotides 1–1,095, numbered according to with HBV genotype B and C). If the partial P region was unsuccessfully amplified, part of Pre-S/S (Position: nucleotides 1–423 and 2,817–3,215, numbered as before) were independently amplified according to a previous study (Nie et al., 2012) (primers were shown in Supplementary Table 1). Eleven highly viremic serum samples of the subgenotype D3 were performed to obtain whole genomes by other two fragments (Position: nucleotides 1,797–274) primers which were shown in Supplementary Table 2. The amplified products were purified and sequenced by Beijing Genomics Institute (BGI, shown in Supplementary Table 2). The final amplicon was approximately 1 kb in length and covered the entire HBsAg region (Position: nucleotides 1–1,095, numbered according to with HBV genotype B and C). If the partial P region was unsuccessfully amplified, part of Pre-S/S (Position: nucleotides 1–423 and 2,817–3,215, numbered as before) were independently amplified according to a previous study (Nie et al., 2012) (primers were shown in Supplementary Table 1). Eleven highly viremic serum samples of the subgenotype D3 were performed to obtain whole genomes by other two fragments (Position: nucleotides 1,797–274) primers which were shown in Supplementary Table 2. The amplified products were purified and sequenced by Beijing Genomics Institute (BGI, Beijing, China). The obtained sequences were then inspected and aligned using the SeqMan program from DNASTAR2 and BioEdit software3.

Evolutionary Rates and Divergence Dates
To co-estimate evolutionary rates, timescale phylogeny, and model parameters, we used the Bayesian Markov chain Monte Carlo (MCMC) method implemented in the BEAST v1.10.4 (Bayesian Evolutionary Analysis by Sampling Trees) package with strict, lognormal, and exponential clock model under a less restrictive Bayesian skyline plot (BSP) coalescent model. Eventually, the GTR + G + I nucleotide substitution and lognormal clock models were chosen as the best models for analysis. The MCMC chains were run for at least 200 million generations and sampled every 2,000 generations. Convergence was assessed based on the effective sampling size (ESS) after a 10% burn-in using Tracer software version 1.5; only ESS values above 200 were accepted. The reconstructed trees were examined and edited using FigTree v1.4.0, which was also used to estimate various nodes’ evolutionary rates and dates on the MCMC tree.

Statistical Analysis
A univariate analysis used the chi-square test to detect associations between HBV infections and participants’ socio-demographic characteristics/risk factors. Measurement data were presented as the mean ± standard error and compared using the two-sample t-test. A multivariate logistic regression analysis was performed to determine the predictors of HBsAg positivity. Furthermore, data were expressed as the mean (standard deviation; SD), percentage (%), and OR (95% confidence interval; CI), where appropriate; P < 0.05 was considered statistically significant. All statistical analyses were performed using SPSS Statistics for Windows version 19.0 (IBM Corp., Armonk, New York).

Nucleotide Sequence Accession Numbers
The nucleotide sequences reported in this study were deposited in GenBank with the following accession numbers MW244442-MW244564 and MW575198-MW575215.

RESULTS
Hepatitis B Virus Infection in the Baisha Li Ethnic Population
Total of 1,682 blood samples from the Baisha Li group were tested for HBsAg, 171 (10.2%) and 1,465 (87.0%) were reactive and non-reactive, respectively, with two independent assays (Table 1). In addition, 30 (1.8%) were only reactive with Monolisa HBsAg ULTRA from Bio-Rad, 16 (1.0%) were only reactive with the HBV ELISA assay from Wantai. 149 HBV DNA + was detected in 171 HBsAg + samples with a mean viral load of 4.67 × 104 ± 223 IU/ml.

Risk Factors for Hepatitis B Virus Infection
Hepatitis B surface antigen positivity was significantly higher in 20–40-year-old (24.6%) and 40–60-year-old (14.1%) subjects

Genotype and Subgenotype Analysis
Hepatitis B virus genotypes and subgenotypes were classified directly from GenBank sequence annotations and phylogenetic trees. The phylogenetic tree was constructed based on the maximum-likelihood method using MEGA-X software4. The reliability of the tree was estimated using 1,000 bootstrap replications. Bootstrap values of the phylogenetic branches greater than 70% were considered as having a high degree of confidence. Simplot programs (available at5) were used to test for recombination. The reference sequences covering HBV genotypes A-J and major subgenotypes were obtained from GenBank and are published in peer-reviewed journals.

Hepatitis B Virus Genotype/Subgenotype Dataset
Hepatitis B Virus-C and HBV-D partial P region sequences were retrieved from the continental available in GenBank to identify the HBV-C and HBV-D subgenotypes circulating in Baisha County and determine their transmission and origin. Sequences without a known County, sampling date, and less than 1 kb were excluded. To avoid the over-representation of unrelated Baisha sequences, we built a phylogenetic tree using MEGA-X to exclude unrelated Baisha reference sequences and very close reference sequences.

1https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/342414/cf6_vrd_PolHBsAgRecommendationsFINAL_HBV_protocol.pdf
2https://www.otofy.com
3https://www.bioedit.software.informer.com
4https://www.megasoftware.net/
5https://simplot.software.informer.com/
TABLE 1 | Seroprevalence of Hepatitis B virus infections of Li ethnic.

| Monolisa HBsAg ULTRA from Bio-Rad | HBV ELISA Assay from Wantai | Total |
|-----------------------------------|-------------------------------|-------|
| Reactive                          | Reactive                     | 171 (10.2) |
| Non-reactive                      | Reactive                     | 16 (1.0)  |
| Total                             | Non-reactive                 | 187 (11.1) |

TABLE 2 | Socio-demographics of subjects according to HBsAg positivity.

| Gender | HBsAg (+) n = 171 | HBsAg (-) n = 1,511 | Univariate Chi-square analysis | Multivariate logistic regression analysis |
|--------|------------------|---------------------|-------------------------------|------------------------------------------|
|        | OR (95% CI)      | P-value             | OR (95% CI)                  | P-value                                  |
| Male   | 89 (13.6)        | 565 (86.4)          | 1.81 (1.32–2.50)             | 1.95E-04                                 |
| Female | 82 (7.98)        | 946 (92.02)         |                              |                                          |
| Age    |                  |                     |                              |                                          |
| <20y   | 17 (5.8)         | 278 (94.2)          | 0.19 (0.06–0.19)             | 2.42E-10                                 |
| 20–40y | 68 (24.6)        | 208 (75.4)          |                              |                                          |
| 40–60y | 60 (14.1)        | 366 (85.9)          | 0.50 (0.34–0.74)             | 4.05E-4                                  |
| >60y   | 26 (3.8)         | 659 (96.2)          | 0.12 (0.08–0.19)             | 7.51E-23                                 |

P-values less than 0.05 are indicated in bold.

compared to younger (10–20 years old, 5.8%) or older subjects (> 60 years old, 3.8%) in our series ($\chi^2 = 1.07E2$, d.f. = 4, $P < 0.01$, chi-square test). The detailed results of the comparison between the two groups (20–40 years old and any other age groups) are shown in Table 2. A logistic regression analysis, with adjusted confounders, confirmed that age was a significant predictor for HBV positivity (OR = 1.27, 95% CI 1.04–1.39, $P < 0.01$). There was a significant increase in the HBsAg positivity in male than female subjects (OR = 1.81, 95% CI 1.32–2.50, $P < 0.01$, chi-square test) (Table 2). Alcohol consumption was more common in HBsAg positive group than in HBsAg negative group (OR = 2.17, 95% CI 1.58–2.99, $P < 0.01$; chi-square test), while there was no difference in transfusion, drug abuse, piercings, acupuncture, tattooing, surgery, dental procedures, or a family history of HBV between HBsAg positive group and HBsAg negative group. A logistic regression analysis, with adjusted confounders, confirmed that alcohol consumption (OR = 2.97, 95% CI 2.03–4.33, $P < 0.01$) was a significant predictor for HBV positivity (Table 3).

Genotype/Subgenotype Distribution of Baisha Hepatitis B Virus Strains

Nest PCR amplification was performed on HBV DNA + samples with different PCR primers. A total of 141 DNA fragments were successfully amplified in 149 HBV DNA + cases, including partial P region (nt 1–1,095) of 123 cases and PreS/S (nt 2,817–3,215 and 1–423) of 18 cases. Neither of the two fragments were amplified in eight samples because of the very low viral loads or limited plasma volumes. The phylogenetic tree constructed from 123 partial P region sequences showed that 79 cases were D3, 10 were B2, 22 were C5, 9 were C1, 2 were new subgenotypes of genotype C, and 1 (BSY173) was a possible recombinant of genotype C and X (Figure 1). The phylogenetic tree of 18 partial Pre-S/S genes showed that 12 cases were D3, 3 were C5 and 3 were C1 (Figure 2). Eleven HBV subgenotype D3 full-length sequences were genotyped to exclude the possibility of C/D hybrid genotypes that reported from the Tibetan ethnic populations living in North-West China (Cui et al., 2002; Wang et al., 2005). The tree showed that a pure genotype D circulated among the Li ethnic populations (Supplementary Figure 1). Furthermore, Simplot also did not show any recombination events (data not shown).

Evolutionary Analyses of Genotypes C and D

Genotypes C (HBV-C) and D (HBV-D) were the main HBV genotypes detected on the Li ethnic in Baisha County (Figure 1). Thus, we performed an evolutionary analysis of HBV-C and HBV-D separately in Baisha County with the global reference sequences. BSY173 was excluded because the possible recombinant events would strongly affect the estimates of both the nucleotide substitution rate and the age of genetic diversity (Zhou and Holmes, 2007).

The evolutionary rate of HBV-C was 2.29 × 10^{-4} substitutions per site per year (s/s/y, 95% CI: 5.99 × 10^{-5}, 4.10 × 10^{-4}). The most recent common ancestor (tMRCA) of the HBV-C in the Li ethnic of Baisha County was from the year 1846 (95% CI: 1739–1932). BSM625 (an undetermined subgenotype) was the most ancient sequence in Baisha County, since it located in the root of all Baisha sequences (Figure 3). The other sequences formed subgenotypes C1, C5, and an undetermined subgenotype (B2O266). There were three transmission routes when C5 was introduced in 1890 (95% CI: 1739–1972) in Baisha County. Group I clustered with the references collected from Malaysia, the Philippines, and Netherlands. Most of the C5 sequences clustered with Thai reference sequences (Group II), while others were related to China and Malaysia (Group III). C1 may have originated from Hong Kong or Vietnam in 1918 (95% CI: 1765–1965) according to the tree topology. BSO266 was located in the root of C2, which was traced to 1938 (95% CI: 1805–1997).

The evolutionary rate of HBV-D was 9.69 × 10^{-5} s/s/y (95% CI: 3.39 × 10^{-5}, 1.62 × 10^{-4}). The tMRCA of subgenotype D3 worldwide was from around 1827 (95% CI: 1673–1975), whereas most D3 introduced in Baisha County was from around
**TABLE 3** | Risk factors for HBV infection.

| Risk Factors                  | HBsAg (+) n = 171 | HBsAg (-) n = 1,511 | Univariate Chi-square analysis | Multivariate logistic regression analysis |
|------------------------------|------------------|---------------------|-------------------------------|-----------------------------------------|
|                              |                  |                     | OR (95%CI)                    | P-value                  | OR (95%CI)                    | P-value                  |
| Alcohol consumption          |                  |                     | 2.17 (1.58–2.99)             | 1.33E-06                  | 2.97 (2.03–4.33)             | 1.73E-06                  |
| Yes                          | 98 (14.5)        | 577 (85.5)          |                              |                          |                            |                          |
| No                           | 73 (7.2)         | 934 (92.8)          |                              |                          |                            |                          |
| Drug abuse                   |                  |                     |                              |                          |                            |                          |
| Yes                          | 0 (0)            | 3 (100)             | 0.99 (0.99–1.00)             | 0.99*                    | -                           | 0.99                     |
| No                           | 171 (10.2)       | 1,506 (89.8)        |                              |                          |                            |                          |
| Tattooing                    |                  |                     |                              |                          |                            |                          |
| Yes                          | 7 (5.2)          | 128 (94.8)          | 0.46 (0.21–1.004)            | 0.052                    | 0.84 (0.36–1.93)            | 0.68                     |
| No                           | 164 (10.6)       | 1,383 (89.4)        |                              |                          |                            |                          |
| Surgery                      |                  |                     |                              |                          |                            |                          |
| Yes                          | 3 (5.5)          | 52 (94.5)           | 0.50 (0.16–1.62)             | 0.24                     | 0.38 (0.09–1.52)            | 0.17                     |
| No                           | 168 (10.3)       | 1,459 (89.7)        |                              |                          |                            |                          |
| Piercings                    |                  |                     |                              |                          |                            |                          |
| Yes                          | 38 (8.9)         | 388 (91.1)          | 0.83 (0.57–1.21)             | 0.33                     | 1.11 (0.71–1.74)            | 0.64                     |
| No                           | 133 (10.6)       | 1,123 (89.4)        |                              |                          |                            |                          |
| Acupuncture                  |                  |                     |                              |                          |                            |                          |
| Yes                          | 1 (3.6)          | 27 (96.4)           | 0.32 (0.04–2.39)             | 0.24                     | 0.48 (0.05–4.22)            | 0.50                     |
| No                           | 170 (10.3)       | 1,484 (89.7)        |                              |                          |                            |                          |
| Dental procedures            |                  |                     |                              |                          |                            |                          |
| Yes                          | 7 (16.3)         | 36 (83.7)           | 1.75 (0.77–3.99)             | 0.18                     | 2.47 (0.99–6.06)            | 0.051                    |
| No                           | 164 (10.0)       | 1,475 (90.0)        |                              |                          |                            |                          |
| Guasha                       |                  |                     |                              |                          |                            |                          |
| Yes                          | 0 (0)            | 8 (100)             | 0.99 (0.991–0.998)           | 0.99*                    | -                           | 0.999                    |
| No                           | 171 (52.8)       | 1,503 (47.2)        |                              |                          |                            |                          |
| Transfusion                  |                  |                     |                              |                          |                            |                          |
| Yes                          | 5 (17.2)         | 24 (82.8)           | 1.87 (0.70–4.96)             | 0.20                     | 2.08 (0.65–6.64)            | 0.22                     |
| No                           | 166 (10.0)       | 1,487 (90.0)        |                              |                          |                            |                          |
| A family history of HBV infection |          |                     |                              |                          |                            |                          |
| Yes                          | 2 (22.2)         | 7 (77.8)            | 2.54 (0.52–2.34)             | 0.23                     | 4.26 (0.73–24.86)           | 0.11                     |
| No                           | 169 (10.1)       | 1,504 (89.9)        |                              |                          |                            |                          |

*Data were obtained by Fisher’s exact test. P-values less than 0.05 are indicated in bold.

1884 (95%CI: 1816–1993). The Baisha sequences formed two groups (I and II), indicating two distinct routes of introduction in Baisha County. Group II contained most of the sequences and was further divided into two branches, suggesting the two transmission events with tMRCA of the main clusters calculated as 1942 (95% CI: 1768–1963) and 1941 (95% CI: 1762–1968), respectively. In addition, as shown in the tree’s topology (Figure 4), Group II was unique, Baisha-specific, and may have originated from Russia and Estonia around 1912 (95%CI: 1847–2010). The Baisha sequences clustered in Group I were very close to two references which collected from northeast China and originated in 1943 (95%CI: 1884–1991), and possibly from Russia and India in 1921 (95%CI: 1865–2001) according to the tree’s topology.

**DISCUSSION**

The prevalence of HBV infection differs across China (MacLachlan and Cowie, 2015). Baisha County, located in Hainan Province of China’s southernmost, is the main residential area of the Li ethnic population. This area is isolated by the surrounded mountains at the entrance to East Asia and have distinctive genetic characteristics (Li et al., 2008). In this study, the HBsAg prevalence was found as high as 10.2%, suggesting that Baisha County was a highly endemic region according to the three categories of endemicity: high (>8%), intermediate (2–7%) and low (<2%) (Li et al., 2008). Epidemiological studies from 2006 to 2016 showed that HBV prevalence decreased from 7.2 to 3.8% (Liang et al., 2013; Zhang et al., 2016; Wang et al., 2018). A cross-sectional study showed that the overall HBsAg prevalence of rural childbearing-aged women was 9.51% in Hainan Province (Zhang et al., 2013). Further study showed that the prevalence of HBsAg in the Li ethnic population (10.56%) was higher than in the Han ethnic populations (9.08%) (Zhang et al., 2013), which was consistent with our study and higher than the prevalence in other areas in China. In previous study, a high HCV prevalence (7.0–9.1%) was found in the same cohort (Xu et al., 2019). However, no HBV/HCV co-infection was found, which was inconsistent with the results that HBV/HCV coinfection is not uncommon in highly endemic areas because of the shared modes of transmission (Dimitris and Melanie, 2015).
We speculated that the different transmission routes were contributed to mono-infection with either HBV or HCV. In this study, the risk factors for HBV infection were 20–40-year-old group and alcohol consumption, while the risk factors for HCV infection were older age (≥ 60 years) and surgery (Xu et al., 2019).

We also found that HBsAg positivity was higher in 20–40-year-old and 40–60-year-old subjects compared to younger and older subjects, consistent with the results previously reported in Turkey, Korean and Togo (Yun et al., 2008; Min et al., 2013; Tozun et al., 2015; Kolou et al., 2017). The higher HBV infection in the 20–60-year-old range derives from lifestyles and behaviors that increase exposure to infections, such as alcohol consumption and sexual relationship. Sexual transmission was shown to be involved in HBV transmission (Nelson et al., 2016). Another possibility to explain the low prevalence observed in over 60 years old subjects is that some of them may die from HBV related complications after progression of chronic infection for years. Chronic HBV infection is considered to significantly increase the risk of liver cancer and the mortality rate of which ranks first among all kinds of malignant tumor in Hainan province especially in Baisha County where Li ethnic has inhabited for centuries (Wu et al., 2015). In addition, liver cancer is the third most common causes of life expectancy loss in Hainan Province (Dou et al., 2021). Effective vaccination can explain the lower prevalence in 10–20-year-old compared to 20–60-year-old subjects. Alcohol consumption was a significant independent determinant of HBsAg positivity. The Li ethnic populations living in the Baisha County commonly consume alcohol, especially middle-aged subjects. Szabo et al. showed that alcohol use was associated with reduced host defense (Szabo and Mandrekar, 2009). Another study showed that HBV replication was enhanced by alcohol consumption with a sevenfold increase in HBsAg and viral DNA levels (Chan and Levitsky, 2016). However, little is known about the potential role of alcohol in HBV infection. Some reports have indicated an association between male gender and greater HBsAg positivity (Kong et al., 2014; Ikezaki et al., 2016). In our study, no correlation between gender and HBV infection in multivariate analysis was found, although the univariate chi-square test showed more male patients to be HBsAg-positive. We speculated that this might due to the strong correlation between gender and alcohol consumption ($\chi^2 = 85.37$, OR = 0.39, 95%CI 0.32–0.48, $P = 2.48E-20$, data not shown) and the rates of alcohol consumption in male was significantly higher than in female [male (51.3%) vs. female (48.7%): $\chi^2 = 80.06$, $P < 0.01$, data not shown].
The mainly prevalent HBV genotypes in China are B and C (Li et al., 2015). Genotype C was predominant in northern China, while genotype B was mainly distributed in southern China (Yin et al., 2010; Li et al., 2015; Shen et al., 2015). In addition, C/D recombinants have been identified in northwest China (Zhou et al., 2011; Wang et al., 2014). Interestingly, the genotype distribution of the Li ethnic in Baisha County differed dramatically from any other regions of China. Even though Baisha County is located in southern China, genotype B is the lowest (7.1%). The proportion of genotype C (28.4%) in the Li ethnic was higher than that of genotype B, which was consistent with previous reports in Hainan Island (Zeng et al., 2009; Wang et al., 2014). Meanwhile, the subgenotypes of genotype C in Baisha County were significantly different from those of other regions. In Baisha, C5 was the main endemic subgenotype, followed by C1. C5 was not found in China except for three cases from Yunnan Province (Wang et al., 2014). Notably, the genotype D was the dominant genotype in the Li ethnic of Baisha County, all of which were subgenotype D3. The genotype D was reported only in a few areas in China, such as Xinjiang, Hainan, Gansu, and Shenyang (Ma et al., 2011; Nie et al., 2012; Zhang et al., 2015; Pu et al., 2016). For example, China's investigation of HBV genotypes in Xinjiang found that most of the Xinjiang Uygur population infected with genotype D (64.2%), and all were D1 subgenotype (Nie et al., 2012). Although the genotype D was found in Shenyang (28%) (Ma et al., 2011), Hainan Island (11.94%) (Zeng et al., 2009) and Gansu (47.19%) (Zhang et al., 2015), none of these sequences have any detailed subgenotype information in text and have not been submitted to websites, so we do not know whether they were D3. We speculated that this unique HBV genotype distribution pattern relates to the location of Baisha County, to be specific, a mountainous rural area on Hainan Island that isolated from mainland China. The native residents in Baisha belong to the Li ethnic that derived from a subgroup of Austronesians. The unique characteristics of this cohort may have a unique HBV transmission pattern and origin.

Therefore, we performed genetic evolution analyses on the specific genotypes C and D to trace the HBV transmission and evolution in Baisha County. The genotype C phylogenetic tree showed that the Li ethnic's HBV-C genome sequences in Baisha County was mainly distributed into two clusters (C1 and C5) and were associated with other C genome sequence branches in the world. The tMRCA of the HBV-C of the Li ethnic was from about the year of 1846 (95% CI: 1739–1932), which was later than China's Opium War in 1840. The population and trade
exchanges between Hainan Island and Southeast Asia can be traced back to ancient times, especially after 1840, when China gradually relaxed its foreign policy restrictions under pressure from European and American countries, allowing migration exchange between Hainan Island and Southeast Asia to a large scale. A previous study showed that genotype C5 might have originated from Southeast Asia (Zhu et al., 2015). However, there is no clear evidence to support this point in the current study as all Baisha C5 sequences were dispersed in references from Southeast Asia such as Thailand, Malaysia and Philippines, suggesting that there have been exchanges occasionally between Baisha County and Southeast Asia. Nevertheless, according to the tree's topological structure in this study, Baisha C1 clustered with sequences from Hong Kong that may have originated from Southeast Asia. Here in our results, Baisha C5 (1890) was earlier than C1 (1918) and C2 (1938), which was in agreement with the previously estimated divergence times for these subgenotypes (Paraskevis et al., 2013).

The subgenotype D3 BEAST tree showed that only four sequences of the Li ethnic in Baisha County were isolated (Group I), while all other sequences were concentrated in one bundle and were not related to D3 genome sequences elsewhere in the world (Group II). The analysis showed that most subgenotype D3 entered the Li ethnic of Baisha County in about 1912 (95% CI: 1847–2010) and became endemic. This event occurred later than the time when subgenotype D3 entered in Argentina and Brazil (Spitz et al., 2019) which was supported by the tree’s topology in this study, as Argentina and Brazil were closer to the tree’s root (Figure 4). Most Baisha D3 sequences formed a Baisha-specific group, suggesting that D3 was introduced from Russia and Estonia and formed a local endemic. There is no population geographic evidence to support such transmission
We speculated that the lack of linkage reference sequences in GenBank failed us to explain how D3 in Baisha originated from Russia or Estonia. More sequences from different regions and countries are needed to perform evolutionary analysis. Another small branch of Baisha D3 was close to the reference sequences collected from China, India, and Russia, indicating that they entered Baisha County by cross-contact between these countries, but did not become endemic for some reasons.

To our knowledge, it is the first study regarding the epidemiology of HBV in the Li ethnic population of Baisha County in Hainan Island, where was also an endemic center for HCV. In addition, the current study also discovered HBV subgenotype D3 in China for the first time. Our results were important for understanding the molecular epidemiological characteristics and risk factors of HBV in the Li ethnic population in China. Whether these observations are generalizable to populations outside of Hainan Island should be considered in the further study. Therefore, more samples from other ethnic population and other regions outside Baisha County even outside of Hainan Island are required to further conduct a more comprehensive analysis and to better understand the epidemiology of HBV among different ethnic populations across China.

CONCLUSION

The current study revealed that the seropositive rate for HBsAg was 10.2% among Li ethnic population in Baisha County, Hainan Island. HBV-positive status was associated with 20–40-year-old subjects and alcohol consumption. The most prevalent HBV subgenotype D3 was first discovered in China, followed by C5, C1, B2 and undetermined subgenotypes. The distribution of HBV genotype
among Li ethnic in Baisha County was unique compared with other regions in China. The tMRCA of the HBV genotype C in the Li ethnic was 1846 (95%CI: 1739–1932), and Baisha-C5 was earlier than Baisha-C1 and Baisha-C2. Most D3 sequences were concentrated in one bundle and unrelated to those D3 genome sequences elsewhere in the world, which was introduced into Baisha County in 1884 (95%CI: 1816–1993) and became a local endemic virus.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Medical Ethics Committee of Guangzhou Blood Center. Written informed consent to participate in this study was provided by the participants’ legal guardian/next of kin.

REFERENCES

Chan, C., and Levitsky, J. (2016). Infection and alcoholic liver disease. Clin. Liver Dis. 20, 595–606. doi: 10.1016/j.cld.2016.02.014

Cui, C., Shi, J., Hui, L., Xi, H., Zhuoma, Quni, et al. (2002). The dominant hepatitis B virus genotype identified in Tibet is a C/D hybrid. J. Gen. Virol. 83(Pt 11), 2773–2777. doi: 10.1099/0022-1317-83-11-2773

Dimitris, K., and Melanie, D. (2015). The spectrum of HBV/HCV coinfection: epidemiology, clinical characteristics, viral interactions and management. Ann. Gastroenterol. 28, 221–228.

Dou, Q., Huang, C., Luo, Q., Cao, X., Huang, C., Wu, G., et al. (2021). Analysis of life expectancy and cause-eliminated life expectancy of Hainan residents in 2018. Med. Soc. 34, 72–75.

Huang, J., Xu, R., Wang, M., Liao, Q., Huang, K., Shan, Z., et al. (2019). Association of HLA-DQβ1*03:01 and DRB1*11:01 with spontaneous clearance of hepatitis C virus in Chinese Li ethnicity, an ethnic group genetically distinct from Chinese Han ethnicity and infected with unique HCV subtype. J. Med. Virol. 91, 1830–1836. doi: 10.1002/jmv.25531

Ikezaki, H., Furuyso, N., Hiramine, S., Ura, K., Mitsumoto-Kaseida, F., Takayama, K., et al. (2016). Association of IL28B rs8099917 genotype and female sex with spontaneous clearance of hepatitis C virus infection: a Japanese cross-sectional study. Arch. Virol. 161, 641–648. doi: 10.1007/s00705-015-2703-9

Kolou, M., Katawa, G., Salou, M., Gozo-Akakpo, K. S., Dossim, S., Kwarteng, A., et al. (2017). High prevalence of hepatitis B virus infection in the age range of 20–39 years old individuals in lome. Open Virol. J. 11, 1–7. doi: 10.2174/1874357901710011001

Kong, F., Pan, Y., Chi, X., Wang, X., Chen, L., Ls, J., et al. (2014). Factors associated with spontaneous clearance of hepatitis C virus in Chinese population. Biomed Res. Int. 2014:527030. doi: 10.1155/2014/527030

Li, D., Li, H., Ou, C., Lu, Y., Sun, Y., Yang, B., et al. (2008). Paternal genetic structure of Hainan aborigines isolated at the entrance to East Asia. PLoS One 3:e2168. doi: 10.1371/journal.pone.0002168

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.837746/full#supplementary-material

AUTHOR CONTRIBUTIONS

YF, XR, and RX conceived the study. RX, DS, MW, and JH conducted the study. RX, DS, QL, and ZS analyzed the data. RX and DS wrote the manuscript. JH, XR, YF, and RX revised and finalized the manuscript. All authors have read and agreed to the published version of the manuscript.

FUNDING

This work was supported by grants from the National Science Foundation of China (32000666), The Key Medical Disciplines and Specialities Program of Guangzhou (2021–2013), and Guangdong Basic and Applied Research Foundation (2020A1515010118).

Li, H. M., Wang, J. Q., Wang, R., Zhao, Q., Li, L., and Zhang, J. P. (2015). Hepatitis B virus genotypes and genome characteristics in China. World J. Gastroenterol. 21, 6684–6697. doi: 10.3748/wjg.v21.i12.6684

Li, X., Liu, Y., Xin, S., Ji, D., You, S., and Hu, J. (2017). Comparison of detection rate and mutational pattern of drug-resistant mutations between a large cohort of genotype b and genotype C hepatitis B virus-infected patients in North China. Microb. Drug Resist. 23, 516–522. doi: 10.1089/mdr.2016.0093

Liang, X., Bi, S., Yang, W., Wang, L., Cui, G., Cui, F., et al. (2013). Reprint of: epidemiological serosurvey of hepatitis B in China—declining HBV prevalence due to hepatitis B vaccination. Vaccine 31(Suppl. 9), J21–J22. doi: 10.1016/j.vaccine.2013.08.012

Lin, S., Liu, C., Shang, H., Chen, H., Yang, B., Chen, J., et al. (2013). HBV serum markers of 49164 patients and heir relationships to HBV genotype in Fujian Province of China. J. Clin. Lab. Anal. 27, 130–136. doi: 10.1002/jcla.21573

Liu, J., Zhang, S., Wang, Q., Shen, H., Zhang, M., Zhang, Y., et al. (2017). Prevalence of HBsAg/HBeAg amongst 1 936 801 couples preparing for pregnancy in rural China: an observational study. J. Viral Hepat. 24, 679–686. doi: 10.1111/jvh.12693

Ma, Y., Ding, Y., Jiao, F., and Dou, X. G. (2011). Genotyping the hepatitis B virus with a fragment of the HBV DNA polymerase gene in Shenyang, China. Virol. J. 8:315. doi: 10.1186/1743-422X-8-315

MacLachlan, J. H., and Cowie, B. C. (2015). Hepatitis B virus epidemiology. Cold Spring Harb. Perspect. Med. 5:a021410. doi: 10.1101/cshperspect.a021410

Maria, B., Montserrat, C., Auriestla, V., Rosendo, J., Francisco, R., and Rafael, E. (2002). Is hepatitis B virus subtype testing useful in predicting virological response and resistance to lamivudine? J. Hepatol. 36, 445–446. doi: 10.1016/s0168-8278(01)00283-5

Mayerat, C., Mantegani, A., and Frei, P. (1999). Does hepatitis B virus (HBV) genotype influence the clinical outcome of HBV infection? J. Viral Hepat. 6, 299–304. doi: 10.1046/j.1365-2893.1999.00174.x

Min, J. A., Yoon, Y., Lee, H. J., Choi, J., Kwon, M., Kim, K., et al. (2013). Prevalence and associated clinical characteristics of hepatitis B, C, and H infections among...
injecting drug users in Korea. J. Med. Virol. 85, 575–582. doi: 10.1002/jmv.23523

Nelson, N. P., Easterbrook, P. J., and McMahon, B. J. (2016). Epidemiology of hepatitis B virus infection and impact of vaccination on disease. Clin. Liver Dis. 20, 607–628. doi: 10.1016/j.cld.2016.06.006

Nie, J., Li, J., Sun, K., Sun, M., Chen, J., Ma, J., et al. (2012). HBV/D1: a major HBV subgenotype circulating in Uygur patients with chronic HBV infection in Xinjiang, China. Arch. Virol. 157, 1541–1549. doi: 10.1007/s00705-012-1325-8

Paraskevis, D., Magiorkinis, G., Magiorkinis, E., Ho, S. Y., Belshaw, R., Allain, J. P., et al. (2013). Dating the origin and dispersal of hepatitis B virus infection in humans and primates. Hepatology 57, 908–916. doi: 10.1002/hep.26079

Polaris Observatory, C. (2018). Global prevalence, treatment, and prevention of hepatitis B virus infection in 2016: a modelling study. Lancet Gastroenterol. Hepatol. 3, 383–403. doi: 10.1016/S2468-1525(18)30056-6

Pu, Z., Li, D., Wang, A., Su, H., Shao, Z., and Zhang, J. (2016). Epidemiological characteristics of the carriers with coexistence of HBsAg and anti-HBs based on a community cohort study. J. Viral Hepat. 23, 286–293. doi: 10.1111/jvh.12492

Shen, Y. Y., Hou, W., Yang, Z. Q., and Xiao, W. (2015). Hepatitis B virus infection and genotype in asymptomatic people from 10 ethnic groups in Yunnan, China. World J. Gastroenterol. 21, 12586–12592. doi: 10.3748/wjg.v21.i44.12586

Shi, Y. H. (2012). Correlation between hepatitis B virus genotypes and clinical outcomes. Jpn. J. Infect. Dis. 65, 476–482. doi: 10.7883/keny.65.476

Spitz, N., Mello, F. C. A., Moreira, A. S., Gusatti, C. S., Martins, R. M. B., Gomes, S. A., et al. (2019). Reconstruction of the spatial and temporal dynamics of hepatitis B virus genotype D in the Americas. PLoS One 14:e0220342. doi: 10.1371/journal.pone.0220342

Su, Q., Zhang, S., Wang, F., Liu, H., Zhang, G., Zheng, H., et al. (2020). Epidemiological distribution of hepatitis B virus genotypes in 1–29-year-olds in the mainland of China. Vaccine 38, 8238–8246. doi: 10.1016/j.vaccine.2020.09.083

Sun, L., Zeng, Z., Chen, H., Fu, Z., Chen, S., Yan, X., et al. (2008). Serological epidemiological survey of hepatitis B in different populations in Hainan Province. China Trop. Med. 8, 1894–1905.

Sunbul, M. (2014). Hepatitis B virus genotypes: global distribution and clinical importance. World J. Gastroenterol. 20, 5427–5434. doi: 10.3748/wjg.v20.i18.5427

Szabo, G., and Mandrekar, P. (2009). A recent perspective on alcohol, immunity, and host defense. Alcohol. Clin. Exp. Res. 33, 220–232. doi: 10.1111/j.1530-0277.2008.00842.x

Tang, X., Allian, J.-P., Wang, H., Rong, X., Chen, J., Huang, K., et al. (2018). Incidence of hepatitis B virus infection in young Chinese blood donors born after mandatory implementation of neonatal hepatitis B vaccination nationwide. J. Viral Hepat. 25, 1008–1016. doi: 10.1111/jvh.12901

Tozun, N., Ozdogan, O., Cakaloglu, Y., Idilman, R., Karasu, Z., and Akarca, U. (2009). Seroprevalence of hepatitis B and C virus infections and risk factors in Turkey: a fieldwork TURKHEP study. Clin. Microbiol. Infect. 21, 1020–1026. doi: 10.1016/j.cmi.2015.06.028

Wang, B., Feng, Y., Li, Z., Duan, H., Zhao, T., and Zhang, A. (2014). Distribution and diversity of hepatitis B virus genotypes in Yunnan, China. J. Med. Virol. 86, 1675–1682. doi: 10.1002/jmv.24002

Wang, H., Men, P., Xiao, Y., Gao, P., Lv, M., Yuan, Q., et al. (2019). Hepatitis B infection in the general population of China: a systematic review and meta-analysis. BMC Infect. Dis. 19:811. doi: 10.1186/s12879-019-4428-y

Wang, S., Tao, Y., Tao, Y., Jiang, J., Yan, L., Wang, C., et al. (2018). Epidemiological study of hepatitis B and hepatitis C infections in Northeastern China and the beneficial effect of the vaccination strategy for hepatitis B: a cross-sectional study. BMC Public Health 18:1088. doi: 10.1186/s12889-018-5984-6

Wang, Z., Liu, Z., Zeng, G., Wen, S., Qi, Y., Ma, S., et al. (2005). A new intertype recombinant between genotypes C and D of hepatitis B virus identified in China. J. Gen. Virol. 86, 985–990. doi: 10.1099/vir.0.07711-0

Wang, Z., Zeng, J., Li, T., Zheng, X., Xu, X., Ye, X., et al. (2016). Prevalence of hepatitis B surface antigen (HBsAg) in a blood donor population born prior to and after implementation of universal HBV vaccination in Shenzhen, China. BMC Infect. Dis. 16:498. doi: 10.1186/s12879-016-1834-2

Wu, J., Deng, C., Chen, R., Wang, J., Wu, H., Wu, T., et al. (2015). An age period cohort modelling study on primary liver cancer in western region of Hainan Province. China J. Chin Mater. Med. 1797–1801.