

Xu, R. et al. (2019) HCV genotype 6 prevalence, spontaneous clearance and diversity amongst elderly members of the Li ethnic minority in Baisha County, China. *Journal of Viral Hepatitis*, 26(5), pp. 529-540.

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

This is the peer reviewed version of the following article:

Xu, R. et al. (2019) HCV genotype 6 prevalence, spontaneous clearance and diversity amongst elderly members of the Li ethnic minority in Baisha County, China. *Journal of Viral Hepatitis*, 26(5), pp. 529-540. (doi:[10.1111/jvh.13062](https://doi.org/10.1111/jvh.13062))

This article may be used for non-commercial purposes in accordance with [Wiley Terms and Conditions for Self-Archiving](#).

<http://eprints.gla.ac.uk/178297/>

Deposited on: 18 February 2019

1 **Hepatitis C virus genotype 6 prevalence, spontaneous clearance and diversity**  
2 **amongst members of the Li ethnic minority in Baisha County, China**

3  
4 **Running title: HCV genotype 6 prevalence in Li minority**

5  
6 Ru Xu<sup>1#</sup>, Yongjuan Yu<sup>2#</sup>, E. Carol McWilliam Leitch<sup>3#</sup>, Min Wang<sup>1</sup>, Ke Huang<sup>1</sup>, Jieting  
7 Huang<sup>1</sup>, Xi Tang<sup>4</sup>, Qiao Liao<sup>1</sup>, Dandan Song<sup>4</sup>, Zhengang Shan<sup>1</sup>, Chengyao Li<sup>4</sup>, John  
8 Mclauchlan<sup>3</sup>, Xia Rong<sup>1</sup>

9  
10 1 Institute of Clinical Blood Transfusion, Guangzhou Blood Center, Guangzhou,  
11 Guangdong, China, 510095

12 2 Department of Clinical Laboratory, People's Hospital of Baisha Li Autonomous  
13 County, Hainan, China, 572800

14 3 MRC-University of Glasgow Centre for Virus Research, Glasgow, United Kingdom,  
15 G61 1QH

16 4 Department of Transfusion Medicine, School of Biotechnology, Southern Medical  
17 University, Guangzhou, Guangdong, China, 510515

18  
19  
20  
21  
22 <sup>#</sup> Ru Xu, Yongjuan Yu and E. Carol McWilliam Leitch should be considered joint first  
23 authors.

24  
25 **Correspondence:** Xia Rong PhD., Institute of Clinical Blood Transfusion, Guangzhou  
26 Blood Center, Guangzhou, Guangdong, China, 510095. E-mail:[joyjoy@126.com](mailto:joyjoy@126.com);  
27 Tel/Fax: +86-20-83595903.

29 **Acknowledgements**

30

31 We are grateful to Professor Ming Zhang (Department of Epidemiology and

32 Biostatistics, Faculty of Infectious Diseases, University of Georgia, Athens, GA, USA)

33 for assistance in modifying the manuscript.

---

**Abstract**

The epidemiology of hepatitis C virus varies widely across geographic regions and ethnic groups. Our previous study showed that 6 strains isolated from Baisha County, Hainan Island, China were all new genotype 6 (gt6) subtypes which differed significantly from subtypes of other regions. In the current study, we conducted a comprehensive epidemiological survey of HCV in the Li ethnic group, native to Baisha County. Anti-HCV antibodies were detected by 2 independent ELISAs in all participants, positive results confirmed by the recombinant immunoblot assay (RIBA) and HCV RNA viral loads were measured. Univariate chi-square test and multivariable logistic regression analyses were used to determine the risk factors for HCV infection and spontaneous clearance rates. Indeterminate RIBA results were excluded or included in analyses, consequently findings were expressed as a range. Direct sequencing of partial regions within NS5B and E1 were employed for genotyping. Among 1682 participants, 117 to 153 were anti-HCV positive (7.0-9.1%), with 42.7-52.6% confirmed to have cleared infection. Anti-HCV positivity was associated with older age ( $\geq 60$ years) (OR=0.02, 95%CI 0.01-0.05,  $P<0.01$ ) and surgery (OR=2.75, 95%CI 1.36-5.57,  $P<0.01$ ), with no significant difference found between the HCV infection group and the HCV spontaneous clearance group. The gt6 subtype distribution characteristics of Baisha County were unique, complex and diverse. The sequences did not cluster with known gt6 subtypes but formed 4 Baisha community-specific groups. HCV infection in members of the Li minority ethnic group is characterized by high prevalence rates in the elderly, high spontaneous clearance rates and broad gt6 diversity.

**Key words:** Hepatitis C; Seroprevalence; Risk factors; Li minority; Genotype

---

## 59 Introduction

60

61 Hepatitis C virus (HCV) is a blood-borne virus that is a major cause of end-stage  
62 liver diseases, including decompensated cirrhosis and hepatocellular carcinoma.  
63 Approximately 6-40 million Chinese people are HCV antibody positive (anti-HCV+),  
64 with a general prevalence of 0.43% to 3.2%<sup>1-3</sup>. Recent reports from the Chinese  
65 Ministry of Health showed that the number of new HCV infections increased from  
66 39,380 cases in 2004 to 214,023 cases in 2017 (<http://en.nhfpc.gov.cn/>). Currently,  
67 there is no vaccine against HCV. Even though new direct acting antivirals (DAAs)  
68 can achieve high sustained virology response (SVR) rates, they are extremely costly  
69 and emerging resistance-associated variants (RAVs) have been linked to DAA  
70 treatment<sup>4-7</sup>. Prevention of new infections is therefore considered the most cost-  
71 effective policy to reduce overall HCV prevalence in China<sup>8</sup>.

72

73 The prevalence of HCV is distributed unevenly in different regions of China and  
74 among different populations. Very high rates of HCV infection were often found in  
75 men who have sex with men (MSM) and in human immunodeficiency virus (HIV)-  
76 infected people in recent years<sup>9,10</sup>. Also, high HCV infection prevalence is often  
77 found in some remote regions of the country. For example, HCV prevalence was as  
78 much as 9.6% in Linxian, a rural community in Henan Province of China<sup>11</sup>. As  
79 reported, HCV is most frequently transmitted by certain medical and non-medical  
80 procedures<sup>12-15</sup>. Understanding the prevalence and risk factors for HCV infection in  
81 regions with high prevalence could effectively prevent virus transmission to other  
82 geographical areas. Following initial exposure to HCV, two outcomes can occur,  
83 either chronic infection or spontaneous clearance of the virus, which occurs in 15%-40%

84 of cases<sup>16-18</sup>. Due to the fact that symptoms occur in only ~15% of HCV-infected  
85 patients<sup>19</sup>, the clearance rate may be underestimated.

86

87 HCV has been classified into 7 genotypes according to genetic variation.  
88 Generally, gt1, gt2, and gt3 are distributed globally, gt4 is prevalent in Middle Eastern  
89 countries, gt5 is restricted to the northern part of South Africa<sup>20</sup>, and gt6 is endemic in  
90 Southeast Asia and south China<sup>21-24</sup>. Previously, we conducted a comprehensive study  
91 <sup>25</sup> and found that HCV gt6a has become the second most prevalent subtype on Hainan  
92 Island. Other than HCV gt6a, we also identified 6 novel unassigned gt6 variants, all  
93 of which were discovered in the Li ethnic within Baisha County, Hainan Island<sup>26</sup>.  
94 Baisha County is located in a mountainous area on Hainan Island and most native  
95 residents belong to the Li ethnic minority that is derived from a sub-group of  
96 Austronesians who are descended from the aborigines of the Neolithic Age. In order  
97 to understand whether the novel gt6 unassigned variants were an unusual phenomenon,  
98 we designed this study to understand the epidemiology of HCV infection in the Hainan  
99 Li ethnic community. Such information could help improve strategies for HCV  
100 prevention and control in this county and potentially avoid transmission to other regions  
101 of Hainan Island. Moreover, it could give insight into the evolution of gt6 in the Li  
102 ethnic group.

103

## 104 **Material and methods**

105

### 106 **Survey cohort**

107

108 From July 2014 to October 2015, 1682 volunteers from all 4 Baisha communes

109 and 7 townships aged 11-95 years were invited to participate in this study. The  
110 participants were selected by a random sampling method, including adults who received  
111 health examinations and children in high school who received a thalassemia test. The  
112 inclusion criteria were: 1) residing in the study locale, 2) having the cognitive ability  
113 required to respond to the questionnaire survey and 3) having no prior HCV treatment.  
114 The physicians ensured that individuals were personally interviewed to assure their  
115 complete understanding of the study and the participants provided written informed  
116 consent prior to enrollment. People who refused to provide the detailed information  
117 contained in questionnaire survey forms were excluded from the study. Participants'  
118 socio-demographic features and the presence of risk factors were recorded and blood  
119 samples were collected for an assessment of HCV markers. This study was approved  
120 by the Medical Ethics Committee of Guangzhou Blood Center and conformed to the  
121 ethical guidelines of the 1975 Declaration of Helsinki. All study subjects signed an  
122 informed consent.

123

#### 124 **Anti-HCV assays and HCV RNA detection**

125

126 The 1682 plasma specimens were tested for HCV antibody using two independent  
127 HCV enzyme-linked immunosorbent assays (ELISA) (Shanghai Kehua, China and  
128 Ortho HCV 3.0 ELISA, Ortho-Clinical Diagnostics, Inc, USA). HCV RNA levels  
129 were tested by quantitative real-time polymerase chain reaction (qPCR) (Diagnostic Kit  
130 for Quantification of Hepatitis C virus RNA, Da An Gene Co, Ltd. of Sun Yat-Sen  
131 University, China) and the limit of detection (LOD) was 50 IU/ml. The presence of  
132 HCV antibody was further confirmed by RIBA-HCV BLOT 3.0 (MP Biomedicals Asia  
133 Pacific Pte. Ltd., Singapore). The RIBA-HCV BLOT 3.0 assay utilizes well-defined

134 antigens derived from HCV immunodominant proteins from core, NS3 helicase, NS4  
135 and NS5 regions. Band reactivity was graded by visual calibration against the human  
136 immunoglobulin (IgG) control band (1+) and anti-IgG control band (3+). A sample  
137 was considered positive when at least two HCV bands had a reactivity of  $\geq 1$  or a  
138 single core band had reactivity of  $\geq 2$ . Indeterminate results were defined when only  
139 a single band had a reactivity (except core reactivity of  $\geq 2$ ), and a result was  
140 considered negative when no single HCV antigen line had a reactivity of  $\geq 1$ .

141

### 142 **Sequence determination**

143

144 HCV sequences from the E1 region [nucleotide (nt) 729 to 1322 according to the  
145 numbering of the H77 genome] and NS5B region (nt 8256 to 8641) were amplified  
146 following previously described protocols<sup>22</sup>. Briefly, viral RNA was extracted using a  
147 QIAamp Viral RNA Mini Kit (QIAGEN Inc, Valencia, CA, USA). Reverse  
148 transcription PCR (RT-PCR) was performed by RevertAid First Strand cDNA Synthesis  
149 Kit (Fermentas, CA, USA). Amplification was completed by a nested PCR with E1  
150 or NS5B-specific primers<sup>22</sup>. The amplified products were bi-directionally sequenced  
151 using an Applied Biosystems (ABI) PRISM Big Dye Terminator Cycle Sequencing  
152 Ready Reaction Kit, Version 3.1 (Applied Biosystems, Foster City, CA, USA). The  
153 bi-directional sequences were inspected and assembled using SeqMan to obtain the  
154 consensus sequence. Sequences were aligned using MegAlign. The above software  
155 used is contained in the Lasergene 8.1 Package (DNASTAR Inc, Madison, WI).

156

### 157 **Phylogenetic analysis**

158



159 Phylogenetic analysis was performed using MEGA 7.0. The maximum  
160 likelihood approach (using the General Time Reversible model and Gamma distributed  
161 with Invariant site (G+I) rate) with 500 iterations of bootstrap sampling was performed.  
162 HCV genotype reference sequences were retrieved from the HCV database  
163 (<http://hcv.lanl.gov/content/sequence/HCV/ToolsOutline.html>). Reference sequences  
164 in the phylogenetic analysis were chosen according to consistent criteria<sup>27</sup> and HCV gt6  
165 unassigned sequences were included (detailed information is shown in Supplementary  
166 Table 1). The significance of the tree topology was evaluated using the distance  
167 program (method: p-distance), implemented in Mega 7.0. Bootstrap values of >70%  
168 were considered significant.

169

#### 170 **Statistical analysis**

171

172 The Chi-square test and multivariate logistic regression analysis were applied to  
173 determine the predictors of anti-HCV positivity. Measurement data were expressed  
174 as the mean  $\pm$  standard deviation (SD) and compared using the two-sample t-test, where  
175 appropriate.  $P < 0.05$  was considered statistically significant. All statistical analyses  
176 were performed with SPSS Statistics for Windows, version 19.0 (IBM Corp., Armonk,  
177 New York).

178

#### 179 **Nucleotide sequence accession numbers**

180

181 The nucleotide sequences reported in this study were deposited in Genbank with  
182 the following accession numbers: MH298078-MH298132 and MH298133-MH298190.

183

---

## 184 **Results**

185

### 186 **HCV infection in the Baisha Li minority**

187

188 Samples from the 1682 Baisha Li subjects were tested for HCV antibodies, and  
189 324 were reactive (19.3%) based on the Ortho ELISA assay. Reactive samples were  
190 retested with Kehua ELISA assay, and 249 (14.8%) were reactive with both  
191 independent assays. HCV RNA was detected in 67 of these samples with a mean viral  
192 load of  $6.76 \times 10^4$  IU/ml  $\pm$  9.55 (Table 1). No samples were found that were both anti-  
193 HCV negative and HCV RNA positive (HCV RNA [+]). When compared with the  
194 viral RNA titer ( $3.90 \times 10^5$  IU/ml  $\pm$  6.92) in chronically HCV-infected subjects in China  
195 in our previous study<sup>28</sup>, the HCV RNA titer in the Li minority of Baisha County was  
196 significantly lower ( $t=20.79$ , d.f.=102,  $P<0.01$ , t test).

197

198 We determined that positivity in both the HCV RNA test and 2 independent  
199 ELISAs was sufficient to confirm HCV infection and no further serological testing was  
200 performed for these 67 samples. For the remaining 182 seropositive samples that were  
201 HCV RNA negative [HCV RNA (-)], an anti-HCV confirmatory assay (RIBA) was used  
202 to further assess the presence of HCV antibodies. Using RIBA, 96 of the 182 HCV  
203 ELISA (+) samples were determined to be negative and subsequently excluded from  
204 the anti-HCV positive group (Table 1). Patients testing HCV RNA (+) ( $n=67$ ) or  
205 RIBA positive ( $n=50$ ) were defined as the anti-HCV (+) group. HCV RNA (+)  
206 individuals were considered as having a persistent infection whereas RIBA positive  
207 individuals but who were HCV RNA (-) were defined as spontaneous clearers. There  
208 were 36 RIBA indeterminate samples which were either included or excluded from

209 subsequent analysis to give both liberal and conservative estimates. Using these  
210 criteria, the prevalence of HCV infection in the Baisha Li minority was 7.0-9.1% (n=  
211 [67+50]/1682 or [67+50+36]/1682). Surprisingly, HCV prevalence varied greatly  
212 between subjects  $\geq 60$  years old and  $< 60$  years old (Table 2). Only 5 of the HCV (+)  
213 samples were from individuals  $< 60$  years old. The prevalence of virus infection for  
214 participants aged 60 or more was 16.4-21.6% (n=112/685, or 148/685 including RIBA  
215 indeterminate results), but only 0.5% (n=5/997) for those  $< 60$  years of age.  
216 Furthermore, we found that HCV spontaneous clearance was high (42.7-52.6%,  
217 n=50/117 or 86/153 minus or plus RIBA indeterminate results respectively) across the  
218 entire infected group. In terms of the antigen reactivity of the 86 HCV RNA (-), RIBA  
219 (+) or indeterminate samples, a wide reactivity was observed with core antigen. 69  
220 samples (80.2%) reacted with core antigen, 65 (75.6%) reacted with NS3 antigen (NS3-  
221 1 and NS3-2), 50 (58.1%) with NS4 antigen and 12 (14.0%) with NS5 antigen  
222 (Supplementary Table 2).

223

### 224 **Risk factors for hepatitis C**

225

226 The prevalence of HCV infection did not differ significantly by univariate analysis  
227 based on gender, transfusions, alcohol consumption, piercings, acupuncture, familial  
228 HCV infections or dental procedures, but did vary significantly for tattooing, surgery  
229 and age  $\geq 60$  years, both including and excluding the RIBA indeterminate samples  
230 (Tables 3 and 4). The results obtained by including or excluding RIBA indeterminate  
231 samples from the analysis were similar; only the inclusion group is shown below and  
232 details for the exclusion group are given in table 4. In the anti-HCV (+) group, there  
233 were more persons who had been tattooed and undergone surgery than in the anti-HCV

234 (-) group (OR=2.82, 95%CI 1.78-4.46, P<0.01; OR=4.05, 95% CI 2.18-7.51, P<0.01,  
235 overall chi-square test). We also found that individuals aged  $\geq 60$  years had a  
236 significantly higher prevalence of HCV infection than people aged <60 years (OR=0.02,  
237 95% CI 0.01-0.05, P<0.01, chi-square test) (Table 3). Multivariate logistic regression  
238 analysis revealed that an age  $\geq 60$  years (OR=0.02, 95% CI 0.01-0.05, P<0.01) and  
239 surgery (OR=2.75, 95%CI 1.36-5.57, P<0.01) were the significant predictors of anti-  
240 HCV positivity (Table 3) with the confounders adjusted. In addition, we found that there  
241 was no significant difference between HCV spontaneous clearance and HCV persistent  
242 infection for risk factors including gender and age, as well as history of transfusion,  
243 alcohol consumption, tattooing, surgery, piercings, acupuncture, a family history of  
244 HCV and dental operations (Supplementary Table 3). There was also no significant  
245 difference in the relative prevalence of persistent and recovered HCV infection  
246 according to gender, either excluding or including the RIBA indeterminate results  
247 (supplementary table 4).

248

### 249 **Molecular characterization of Baisha HCV strains**

250

251 In order to determine the HCV genotype/subtype distribution in the study subjects,  
252 we amplified and sequenced the 67 HCV RNA (+) samples. 55 sequences from the  
253 E1 region and 58 from the NS5B region were obtained, while neither E1 nor NS5B  
254 sequences were amplified from 8 samples because of their very low viral loads and  
255 limited plasma volume. For these 8 samples, we tried to amplify the 5'UTR and Core  
256 regions, but only 1 sequence from the 5'UTR region was amplified which was  
257 determined as gt6new by phylogenetic analysis (data not shown). No amplified  
258 products were obtained from sample 114 in the NS5B region and samples 257, 292,

259 429 and 555 in the E1 region despite several attempts, suggesting a lack of primer  
260 specificity for these HCV strains which potentially represent new subtypes. Indeed all  
261 the amplified products in our analysis did not cluster clearly with any of the previously  
262 assigned gt6a-6xd or other unassigned gt6 variants which were collected from some  
263 south Asian countries such as Thailand, Vietnam, Cambodia, Laos and Myanmar<sup>23,29</sup>  
264 (Fig 1). Interestingly, the NS5B Baisha sequences formed four specific groups  
265 (designated I , II, III and IV) which included 6 database sequences collected from  
266 Baisha County of Hainan Island and 1 database sequence obtained from the Guangdong  
267 province of China. 24.1% (14/58) of the Baisha strains clustered within group I , 19.0%  
268 (11/58) clustered with group II , 36% (21/58) clustered with group III and 15.5%  
269 (9/58) clustered with group IV , all with significant bootstrap values. Isolates BSM397,  
270 371 and 396 were distinctly divergent from the above four groups and the gt6 reference,  
271 but their bootstrap values were lower than 70%. It should be noted that sequence  
272 homology between these three isolates and gt6 reference strains ranged between 61-  
273 83%. Full length viral genome sequencing is required to confirm their diversity in a  
274 future study. Although the subtype 6w (EU643834, DQ278892) database sequences  
275 were the closest strains to group I , these sequences cannot be considered as belonging  
276 to 6w because of high variability with the database strains (ranging between 19.4%-  
277 22.9%), which is greater than the recognized within-subtype distance of 15%. Six  
278 isolates (KJ470620, KJ470621, KJ470622, KJ470623, KJ470624, KJ470625;  
279 designated gt6new) that were previously collected from Baisha County clustered with  
280 group II and III. The distance between KJ470623 and samples 9, 41, 46, 22, 239 and  
281 525 was 9.3%-12.4%, which is within the acceptable subtype range. However at least  
282 one full-length (or at least complete open reading frame) is needed to designate a new  
283 subtype according to the HCV nomenclature criteria of the International Committee on

284 Taxonomy of Viruses (ICTV). The distances between other database sequences and  
285 the Baisha sequences were greater than 15%. Even KC844040, the gt6new sequence  
286 which was collected from Guangdong province of China that clustered with group IV,  
287 was 15%-19.3% distant from the group IV Baisha isolates, indicating that the Baisha  
288 sequences are divergent from all other current gt6 sequences. To illustrate the  
289 diversity within the Baisha sequences, strains 282 and 305 were identified from a  
290 couple but clustered with different groups and the sequence divergence was 29%. No  
291 other participants within the study were related according to our questionnaire survey.

292

293 We also analyzed E1 sequences using the same methodology as that for NS5B.  
294 The phylogenetic tree inferred by E1 sequences (Fig 2) had a similar topology to the  
295 NS5B phylogenetic tree. All the E1 sequences in the current study also did not clearly  
296 cluster with any of the currently assigned gt6a-6xd subtypes or with unassigned gt6  
297 variants which were collected from some south Asian countries<sup>23,29</sup>. Similar to the  
298 results with NS5B, the E1 sequences demonstrated a trend and formed four specific  
299 groups (groups I, II, III and IV) although the bootstrap values of groups I and II were  
300 lower than 70%. Full length viral genome sequencing is required to confirm the  
301 diversity of these two groups in a future study. The sequences from the couple (strains  
302 282 and 305) that were divergent in the NS5B region also clustered with different  
303 groups in the E1 region.

304

### 305 **Characterization of highly divergent Baisha HCV viral strains**

306

307 In order to determine whether or not the high sequence divergence observed for  
308 samples 282 and 305 was the consequence of a complex mixture of multiple infections,

309 the E1 amplified product was cloned and 50 clones from each sample were sequenced  
310 and phylogenetically analyzed (Supplementary Fig 1). Strain 409 was cloned and  
311 similarly sequenced as a control. Within each sample tested, the cloned sequences  
312 were closely related, with a sequence homology of 99-100% for sample 282, 95-100%  
313 for 305 and 98-99% for 409. Phylogenetic analysis showed that the cloned sequences  
314 for each isolate clustered together and with the corresponding PCR-derived sequence  
315 (Supplementary Fig 1). When the cloned sequences were included in the E1  
316 phylogenetic analysis, the topology and bootstrap values were similar to those obtained  
317 with the PCR-derived consensus sequences (data not shown). These data confirmed  
318 the genomic divergence of the strains from samples 282 and 305, indicating that this  
319 couple had been independently infected and sexual transmission can be excluded.

320

## 321 **Discussion**

322 In the current study, high HCV prevalence (7.0-9.1%), high spontaneous clearance  
323 rates (42.7-52.6%) and a broad gt6 diversity were observed in Baisha County, which is  
324 remarkably different from that observed in other regions. The specific epidemiology  
325 resulting in such high divergence is unknown but may be indicative of transmission  
326 trends or specific HCV evolution.

327

328 The prevalence of HCV in the Li minority of Baisha County in our study was  
329 between 7.0% and 9.1%, which is higher than previous reports in China (3.2% in 1992<sup>30</sup>,  
330 2.2% in 1997<sup>2</sup>, 0.43% in 2006<sup>3</sup> and 3.0% reported by a more recent study whose  
331 subjects were recruited in Jilin province from 2010 to 2013<sup>1</sup>). One reason for the  
332 varied HCV prevalence between studies was the testing procedures used, i.e. utilizing  
333 one or two ELISA screening assays, and whether a confirmatory assay (RIBA) was

334 applied. For example, the 0.43% positive rate was based on the utilization of two  
335 ELISA screening assays plus RIBA<sup>1</sup> whereas the 3.0% prevalence in Jilin province was  
336 based on two ELISA screening assays only<sup>3</sup>. In our study, after performing two  
337 independent ELISA screening assays, we further employed a highly specific RIBA  
338 assay to exclude false positive reactions from the ELISA screening assays which have  
339 been reported to have up to 35% false-positive results in immunocompetent individuals,  
340 blood donors, the military and the general population<sup>31</sup>. RIBA indeterminate results  
341 indicate that the clinical status of such individuals cannot be determined. Previous  
342 studies showed that a RIBA indeterminate test may be observed in recently infected  
343 individuals who are in the seroconversion phase or in elderly individuals who have  
344 recovered from previous or distant HCV infection<sup>32-34</sup>. Makuria et al<sup>33</sup> reported that  
345 RIBA-indeterminate blood donors were older than chronic HCV carriers. All of the  
346 RIBA indeterminate individuals in our study were  $\geq 60$  years old and there is no other  
347 available RIBA assay to perform a second test. Therefore, we considered these RIBA  
348 indeterminate samples as both recovered and not recovered from infection in our  
349 calculations. Thus, the 7.0 to 9.1% prevalence of HCV infection in the Li minority  
350 group in this remote, rural county is relatively reliable. Another likely reason for  
351 discrepancies between studies is the socio-demographics of the participants within each  
352 cohort, including geography and age. Interestingly, HCV infection in the Baisha Li  
353 minority demonstrates a birth cohort effect, with 95.7-96.7% (n=[112/117]-[148/157])  
354 of HCV infections occurring in individuals  $\geq 60$  years old and the prevalence rate in  
355 this age group was 16.4 to 21.6%. Of the five infected individuals whose age was <60  
356 years old, four sequences (BSY159, BSY517, BSM217 and BSM694) had a scattered  
357 distribution among the elder age sequences (Fig 1 and Fig 2); the fifth was HCV RNA  
358 (-) and no viral sequences were obtained from PCR amplification. Both univariate



359 Chi-square and multiple logistic regression analysis showed that age  $\geq 60$  years old  
360 was strongly associated with anti-HCV positivity. We speculate that perhaps specific  
361 lifestyle behaviors or poor medical practices in the past led to HCV transmission in the  
362 Li ethnic group from Baisha County. Furthermore, surgery was an important risk factor  
363 for HCV infection both by univariate Chi-square and multiple logistic regression  
364 analysis, probably due to the use of unsterile devices, as demonstrated in previous  
365 studies<sup>35,36</sup>. Another important risk factor was tattooing based on the results from the  
366 univariate Chi-square between HCV infection prevalence in tattooed versus non-  
367 tattooed persons (Table 3 and 4). Unsafe practices could have exposed individuals to  
368 contaminated tattooing tools because of a lack of awareness of infection control  
369 practices, lack of sterilization equipment, not purchasing disposable tattoo tools and so  
370 forth. There is evidence that tattooing is a route for HCV transmission<sup>37,38</sup>. The Li  
371 minority of Baisha County has reportedly had less contact with individuals from other  
372 ethnic groups, and tattooing as a custom persists throughout their life, especially in  
373 elderly individuals, as demonstrated by the average age of the persons who were  
374 tattooed (73.33 $\pm$ 14.54 years old) being greater than non-tattooed individuals  
375 (47.77 $\pm$ 22.41 years old) ( $t=98.98$ ,  $P<0.01$ ,  $t$  test, data not shown). However, in terms  
376 of the multivariate analysis, HCV prevalence was not significantly different between  
377 the tattooed and non-tattooed participants. These inconsistent results may be  
378 attributable to the strong association between age and tattooing ( $\chi^2=1.37E2$ ,  $OR=12.89$ ;  
379  $95\%CI = 7.57-21.95$ ,  $P=1.37E-31$ , chi-square test, data not shown). Furthermore,  
380 some reports indicate that the male sex is associated with HCV infection<sup>39-41</sup>. Our  
381 results lack an association between males and HCV infection, which differs from the  
382 literature on this topic. One reason for this discrepancy could be that the prevalence  
383 of tattooing in females was significantly higher than in males (female (11.0%) vs male

384 (3.4%):  $\chi^2=31.50$ ,  $P<0.01$ , data not shown), and that the custom of tattooing leads to  
385 higher HCV prevalence (tattooed (20.0%) vs non-tattooed (8.1%): ( $\chi^2=21.10$ ,  $P<0.01$ ,  
386 data not shown)) in females.

387 The published data reveals that the HCV clearance rate varies in different  
388 populations<sup>16</sup>. HCV spontaneous clearance of 18% occurred in children in a  
389 contemporary US cohort after 6 months of infection<sup>42</sup>. Two recent studies from north  
390 and southwest China involving a limited number of patients with acute HCV infection  
391 reported recovery rates of 28.2%<sup>43</sup> and 30.2%<sup>44</sup>. The Irish Hepatology Research Group  
392 revealed that there was up to 50% recovery among a population of Irish women who  
393 were exposed to HCV-contaminated anti-D immune globulin<sup>45,46</sup>. Therefore, the HCV  
394 spontaneous clearance rate in Baisha Li ethnic group is relatively high (42.7 to 56.2%).  
395 There is a possibility that a lower spontaneous clearance rate is masked by fast  
396 progression to disease and HCV-related mortality thus increasing the observed  
397 clearance rate among elderly individuals. Compared to a population of Chinese  
398 chronically HCV-infected subjects, the significantly lower viral load observed in the Li  
399 population provides indirect evidence to support a higher spontaneous clearance rate  
400 which would correlate with a previous study<sup>39</sup>. No significant association was found  
401 with gender in our study, although it has been well-established that women are more  
402 likely to clear HCV than males<sup>41,43,44,47</sup>. However, three other studies, including U.S.  
403 injection drug users, U.S. blood donors and the Italian general population, also found  
404 no significant association of HCV clearance with gender<sup>48</sup>. One possible interpretation  
405 for these differing results is that the association with gender may be confounded by  
406 tattooing, the probable route of HCV transmission in Baisha County, which is practiced  
407 significantly more in females than males as discussed above. In addition the genotype  
408 6 strains from our study may differ in this respect compared to the genotypes that

409 predominate in other studies that have evidence linking gender and clearance rates.  
410 According to reports, the different HCV spontaneous clearance rates might be related  
411 to host and/or viral factors. Host factors include humoral and cellular immune  
412 responses. In this study, only the humoral response was examined, particularly in the  
413 reactivity to individual viral proteins. From the data collected here, the anti-core  
414 reactivity seemed to persist, while the anti-NS5 reactivity appeared to be short-lived.  
415 It has been reported that reactivity to core protein may represent either non-specific  
416 reactivity or resolved infection with antibodies to other HCV proteins that are no longer  
417 reactive. In addition, NS3 and core antibodies are the last two remaining reactive  
418 proteins in cases of seroconversion<sup>49</sup>.

419 There is a geographical distribution of the HCV genotypes in China. For  
420 example, subtypes 1b and 2a were predominant in north China<sup>50</sup>, subtype 6a was first  
421 found in Guangdong province and transmitted to other regions<sup>40</sup>, and subtypes 3a and  
422 3b were found in south China<sup>51</sup>. Subtype 6a was initially detected in Guangdong  
423 province and has become the second most prevalence subtype since 2008<sup>22</sup>.  
424 Importantly, subtype 6a has rapidly spread from Guangdong to other regions in China  
425 over the past 10 years<sup>22</sup>. The distribution of HCV genotypes on Hainan Island which  
426 seceded from Guangdong province in 1988  
427 ([http://www.npc.gov.cn/npc/dbdhhy/content\\_5602.htm](http://www.npc.gov.cn/npc/dbdhhy/content_5602.htm)), is still largely unknown  
428 especially in the Li ethnic minority. Fig 3 shows the HCV genotype distribution of  
429 previous studies on Hainan Island from 2009 to 2016; subtype 1b has gradually  
430 decreased, while subtype 6a significantly increased from no detected cases to 35% of  
431 cases in the Han ethnic population<sup>24,26,52,53</sup>. Only 6 isolates were from the Li minority,  
432 and interestingly, they were novel gt6 variants<sup>20</sup>. According to the above results, we  
433 speculate that there is a different HCV genotype distribution between the Li minority

434 and Han ethnic groups on Hainan Island. We found that novel HCV gt6 predominate  
435 in the Li minority group of Baisha County according to phylogenetic trees constructed  
436 from NS5B and E1 sequences. The predominant gt6 strains were diverse and did not  
437 cluster into any assigned subtypes from other regions in China, including the rest of  
438 Hainan island<sup>24,52,53</sup>. The reason for this discrepancy could be the distinct  
439 geographical location of Li and Han populations or a lack of interaction between the  
440 groups. In previous studies, samples were collected from the Han population of  
441 Haikou city, a modern city. Samples in our study were collected from Baisha County,  
442 a remote community, most of whom are of the Li minority. We concluded that the  
443 broad diversity of gt6 is a likely characteristic of HCV infection in the Li minority group  
444 of Baisha County. One possible biological interpretation for this tree topology (Fig 1  
445 and Fig 2) is that most of the variants have evolved over time rather than a recent  
446 introduction into Baisha County. Further support for this hypothesis is that most of  
447 the HCV variants were isolated from elderly ( $\geq 60$  years) individuals with limited  
448 exposure to HCV transmission routes outside their own community. A similar  
449 endemic spread seems to have occurred in Ghana, west Africa with gt2<sup>54</sup>. However,  
450 there is no evolutionary data to support this hypothesis in our current analysis.  
451 Another possible reason is that some isolates were transmitted from Guangdong  
452 province, as KC844040 was the closet sequence from group IV. To test these two  
453 hypotheses, comprehensive evolutionary and diversity analyses will be required in a  
454 future study. Moreover, full-length genome sequencing will be required to provide  
455 more detailed sequence information and meet the requirement of new subtype  
456 designation according to the HCV nomenclature criteria of ICTV<sup>27</sup>.

457

458 In summary, HCV is common in a rural Li ethnic minority community among its

459 older residents ( $\geq 60$  years), demonstrating a birth cohort effect. Moreover, the anti-  
460 HCV confirmatory assay revealed a high rate of spontaneous HCV clearance and  
461 infection in the Baisha Li minority group. HCV infection within the community is  
462 characterized by the predominance of gt6, with broad genetic diversity that potentially  
463 contains new subtypes for this genotype.

464

#### 465 **Conflict of interest**

466 All authors declare that they have no competing interest.

467

#### 468 **Funding:**

469 This research was supported by the grants from National Natural Science Foundation  
470 of China (No.81772208), ECML and JM were supported by funding from the UK  
471 Medical Research Council (MC\_UU\_12014/1).

472

473 **Author Contributions:** Ru Xu drafted the main manuscript. Yongjuan Yu, Min Wang,  
474 Jieting Huang, Xi Tang, Qiao Liao, Dandan Song and Zhengang Shan contributed to  
475 donor recruitment and data collection. Ru Xu and Ke Huang performed statistical  
476 analysis and the interpretation of the data. E. Carol McWilliam Leitch, Chengyao Li  
477 and John Mclauchlan revised the manuscript. Xia Rong conceived the study and  
478 approved the manuscript. All authors reviewed and approved the final version of the  
479 manuscript.

480

#### 481 **References**

482

- 483 1. Fu-qiang CY-sLLC. A sero-epidemiological study on hepatitis C in China. Chin J  
484 Epidemiol 2011;32:888-91.

- 
- 485 2. Kang LY SY, Hao LJ, Cao XY, Pan QC. Studies on epidemiology of the population  
486 with HCV and HEV infections and their epidemic factors in China. Chinese Journal of  
487 Infectious Diseases 1997;15(2):71-5.
- 488 3. Zhang Q, Qi W, Wang X, et al. Epidemiology of Hepatitis B and Hepatitis C  
489 Infections and Benefits of Programs for Hepatitis Prevention in Northeastern China: A  
490 Cross-Sectional Study. Clin Infect Dis 2015.
- 491 4. Feld JJ, Kowdley KV, Coakley E, et al. Treatment of HCV with ABT-450/r-  
492 ombitasvir and dasabuvir with ribavirin. The New England journal of medicine  
493 2014;370:1594-603.
- 494 5. Fried MW, Buti M, Dore GJ, et al. Once-daily simeprevir (TMC435) with  
495 pegylated interferon and ribavirin in treatment-naive genotype 1 hepatitis C: the  
496 randomized PILLAR study. Hepatology 2013;58:1918-29.
- 497 6. Jacobson IM, Gordon SC, Kowdley KV, et al. Sofosbuvir for hepatitis C genotype  
498 2 or 3 in patients without treatment options. The New England journal of medicine  
499 2013;368:1867-77.
- 500 7. Jacobson IM, McHutchison JG, Dusheiko G, et al. Telaprevir for previously  
501 untreated chronic hepatitis C virus infection. The New England journal of medicine  
502 2011;364:2405-16.
- 503 8. Martin NK, Vickerman P, Grebely J, et al. Hepatitis C virus treatment for  
504 prevention among people who inject drugs: Modeling treatment scale-up in the age of  
505 direct-acting antivirals. Hepatology 2013;58:1598-609.
- 506 9. Dong Y1 QC, Xia X, Wang J, Zhang H, Zhang X, Xu J. Hepatitis B virus and  
507 hepatitis C virus infection among HIV-1-infected injection drug users in Dali, China:  
508 prevalence and infection status in a cross-sectional study. Arch Virol 2015;160:929-36.
- 509 10. van de Laar TJ MG, Prins M, Danta M. Acute hepatitis C in HIVinfected men who

- 510 have sex with men: an emerging sexually transmitted infection. *AIDS* 2010;24:1799–  
511 812.
- 512 11. Zhang M, Sun XD, Mark SD, et al. Hepatitis C virus infection, Linxian, China.  
513 *Emerg Infect Dis* 2005;11:17-21.
- 514 12. Delarocque-Astagneau E, Pillonel J, De Valk H, Perra A, Laperche S, Desenclos  
515 JC. An incident case-control study of modes of hepatitis C virus transmission in France.  
516 *Ann Epidemiol* 2007;17:755-62.
- 517 13. Esteban JI, Sauleda S, Quer J. The changing epidemiology of hepatitis C virus  
518 infection in Europe. *J Hepatol* 2008;48:148-62.
- 519 14. Santantonio T ME, Ferrari C. Risk factors and outcome among a large patient  
520 cohort with community-  
521 acquired acute hepatitis C in Italy. *Clin Infect Dis* 2006;43:1154-9.
- 522 15. Sohn HS, Kim JR, Ryu SY, et al. Risk Factors for Hepatitis C Virus (HCV)  
523 Infection in Areas with a High Prevalence of HCV in the Republic of Korea in 2013.  
524 *Gut Liver* 2016;10:126-32.
- 525 16. Micallef JM, Kaldor JM, Dore GJ. Spontaneous viral clearance following acute  
526 hepatitis C infection: a systematic review of longitudinal studies. *J Viral Hepat*  
527 2006;13:34-41.
- 528 17. Strasak AM, Kim AY, Lauer GM, et al. Antibody dynamics and spontaneous viral  
529 clearance in patients with acute hepatitis C infection in Rio de Janeiro, Brazil. *BMC*  
530 *Infect Dis* 2011;11:15.
- 531 18. Tobler LH, Bahrami SH, Kaidarova Z, et al. A case-control study of factors  
532 associated with resolution of hepatitis C viremia in former blood donors (CME).  
533 *Transfusion* 2010;50:1513-23.
- 534 19. Maheshwari A, Ray S, Thuluvath PJ. Acute hepatitis C. *Lancet* 2008;372:321-32.

- 535 20. Gower E, Estes C, Blach S, Razavi-Shearer K, Razavi H. Global epidemiology and  
536 genotype distribution of the hepatitis C virus infection. *J Hepatol* 2014;61:S45-57.
- 537 21. Fu Y, Qin W, Cao H, et al. HCV 6a prevalence in Guangdong province had the  
538 origin from Vietnam and recent dissemination to other regions of China:  
539 phylogeographic analyses. *PLoS One* 2012;7:e28006.
- 540 22. Rong X, Xu R, Xiong H, et al. Increased prevalence of hepatitis C virus subtype  
541 6a in China: a comparison between 2004-2007 and 2008-2011. *Arch Virol*  
542 2014;159:3231-7.
- 543 23. Wan Z, Chen Q, Chen X, et al. HCV Diversity among Chinese and Burmese IDUs  
544 in Dehong, Yunnan, China. *PLoS One* 2016;11:e0163062.
- 545 24. Wu T, Xiong L, Wang F, et al. A Unique Pattern of HCV Genotype Distribution on  
546 Hainan Island in China Revealed by Evolutionary Analysis. *Cell Physiol Biochem*  
547 2016;39:316-30.
- 548 25. Huang K, Chen J, Xu R, et al. Molecular evolution of hepatitis C virus in China: A  
549 nationwide study. *Virology* 2018;516:210-8.
- 550 26. An Y, Wu T, Wang M, et al. Conservation in China of a novel group of HCV  
551 variants dating to six centuries ago. *Virology* 2014;464-465:21-5.
- 552 27. Smith DB, Bukh J, Kuiken C, et al. Expanded classification of hepatitis C virus  
553 into 7 genotypes and 67 subtypes: updated criteria and genotype assignment web  
554 resource. *Hepatology* 2014;59:318-27.
- 555 28. Rong X, Lu L, Wang J, et al. Correlation of viral loads with HCV genotypes: higher  
556 levels of virus were revealed among blood donors infected with 6a strains. *PLoS One*  
557 2012;7:e52467.
- 558 29. Lu L, Wu T, Xiong L, Li C, Nguyen MH, Murphy DG. Analysis of HCV-6 isolates  
559 among Asian-born immigrants in North America reveals their high genetic diversity



- 560 and a new subtype. *Virology* 2016;492:25-31.
- 561 30. Xia GL, Liu, C. B., Cao, H. L., et al. Prevalence of hepatitis B and C virus  
562 infections in the general Chinese population. Results from a nationwide cross-sectional  
563 seroepidemiologic study of hepatitis A, B, C, D, and E virus infections in China, 1992.  
564 *Int Hepatol Commun* 1996;5:62-37.
- 565 31. Alter MJ, Kuhnert WL, Finelli L, Centers for Disease C, Prevention. Guidelines  
566 for laboratory testing and result reporting of antibody to hepatitis C virus. Centers for  
567 Disease Control and Prevention. *MMWR Recommendations and reports : Morbidity  
568 and mortality weekly report Recommendations and reports* 2003;52:1-13, 5; quiz CE1-  
569 4.
- 570 32. Bes M, Esteban JI, Casamitjana N, et al. Hepatitis C virus (HCV)-specific T-cell  
571 responses among recombinant immunoblot assay-3-indeterminate blood donors: a  
572 confirmatory evidence of HCV exposure. *Transfusion* 2009;49:1296-305.
- 573 33. Makuria AT, Raghuraman S, Burbelo PD, et al. The clinical relevance of persistent  
574 recombinant immunoblot assay-indeterminate reactions: insights into the natural  
575 history of hepatitis C virus infection and implications for donor counseling. *Transfusion*  
576 2012;52:1940-8.
- 577 34. Ponde RA. Enzyme-linked immunosorbent/chemiluminescence assays,  
578 recombinant immunoblot assays and nucleic acid tests in the diagnosis of HCV  
579 infection. *European journal of clinical microbiology & infectious diseases : official  
580 publication of the European Society of Clinical Microbiology* 2013;32:985-8.
- 581 35. Moosavy SH, Davoodian P, Nazarnezhad MA, Nejatizadeh A, Eftekhar E,  
582 Mahboobi H. Epidemiology, transmission, diagnosis, and outcome of Hepatitis C virus  
583 infection. *Electron Physician* 2017;9:5646-56.
- 584 36. Zamani F, Sohrabi M, Poustchi H, et al. Prevalence and risk factors of hepatitis C

- 585 virus infection in amol city, north of iran: a population-based study (2008-2011). *Hepat*  
586 *Mon* 2013;13:e13313.
- 587 37. Hellard ME, Aitken CK, Hocking JS. Tattooing in prisons--not such a pretty picture.  
588 *Am J Infect Control* 2007;35:477-80.
- 589 38. Jafari S CR, Baharlou S, Etminan M, Buxton J. Tattooing and the risk of  
590 transmission of hepatitis C: a systematic review and meta-analysis. *Int J Infect Dis*  
591 2010;14:e928±40.
- 592 39. Fabricio-Silva GM, Poschetzky BS, de Mello Perez R, Dos Santos RC, Cavalini  
593 LT, Porto LC. Association of cytokine gene polymorphisms with hepatitis C virus  
594 infection in a population from Rio de Janeiro, Brazil. *Hepat Med* 2015;7:71-9.
- 595 40. Fu Y, Wang Y, Xia W, et al. New trends of HCV infection in China revealed by  
596 genetic analysis of viral sequences determined from first-time volunteer blood donors.  
597 *J Viral Hepat* 2011;18:42-52.
- 598 41. Inoue G, Horiike N, Michitaka K, Onji M. Hepatitis C virus clearance is prominent  
599 in women in an endemic area. *J Gastroenterol Hepatol* 2000;15:1054-8.
- 600 42. Wang CC, Krantz E, Klarquist J, et al. Acute hepatitis C in a contemporary US  
601 cohort: modes of acquisition and factors influencing viral clearance. *The Journal of*  
602 *infectious diseases* 2007;196:1474-82.
- 603 43. Xu CJ, Zhang CP, Luo BF, et al. Prevalence and characterization of hepatitis B and  
604 C virus infections in a needle-sharing population in Northern China. *BMC Public*  
605 *Health* 2015;15:460.
- 606 44. Liu S, Cheng M, Mu M, Yang Q. [Natural clearance of hepatitis C virus in 96  
607 patients with infection acquired by blood transfusion from a single donor in Guizhou].  
608 *Zhonghua Gan Zang Bing Za Zhi* 2014;22:251-4.
- 609 45. Kuniholm MH, Kovacs A, Gao X, et al. Specific human leukocyte antigen class I

- 610 and II alleles associated with hepatitis C virus viremia. *Hepatology* 2010;51:1514-22.
- 611 46. Kenny-Walsh E. Clinical outcomes after hepatitis C infection from contaminated  
612 anti-D immune globulin. Irish Hepatology Research Group. *The New England journal*  
613 *of medicine* 1999;340:1228-33.
- 614 47. Bakr I, Rekacewicz C, El Hosseiny M, et al. Higher clearance of hepatitis C virus  
615 infection in females compared with males. *Gut* 2006;55:1183-7.
- 616 48. Murphy EL, Fang J, Tu Y, et al. Hepatitis C virus prevalence and clearance among  
617 US blood donors, 2006-2007: associations with birth cohort, multiple pregnancies, and  
618 body mass index. *The Journal of infectious diseases* 2010;202:576-84.
- 619 49. Rafik M, Bakr S, Soliman D, et al. Characterization of differential antibody  
620 production against hepatitis C virus in different HCV infection status. *Virol J*  
621 2016;13:116.
- 622 50. Lu L, Nakano T, He Y, Fu Y, Hagedorn CH, Robertson BH. Hepatitis C virus  
623 genotype distribution in China: predominance of closely related subtype 1b isolates and  
624 existence of new genotype 6 variants. *J Med Virol* 2005;75:538-49.
- 625 51. Wu Z, Cui L, Zhao W, et al. Molecular epidemiology of hepatitis C infections in  
626 Ningxia, China: genotype, phylogeny and mutation analysis. *Virol J* 2016;13:172.
- 627 52. Junxia Zhou JX. The research of Hepetitis C Virus genotyping in Haikou region.  
628 *Modern preventive medicine* 2012;39:4507-10.
- 629 53. Tao wu JW, Xiaozhen Xu, Lu Xiong, Jiao Wang, Fuli Wang, Feng Lin. Genotype  
630 distribution of Hepetitis C Virus among 163 isolates from ethic Han in Hainan Province.  
631 *China tropical medicine* 2016;16:873-6.
- 632 54. Candotti D, Temple J, Sarkodie F, Allain JP. Frequent recovery and broad genotype  
633 2 diversity characterize hepatitis C virus infection in Ghana, West Africa. *J Virol*  
634 2003;77:7914-23.

---

636 **Supplementary material**

637

638 **Supplementary Fig 1 Phylogenetic tree of isolates 282, 305 and 409 E1 sequences.**

639 Solid yellow circles, solid green circles and solid red circles represent clones of the 282  
640 isolate, 305 isolate and 409 isolate respectively. A maximum Likelihood tree was  
641 constructed using Mega 7.

642

643 **Supplementary table 1 Supplementary table 1 Detailed information of reference  
644 sequences in the phylogenetic analysis**

645

646 **Supplementary table 2 Details of RIBA reactivity to core, NS3, NS4 and NS5  
647 antigens**

648

649 **Supplementary table 3 Socio-demographics of subjects according to HCV  
650 Persistent infection and HCV Spontaneous clearance**

651 **Figure legends:**

652

653 **Fig 1 The estimated maximum likelihood phylogeny for 58 Baisha Li minority**

654 **NS5B sequences.** Sequences determined in this study are shown in solid green circle.

655 Solid yellow circle represents unassigned HCV gt6 references which were collected

656 from China and unassigned reference sequences from other south Asian countries

657 except China are indicated by solid red circle. Other reference sequences which were

658 named by subtype and accession number represent assigned HCV gt1-gt7.

659

660 **Fig 2 The estimated maximum likelihood phylogeny for 55 Baisha Li minority E1**

661 **sequences.** Sequences determined in this study are shown in solid green circle. Solid

662 yellow circles represent unassigned HCV gt6 references which were collected from

663 China and unassigned reference sequences from other south Asian countries except

664 China are indicated by solid red circle. Other reference sequences which were named

665 by subtype and accession number represent assigned HCV gt1-7.

666

667 **Fig 3 The location of Hainan island in China and the HCV genotype distribution**

668 **in previous studies.** The Li minority population reside in the central southern

669 mountainous region (colors) whereas the Han ethnic population occupy the northern

670 coastal region (blank).

671 Table 1 Confirmatory results of HCV antibody by RIBA

	Both Ortho and Kehua ELISA positive (n=249)				ELISA negative,
	RNA positive	RNA negative, ELISA positive (n=182)			RNA negative
		RIBA positive	RIBA indeterminate	RIBA negative	
Total Samples = 1682	67	50	36	96	1433
HCV RNA titer (IU/ml)	6.76x10 <sup>4</sup> (± 9.55)	-	-	-	-
Anti-HCV positive		117 <sup>§</sup>	153 <sup>§</sup>		
Prevalence of HCV infection		7.0 (117/1682) <sup>§</sup>	9.1% (153/1682) <sup>§</sup>		
Prevalence of HCV spontaneous clearance		42.7 (50/117) <sup>§</sup>	56.2% (86/153) <sup>§</sup>		

672 <sup>§</sup> Lower value refers to the exclusion of RIBA indeterminate data and higher value to the  
673 inclusion of RIBA indeterminate data.

674

675 Table 2 Age-related HCV infection in the Baisha Li minority community

ages	Anti-HCV (-) (%)	Anti-HCV (+) (%)	total
10-20y	293(99.3)	2(0.7)	295
20-30y	148(100)	0(0)	148
30-40y	128(100)	0(0)	128
40-50y	225(99.1)	2(0.9)	227
50-60y	198(99.5)	1 (0.5)	199
60-70y	189-190 (89.6–90.0)	21-22 (10.0-10.4)	211
70-80y	262–284 (76.8–83.3)	57-79(16.7-23.2)	341
80-95y	86–99 (64.7–74.4)	34-47(25.6-35.3)	133
total	1529–1565 (90.9–93.0)	117-153(7.0-9.09)	1682

676

677 Table 3 Socio-demographics of subjects according to anti-HCV positivity including  
 678 RIBA indeterminate results

	Univariate chi-square analysis		Multivariate analysis			
	Anti-HCV(+) <sup>§</sup> n=153	Anti-HCV(-) n=1529	OR (95%CI)	P-value	OR (95%CI)	P-value
<b>Gender</b>						
Male	57(37.3)	597(39.0)				
Female	96(62.7)	932(61.0)	0.93 (0.28-1.28)	0.78	0.73 (0.47-1.13)	0.16
<b>Age</b>						
<60y	5(3.3)	982(64.2)				
>60y	148(96.7)	547(35.8)	0.02 (0.01-0.05)	4.11E-54	0.02 (0.01-0.05)	6.66E-18
<b>Transfusion</b>						
Yes	4(2.6)	25(1.6)				
No	149(97.4)	1504(98.4)	1.62 (0.56-4.73)	0.51	1.17 (0.34-3.98)	0.81
<b>Alcohol Consumption</b>						
Yes	61(39.9)	614(40.2)				
No	92(60.1)	915(59.8)	0.99 (0.7-1.39)	0.95	0.97 (0.67-1.43)	0.89
<b>Tattooing</b>						
Yes	27(17.6)	108(7.1)				
No	126(82.4)	1421(92.9)	2.82 (1.78-4.46)	1.35E-06	1.01 (0.57-1.81)	0.96
<b>Surgery</b>						
Yes	15(9.8)	40 (2.6)				
No	138 (90.2)	1489 (97.4)	4.05 (2.18-7.51)	1.88E-06	2.75 (1.36-5.57)	0.005
<b>Piercings</b>						
Yes	44(28.8)	382(25)				
No	109(71.2)	1147(75)	1.21 (0.84-1.75)	0.31	1.46 (0.91-2.35)	0.12
<b>Acupuncture</b>						
Yes	4(1.6)	24(2.6)				
No	149(98.4)	1505(97.4)	1.68 (0.58-4.92)	0.34	0.54 (0.17-1.72)	0.30
<b>Family history of HCV</b>						
Yes	2(1.3)	7 (0.5)				
No	151(98.7)	1522 (99.5)	2.88 (0.59-13.99)	0.17	1.28(0.23-7.23)	0.78
<b>Dental procedures</b>						
Yes	4(2.6)	39(2.6)				
No	149(97.4)	1490(97.4)	0.98 (0.36-2.9)	0.96	0.70(0.22-2.24)	0.55

679

680 §: HCV RNA positive, RIBA positive and RIBA indeterminate samples  
 681 constituted the anti-HCV (+) group

682

683 Table 4 Socio-demographics of subjects according to anti-HCV positivity excluding  
 684 RIBA indeterminate results

	Univariate Chi-square analysis				Multivariate analysis	
	Anti-HCV(+) <sup>#</sup> n=117	Anti-HCV(-) n=1565	OR (95%CI)	P-value	OR (95%CI)	P-value
<b>Gender</b>						
Male	44(37.6)	610(39.0)	0.94 (0.64-1.39)	0.77	0.69 (0.42-1.12)	0.14
Female	73(62.4)	955(61.0)				
<b>Age</b>						
<60y	5(4.3)	992(62.4)	0.03 (0.01-0.06)	<b>3.82E-36</b>	0.03 (0.01-0.06)	<b>2.39E-15</b>
≥60y	112(95.7)	573(37.6)				
<b>Transfusion</b>						
Yes	3(2.6)	26(1.7)	1.56 (0.46-5.22)	0.47	1.24 (0.32-4.82)	0.76
No	114(97.4)	1539(98.3)				
<b>Alcohol consumption</b>						
Yes	46(39.3)	629(40.2)	0.96 (0.66-1.42)	0.85	0.96 (0.63-1.47)	0.86
No	71(60.7)	936(59.8)				
<b>Tattooing</b>						
Yes	20(17.1)	115(7.3)	<b>2.60 (1.55-4.36)</b>	<b>1.82E-04</b>	1.01 (0.57-1.81)	0.96
No	97(82.9)	1450(92.7)				
<b>Surgery</b>						
Yes	11(9.4)	44 (2.8)	<b>3.59 (1.80-7.15)</b>	<b>1.11E-04</b>	<b>2.36 (1.09-5.10)</b>	<b>0.03</b>
No	106 (90.6)	1521 (97.2)				
<b>Piercings</b>						
Yes	35(29.9)	391(25)	1.28 (0.85-1.94)	0.24	1.63 (0.96-2.75)	0.07
No	82(70.1)	1174(75)				
<b>Acupuncture</b>						
Yes	4(3.4)	24(1.5)	2.27 (0.78-6.67)	0.12	1.08 (0.31-3.70)	0.91
No	113(96.6)	1541(98.5)				
<b>Family history of HCV</b>						
Yes	2(1.7)	7(0.4)	3.87 (0.80-18.85)	0.07	1.82(0.32-10.27)	0.49
No	115(98.3)	1558(99.6)				
<b>Dental procedures</b>						
Yes	2(1.7)	41(2.6)	0.65 (0.15-2.71)	0.55	0.37(0.08-1.76)	0.21
No	115(98.3)	1524(97.4)				

685 §: HCV RNA positive and RIBA positive samples only constituted the anti-HCV  
 686 (+) group

687



688 **Supplementary Table 4**

	Male n (%)	Female n (%)	P Value*	RIBA indeterminate results
HCV(+)	57 (8.7)	96 (9.3)	0.78	Included
HCV(-)	597 (91.3)	932 (90.7)		
Total	654	1028		
HCV(+)	44 (6.7)	73 (7.1)	0.77	Excluded
HCV(-)	610 (93.3)	955 (92.9)		
Total	654	1028		
Spontaneous clearance	29 (50.9)	57 (59.4)	0.31	Included
Persistent infection	28 (49.1)	39 (40.6)		
Total	57	96		
Spontaneous clearance	16 (36.4)	34 (46.6)	0.28	Excluded
Persistent infection	28 (63.6)	39 (53.4)		
Total	44	73		

689 \* Statistical analysis was performed with a Chi-square test.

690

691