JVI Accepts, published online ahead of print on 3 September 2014
J. Virol. doi:10.1128/JVI.01594-14

Title: Recent Population Expansions of Hepatitis B Virus in the United States

Short title: Selective HBV expansion among cases of acute hepatitis B virus infection in USA

Authors: Sumathi Ramachandran*, Michael A. Purdy, Guo-liang Xia, David S. Campo, Zoya E. Dimitrova, Eyasu H. Teshale, Chong Gee Teo, Yury E. Khudyakov

Division of Viral Hepatitis, Centers for Disease Control and Prevention, Atlanta, GA 30333, USA

Correspondence: *Corresponding author email: dcq6@cdc.gov; Tel: 4046391403; Fax: 4046391563

Footnote Page

Disclaimer: This information is distributed solely for the purpose of pre dissemination peer review under applicable information quality guidelines. It has not been formally disseminated by the Centers for Disease Control and Prevention/Agency for Toxic Substances and Disease Registry. It does not represent and should not be construed to represent any agency determination or policy. No conflict of interest exists for any authors or financial funding.

Competing financial interests: The authors declare no competing financial interests.

Word count: Text - 3085; Abstract - 243

Abbreviations: Hepatitis B virus (HBV), whole-genome (WG), men who have sex with men (MSM), injecting drug use (IDU)

Key words: disease transmission, infectious; sequence analysis, DNA; homosexuality, male; substance abuse, intravenous; phylodynamics; risk factors

ABSTRACT

The recent epidemic history of hepatitis B virus (HBV) infections in the United States is complex, as indicated by current disparity in HBV genotype distribution between acute and chronic hepatitis B cases and rapid decline in hepatitis B incidence since the 1990s. We report temporal changes in genetic composition of the HBV population using whole-genome sequences (n=179) from acute hepatitis B cases (n=1206) identified through the Sentinel County Surveillance for Acute Hepatitis (1998-2006). HBV belonged mainly to subtypes A2 (75%) and D3 (18%), with times of their most recent common ancestors being, respectively, 1979 and 1987, respectively. A2 underwent rapid population expansions in *ca.* 1995 and *ca.* 2002, coinciding with transient rises in acute hepatitis B notification rates among adults; D3 underwent expansion in *ca.* 1998. A2 strains from cases identified after 2002, compared to those before 2002, tended to cluster phylogenetically, indicating selective expansion of specific strains, and were significantly reduced in genetic diversity ($p = 0.001$) and frequency of drug-resistance mutations ($p = 0.001$). The expansion of genetically close HBV A2 strains was associated with risk of infection among male homosexuals ($p = 0.03$). Incident HBV strains circulating in the US were recent in origin, and restricted in genetic diversity. Disparate transmission dynamics among phylogenetic lineages affected the genetic composition of HBV populations and their capacity to maintain drug-resistance mutations. The tendency of selectively expanding HBV strains to be transmitted among male homosexuals highlights the need to improve hepatitis B vaccination coverage among at-risk adults.

55 **IMPORTANCE**

56 *Hepatitis B virus (HBV) remains an important cause of acute and chronic liver disease*
57 *globally, and in the United States. Genetic analysis of* ***HBV whole genomes from cases of acute***
58 ***hepatitis B identified from 1998-2006 in the United States*** *showed dominance of genotype A2*
59 *(75%), followed by D3 (18%). Strains of both subtypes were recent in origin and underwent*
60 *rapid population expansions from 1995-2000, indicating increase in transmission rate for*
61 *certain HBV strains during a period of decline in the reported incidence of acute hepatitis B in*
62 *the US. HBV A2 strains from a particular cluster that experienced the most recent population*
63 *expansion were more commonly detected among men who have sex with men. Vaccination needs*
64 *to be stepped up to protect persons who remain at risk of HBV infection.*

65

66

INTRODUCTION

Hepatitis B virus (HBV) is one of the major causative agents of liver disease, including acute and chronic hepatitis, cirrhosis and hepatocellular carcinoma. Infection with HBV is common, with an estimated prevalence of 240 million infected persons globally (1). In the United States (US), it is estimated that 1.4 million people are infected with HBV (2) and 43,000 new infections occur annually (3). The estimated incidence rate of acute HBV infection has declined from 13.5 cases per 100,000 population in 1987 to 2.8 cases per 100,000 population in 2002 (4, 5), which is attributable in part to nationwide implementation of vaccination against hepatitis B. However, this decline was not uniform: between 1999 and 2002, incidences were observed to increase transiently among adults in certain age groups (5). Despite the availability of hepatitis B vaccine and the implementation of comprehensive national guidelines, immunization coverage remains suboptimal in certain demographic and risk groups (6-9).

HBV is genetically diverse, and its strains are classified into 8 major genotypes (A to H) and numerous subtypes (10). The Sentinel Counties Study of Acute Viral Hepatitis was a population-based study that enrolled acute viral hepatitis patients from 6 city/county health departments in the US from 1982 through 2006 (11, 12). We recently showed that 75% of acute hepatitis B cases from 1998-2006 were infected by HBV genotype A2 and 17% by genotype D (13). This distribution contrasts with findings from a study of persons in the US with chronic hepatitis B showing that only 37.4% were infected by genotype A2, and 10.4% by genotype D (14). Strains identified among the acute cases appeared to be of limited genetic diversity, with many strains sharing identical sequences (13). These observations might, however, have reflected usage of the relatively conserved *S* gene as basis of HBV strain comparison. We and others have shown that analysis of whole-genome (WG) sequences of HBV confers significantly

finer resolution to the study of viral diversity and better assessments of HBV evolution during natural infection and under selection pressure from antiviral drugs (15-17).

The disparity of genotype distribution between acute (13) and chronic (14) cases of HBV infection reflects complexity of epidemiological factors associated with patterns of HBV introduction, transmission and maintenance in the US. The disproportionately high representation of genotypes A2 and D in incident HBV infections imply greater efficacy of transmission from chronic infections with these genotypes than with other HBV genotypes, or more frequent transmission of these 2 genotypes from acutely infected persons. Taken together with the observation of a substantial decline in incident HBV infection, the genotypic disparity between acute and chronic hepatitis B indicates a dynamic epidemic history of HBV infections over the last 2 decades. To gain insight into the evolutionary history and molecular epidemiology of incident HBV infection in the US, we characterized HBV WG sequences from cases of acute HBV infection.

MATERIALS AND METHODS

In the Sentinel Counties Study (11), a case of acute hepatitis B was defined as having history of sudden onset of signs and symptoms consistent with hepatitis together with serological detection of immunoglobulin M antibody to hepatitis B core antigen, or newly detected hepatitis B surface antigen within the window period of 60 days. The cases were treatment-naive at the time of sampling. The 6 Sentinel sites were: Jefferson County, Alabama; Pinellas County, Florida; Pierce County, Tacoma, Washington; Multnomah County, Oregon; San Francisco, California; and Denver County, Denver, Colorado. The study was approved by institutional review boards of the Centers for Disease Control and Prevention (CDC). For the current study,

only sera originating from cases enrolled between 1998 through to 2006 that contained sufficient HBV titer to allow for WG amplification and sequencing were included.

HBV WG amplification, sequencing and analyses

HBV WG sequences were amplified using 2 rounds of PCR and sequenced, as previously described (15). Detection sensitivity of this approach is $5x10^2$ IU/ml, using the 3rd World Health Organization International standard for HBV DNA. HBV genotyping was performed by nucleotide (nt) sequencing of a 435–base pair DNA segment amplified from the HBV *S* gene (from nt position 222 to 656 of HBV genome). Phylogenetic trees were constructed using a maximum likelihood algorithm. Nt diversity and the distribution of nt distances were evaluated among A2 and D3 sequences. Analysis of Molecular variance (AMOVA) was conducted to measure the fraction of heterogeneity in genetic distances among the A2 sequences that arose due to differences between the A2 time clusters. WG sequences of the HBV isolates have been deposited in the National Center for Biotechnology Information GenBank database (accession numbers KF779209 - KF779386).

Estimation of evolutionary dates and demographic history

The time to the most recent common ancestor (tMRCA) for each genotype was calculated using BEAST (ver. 1.7.1) (18). Estimates were calculated using the HKY substitution model with 4 gamma rate categories and invariant sites. The calculations used a coalescent constant size tree prior and an uncorrelated lognormal molecular clock with an initial substitution rate estimate of $5x10^{-3}$ substitutions per site per year. The nt mean was estimated with a uniform prior distribution. Bayesian skyline plot analysis was done separately for genotype A2 and D3 using mean substitution rate estimates from the tMRCA estimates and a constant rate uncorrelated lognormal molecular clock with a coalescent Bayesian skyline prior and 10 groups

in a piecewise-constant skyline model. All calculation was run until the effective sample size (ESS) was greater than 200.

Statistical analysis

Risk-ratio estimates were calculated to determine the association between patient characteristics and phylogenetic distribution of WG sequences. We used univariate and multivariate analyses to determine factors associated with infection by HBV genotypes. We also compared genetic distances among HBV strains by geographic or temporal distributions. SAS for Windows Version 9.3 was used for statistical analysis. Differences in strain mutations were tested by Fisher's exact test (Supplementary Information).

RESULTS

HBV genetic diversity

From 1998 through 2006, 1,206 acute hepatitis B cases were reported from the 6 survey sites. The HBV *S* gene could be sequenced from 614 cases (13). As we have shown earlier (13), its diversity was restricted: among the A2 strains (75%), 47% shared 3 sequences, with 32% sharing a single sequence; and among the D strains (18%), 41% shared a single sequence (Fig.1a). WG sequences were obtained from 179 HBV strains, of which 134 (75%) belonged to A2, 32 (18%) to D3, and the remaining 13 (8%) to genotypes B, C, E, F, G and H. There are some unavoidable limitations to our study. We did not have sufficient serum volume or titers for HBV WG genome sequencing of all incident cases genotyped using the *S* gene amplicon. However, the % distribution of genotypes A, D and the others was similar among HBV strains, from which WG or only *S* gene sequences were obtained, confirming a fair representation of the sampled population by the 179 WG sequences. The WG sequences were unique (Fig.1b), except

158 for those from 2 A2 and 2 D3 strains; the identical A2 strains were from cases in the same
159 county and identified from the same year, and the identical D3 strains were from different
160 counties but identified in the same year. Five strains were recombinant: 2 between A1 and A2,
161 and 3 between D3 and A2. Cases infected by these recombinants were excluded from analysis.

162 The WG sequences were tightly clustered within A2 and D3 subtypes, producing a star-
163 like phylogeny, indicating close genetic relatedness within each subtype (Fig.1b) and recent
164 selective sweep or population expansion for each subtype. Genetic relatedness among A2 strains
165 varied from 97.4%-100%, and among D3 from 97.6%-100% (Fig.5b). Among A2 strains from
166 any county, the genetic relatedness ranged between 97.6% and 100%. The proportion of pairs of
167 sequences differing at > 3 nt positions was 99.8% for A2 strains, and 97.5% for D3 strains.

168 **Geographic and temporal distributions**

169 No correlation between genetic distance and geographic distance for either subtype A2 or
170 D3 was observed, although there was some degree of clustering of A2 and D3 strains from
171 Pinellas County, Florida, and Multnomah County, Oregon in phylogenetic tree (sequences
172 highlighted in yellow, Fig. 2a). No correlation was observed between genetic distances among
173 HBV strains and year of diagnosis. For example, the 8 sequences (highlighted in yellow, Fig. 2b)
174 that formed a tight cluster in the tree were diagnosed 1-4 years apart.

175 **Recent origin of A2 and D3 strains**

176 The tight sequence clustering (Fig.1) and the close genetic relatedness (Fig.5b) among
177 HBV strains of same subtypes indicate their recent origins. Bayesian analysis (Fig. 3) showed
178 that the MRCA for A2 existed in *ca*. 1979 (range, 1961-1989), and for D3 in *ca*.1987 (range,
179 1971-1998).

A2 and D3 population dynamics

Skyline plot analysis of all HBV A2 sequences indicated expansion of the A2 strains between 1994 and 1996, followed by a transient decline, and a second expansion during 2002. For D3 strains, there was a trend towards population expansion between 1998 and 2000 (Fig. 3).

Differential A2 expansion

Inspection of the A2 phylogenetic tree identified 2 rapidly evolving viral lineages that together form a cluster (designated Cluster 1, yellow circle in Fig. 4a). Although Cluster 1 contained HBV strains sampled from 1998 to 2006, ~80% of strains originated from cases diagnosed after 2002. A skyline plot constructed using sequences from Cluster 1 showed a sharp population expansion during 2002 (Fig. 4b). The other A2 sequences (designated Cluster 2, Fig. 4a) were mainly from samples collected before 2002. Analysis of these sequences showed an earlier expansion between 1994 and 1996 (Fig. 4c). These observations indicate that the HBV A2 strains experienced 2 recent expansions (Fig. 3a). The first expansion, mainly of Cluster 2 strains, occurred during the mid-1990s. The second, of Cluster 1 strains, occurred in 2002 and was accompanied with population decline among the Cluster 2 strains.

Changes in A2 and D3 heterogeneity

Differential expansion of specific phylogenetic lineages should have a significant effect on genetic heterogeneity of HBV population over time. Indeed, the mean nt diversity across all positions of the HBV genome were significantly greater ($p = 0.001$) in the A2 sequences of the 1998-2001 cases than those of the 2002-2006 cases (Fig. 5). However, the genetic differences were not distributed uniformly along the HBV genome, with regions at positions 1-400, 600-1000, 1500-1900 and 2200-2700 showing more diversity for A2 strains sampled between 1998 and 2001 than between 2002 and 2006 (Fig. 5a). The observation suggests that these genomic

regions contribute more than other to differentiation among HBV A2 strains sampled here. Comparison of sequences within and between Clusters 1 and 2 by AMOVA showed that 91.4% of all genetic variations occur within each cluster ($p = 0.0001$), and only 8.7% between clusters (Fig.5).

WG sequences were inspected for the presence of clinically important nt changes (Table 2). None of the substitutions known to confer vaccine escape were identified. The A1896 (stop codon at codon 28) pre-core mutation was found in 3 D3 strains and none in A2 strains. There were 18 strains carrying aa substitutions known to be associated with drug resistance: rtA194 (to adefovir/tenofovir; n=17), rtM204 and rtV207 (to lamivudine, n=16) and rtS202 and rtM250 (to entecavir, n=3). These drug-resistance associated changes were observed in 34% of D3 strains, 28% of A2 Cluster 2 strains, and none from Cluster 1 strains ($p < 0.0001$) (Table 2). Analysis of mutations within the HBsAg "*a*" determinant at positions 120-165 aa identified a single HBV A2 strain with the known vaccine-escape mutation T126I (Table 2).

Factors associated with selective A2 expansion

Analysis of demographic data of acute hepatitis B cases (Table 1) showed that HBV A2 strains comprising Cluster 1 (Fig. 4a) tended to be carried by persons: of African-American origin compared to whites (OR, 3.22; 95% CI, 1.44-7.2; $p = 0.005$); from Jefferson County, Alabama, compared to those from Denver, Colorado (OR, 5.85; 95% CI, 1.37-24.89; $p = 0.019$); and who were diagnosed after 2002 compared to those earlier than 2002 (OR, 4.47; 95% CI 95%, 2-9.98; $p = 0.001$). Among the risk factors examined, Cluster 1 strains were more common among cases who were MSM compared to those who were not (OR, 2.05; 95% CI, 1.03-4.05; $p = 0.03$). Multivariate analysis was conducted but owing to small sample sizes, no statistically significant conclusions could be made.

DISCUSSION

Genetic analysis of HBV WG sequences confirmed significant prevalence of genotypes A and D among acute hepatitis B cases in the US (1998-2006). All HBV genotype A variants belonged to subtype A2. Although the *S* gene sequences used in the previous study did not allow for a confident subtyping of the genotype D variants(13), phylogenetic analysis of the WG sequences conducted here showed for the first time that all HBV genotype D variants belonged to subtype D3. The high incidence of A2 in acute cases in the US parallels other global reports. In Europe, the majority of incident HBV strains belong to A2 (19, 20), with specific strains identified in English prisons (21), in outbreaks of nosocomial HBV transmissions in Germany (22) and Belgium (23), and among MSM in The Netherlands (24). In Japan, HBV A2 has been observed to be spreading not only within the MSM community (25), but also from there to the general population (26). The National Heart, Lung and Blood Institute Retrovirus Epidemiology Donor Study-II (REDS-II) of 34 million US blood donations between 2006 and 2009 reported that the majority of 193 donor HBV strains (37%) consisted of A2, with incident donors carrying higher frequencies of A2 (67%) compared to those in prevalent donors (27%) (27). A similar difference in genotype distribution between acute hepatitis and chronic hepatitis has been reported in Japan (28). There is preliminary evidence suggesting that the apparent high transmission rate of A2 may reflect the longer duration of HBV viremia in infected patients (29, 30).

Almost all HBV WG sequences were unique, with only 2 A2 and 2 D3 variants being identical. It is important to note that the observations made in this study show that sharing the *S* gene sequence (13) is not an indication of the close genetic relatedness among HBV A2 or D3 variants, since HBV variants carrying identical *S* gene sequences were scattered without

249 clustering within the subtype-specific branches in the phylogenetic tree constructed using WG
250 sequences (Fig 1). The identical *S* gene sequences shown as a single red node in Fig 1a (32% of
251 total sequences) are found in many different WG sequences in Fig 1b. Thus, the extent of HBV
252 genetic relatedness estimated using the *S* gene sequences should be interpreted with caution, and
253 limited to genotype assignment. Comprehensive analysis using WG rather than the *S* gene of
254 HBV should lend confidence to establishing genetic relatedness of incident HBV, and also to
255 detection of recombinants and clinically significant mutations.

256 Both A2 and D3 strains were of relatively recent origin, with the MRCA for each subtype
257 predicted to exist only 30-50 years ago, observations that also substantiate the close genetic
258 relatedness observed within each subtype. Despite their short evolutionary histories, both
259 genotypes experienced population expansions. The effective population size of D3 increased
260 between 1998 and 2000. The A2 strains underwent 2 rounds of population expansion (Fig. 3 and
261 4). The first expansion, between 1994 and 1996, coincided with a transient rise in notifications of
262 acute hepatitis B from 1994-1998 (4). The second expansion, during 2002, was observed for the
263 2 lineages that comprise Cluster 1. It was preceded or accompanied by a modest decline in the
264 effective population size for the other A2 strains (Cluster 2, Fig. 4). The second A2 expansion
265 coincided with increased acute hepatitis B notifications reported by the US National Notifiable
266 Disease Surveillance (NNDSS) between 1999 and 2002 among men aged > 19 years old (5%),
267 and among men and women aged > 40 years old (20% and 31%, respectively) (5).

268 The dynamic nature of HBV's epidemic history during the study period has resulted in
269 significant changes in genetic composition of the HBV A2 population carried by acute hepatitis
270 B cases. The A2 strains sampled after 2002 were restricted in genetic heterogeneity, with
271 reduction in the number of polymorphic sites unevenly distributed across the HBV genome (Fig.

5a). No known drug-resistance substitutions were found in these recently expanded strains, contrasting with almost a third of strains sampled before 2002 that carried such substitutions (Table 2). Consistent with the differential expansion of a single A2 cluster (Fig. 4a), the findings indicate that incident HBV populations circulating after 2002 were genetically distinct from those before 2002.

Reduction in genetic heterogeneity and the number of pre-existing drug-resistance mutations over time suggests the potentially changing capacity of HBV populations to respond to antiviral treatment. Predisposition to drug resistance is a convergent trait encoded in epistatic connectivity among HBV genomic sites (16,31). The shift in the HBV population structure observed in the current study likely reflects changes in epistatic connectivity pertaining to the presentation of drug-resistance mutations. Additionally, the rapid population expansion of the closely related HBV strains is likely consequence of frequent opportunities for transmission from persons with primary infections, who tend to carry fewer intra-host HBV variants (32), than from persons with chronic infection. Thus, reduction in frequency of the detected drug-resistance mutations after the second expansion period is likely associated with changing genetic structure and declining heterogeneity of the HBV population than with specific selection pressures related to therapeutic treatment.

Analysis of demographic characteristics of the cases showed that the Cluster 1 HBV A2 strains tended to be detected among African-Americans, persons from Jefferson County in Alabama, those diagnosed after 2002, and MSM (Table 1). Multivariate analysis did not produce statistically significant results, due to the relative small sample size. Nonetheless, although each of the 3 factors was shown to be independently associated with infection by Cluster 1 strains, MSM activity is the behavioral risk factor identified that would account for the observed increase

295 in the rate of transmission of these genetically close strains. While locally circulating HBV
296 strains may be expected to be genetically related, geographical location or race by itself cannot
297 explain increase in the rate of transmission. Furthermore, race cannot be independently
298 associated with genetic relatedness among incident HBV strains.

299 The demographic and risk associations to infection by Cluster 1 HBV A2 strains are
300 consistent with NNDSS data showing a rise in incidence rate of acute hepatitis B between 1999-
301 2002, particularly in southern USA, and among persons reporting engagement in high-risk
302 practices, such as MSM activity and IDU (4). Contemporaneous reports from other parts of the
303 world have shown sharp increases of acute infections with genotype A2 strains linked to similar
304 high-risk behaviors (20, 21). As noted, rapid transmissions among acutely infected MSM in the
305 Netherlands and Japan (25, 26) were associated with genetically identical or closely related A2
306 strains. Data from Japan also indicate a greater propensity of primary infection with A2,
307 compared to other genotypes, associated with persistent HBV infection (29,30). Although there
308 are currently no data indicating such propensity of the US A2 strains that underwent selective
309 transmission, it would be important to monitor the natural history of infection by these strains.
310 Our results indicate the possibility of missed opportunities to vaccinate MSM. Gaps in
311 implementing existing vaccination strategies must be addressed to increase hepatitis B
312 vaccination coverage for MSM.

313 This study highlights that full-length genomes are critical for complete characterization
314 of the biological properties of HBV variants and understanding the epidemiology of the disease.
315 Despite decline in the rate of incident HBV infection in the US, HBV strains identified among
316 acute hepatitis B cases have experienced 3 population expansions. The expansions altered
317 significantly the viral genetic composition, and affected capacity of HBV populations to

maintain resistance to antiviral treatment. A strong association between phylogeny and transmission rates reflected in selective viral expansion of specific HBV lineages, particularly of A2 strains, suggests biological differences among HBV variants affecting their dissemination, or the existence of host contact networks linked to specific risks, such as among MSM. Integration of hepatitis B vaccine into routine childhood vaccination schedules has dramatically increased immunization coverage among younger age groups, but similar gains in coverage have not been demonstrated among high-risk adults (33,34), a population that accounts for an estimated 75–95% of all incident HBV cases in the US (4,33). Increased HBV transmission rate among MSM reflects inadequate hepatitis B vaccination coverage, highlighting the need to improve coverage of at-risk adults (9,35,36).

ACKNOWLEDGEMENTS

We would like to acknowledge the contributions of JW Ward, D Holtzman, SD Holmberg, H Roberts, J Groeger and V Barry for their support.

REFERENCES

1. **Ott JJ, Stevens GA, Groeger J, Wiersma ST.** 2012. Global epidemiology of hepatitis B virus infection: new estimates of age-specific HBsAg seroprevalence and endemicity. Vaccine **30:**2212-2219.
2. **Weinbaum CM, Williams I, Mast EE, Wang SA, Finelli L, Wasley A, Neitzel SM, Ward JW, Centers for Disease C, Prevention.** 2008. Recommendations for identification and public health management of persons with chronic hepatitis B virus infection. MMWR. Recommendations and reports : Morbidity and mortality weekly report. Recommendations and reports / Centers for Disease Control **57:**1-20.
3. **Lu PJ, Byrd KK, Murphy TV, Weinbaum C.** 2011. Hepatitis B vaccination coverage among high-risk adults 18-49 years, U.S., 2009. Vaccine **29:**7049-7057.
4. **Goldstein ST, Alter MJ, Williams IT, Moyer LA, Judson FN, Mottram K, Fleenor M, Ryder PL, Margolis HS.** 2002. Incidence and risk factors for acute hepatitis B in the United States, 1982-1998: implications for vaccination programs. J Infect Dis **185:**713-719.
5. **Centers for Disease C, Prevention.** 2004. Incidence of acute hepatitis B--United States, 1990-2002. MMWR. Morbidity and mortality weekly report **52:**1252-1254.
6. **Ladak F, Gjelsvik A, Feller E, Rosenthal SR, Montague BT.** 2012. Hepatitis B in the United States: ongoing missed opportunities for hepatitis B vaccination, evidence from the Behavioral Risk Factor Surveillance Survey, 2007. Infection **40:**405-413.
7. **Lum PJ, Hahn JA, Shafer KP, Evans JL, Davidson PJ, Stein E, Moss AR.** 2008. Hepatitis B virus infection and immunization status in a new generation of injection drug users in San Francisco. J Viral Hepat **15:**229-236.
8. **Centers for Disease C, Prevention.** 2012. Adult vaccination coverage--United States, 2010. MMWR. Morbidity and mortality weekly report **61:**66-72.
9. **Pitasi MA, Bingham TA, Sey EK, Smith AJ, Teshale EH.** 2014. Hepatitis B virus (HBV) infection, immunity and susceptibility among men who have sex with men (MSM), Los Angeles County, USA. AIDS and behavior **18 Suppl 3:**248-255.
10. **Tellinghuisen TL, Evans MJ, von Hahn T, You S, Rice CM.** 2007. Studying hepatitis C virus: making the best of a bad virus. J Virol **81:**8853-8867.
11. **Alter MJ, Hadler SC, Margolis HS, Alexander WJ, Hu PY, Judson FN, Mares A, Miller JK, Moyer LA.** 1990. The changing epidemiology of hepatitis B in the United States. Need for alternative vaccination strategies. JAMA : the journal of the American Medical Association **263:**1218-1222.
12. **Nainan OV, Alter MJ, Kruszon-Moran D, Gao FX, Xia G, McQuillan G, Margolis HS.** 2006. Hepatitis C virus genotypes and viral concentrations in participants of a general population survey in the United States. Gastroenterology **131:**478-484.
13. **Teshale EH, Ramachandran S, Xia GL, Roberts H, Groeger J, Barry V, Hu DJ, Holmberg SD, Holtzman D, Ward JW, Teo CG, Khudyakov Y.** 2011. Genotypic distribution of hepatitis B virus (HBV) among acute cases of HBV infection, selected United States counties, 1999-2005. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America **53:**751-756.
14. **Chu CJ, Keeffe EB, Han SH, Perrillo RP, Min AD, Soldevila-Pico C, Carey W, Brown RS, Jr., Luketic VA, Terrault N, Lok AS.** 2003. Hepatitis B virus genotypes in the United States: results of a nationwide study. Gastroenterology **125:**444-451.

15. **Ramachandran S, Zhai X, Thai H, Campo DS, Xia G, Ganova-Raeva LM, Drobeniuc J, Khudyakov YE.** 2011. Evaluation of intra-host variants of the entire hepatitis B virus genome. PLoS One **6:**e25232.
16. **Thai H, Campo DS, Lara J, Dimitrova Z, Ramachandran S, Xia G, Ganova-Raeva L, Teo CG, Lok A, Khudyakov Y.** 2012. Convergence and coevolution of hepatitis B virus drug resistance. Nature communications **3:**789.
17. **Utsumi T, Lusida MI, Yano Y, Nugrahaputra VE, Amin M, Juniastuti, Soetjipto, Hayashi Y, Hotta H.** 2009. Complete genome sequence and phylogenetic relatedness of hepatitis B virus isolates in Papua, Indonesia. J Clin Microbiol **47:**1842-1847.
18. **Drummond AJ, Rambaut A.** 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC evolutionary biology **7:**214.
19. **Sloan RD, Strang AL, Ramsay ME, Teo CG.** 2009. Genotyping of acute HBV isolates from England, 1997-2001. J Clin Virol **44:**157-160.
20. **Laoi BN, Crowley B.** 2008. Molecular characterization of hepatitis B virus (HBV) isolates, including identification of a novel recombinant, in patients with acute HBV infection attending an Irish hospital. J Med Virol **80:**1554-1564.
21. **Hallett RL, Ngui SL, Meigh RE, Mutton KJ, Boxall EH, Teo CG.** 2004. Widespread dissemination in England of a stable and persistent hepatitis B virus variant. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America **39:**945-952.
22. **Dreesman JM, Baillot A, Hamschmidt L, Monazahian M, Wend UC, Gerlich WH.** 2006. Outbreak of hepatitis B in a nursing home associated with capillary blood sampling. Epidemiology and infection **134:**1102-1113.
23. **Pourkarim MR, Verbeeck J, Rahman M, Amini-Bavil-Olyaee S, Forier AM, Lemey P, Maes P, Van Ranst M.** 2009. Phylogenetic analysis of hepatitis B virus full-length genomes reveals evidence for a large nosocomial outbreak in Belgium. J Clin Virol **46:**61-68.
24. **van Houdt R, Bruisten SM, Geskus RB, Bakker M, Wolthers KC, Prins M, Coutinho RA.** 2010. Ongoing transmission of a single hepatitis B virus strain among men having sex with men in Amsterdam. J Viral Hepat **17:**108-114.
25. **Fujisaki S, Yokomaku Y, Shiino T, Koibuchi T, Hattori J, Ibe S, Iwatani Y, Iwamoto A, Shirasaka T, Hamaguchi M, Sugiura W.** 2011. Outbreak of infections by hepatitis B virus genotype A and transmission of genetic drug resistance in patients coinfected with HIV-1 in Japan. J Clin Microbiol **49:**1017-1024.
26. **Tamada Y, Yatsuhashi H, Masaki N, Nakamuta M, Mita E, Komatsu T, Watanabe Y, Muro T, Shimada M, Hijioka T, Satoh T, Mano Y, Komeda T, Takahashi M, Kohno H, Ota H, Hayashi S, Miyakawa Y, Abiru S, Ishibashi H.** 2012. Hepatitis B virus strains of subgenotype A2 with an identical sequence spreading rapidly from the capital region to all over Japan in patients with acute hepatitis B. Gut **61:**765-773.
27. **Delwart E, Slikas E, Stramer SL, Kamel H, Kessler D, Krysztof D, Tobler LH, Carrick DM, Steele W, Todd D, Wright DJ, Kleinman SH, Busch MP, Group N-R-IS.** 2012. Genetic diversity of recently acquired and prevalent HIV, hepatitis B virus, and hepatitis C virus infections in US blood donors. J Infect Dis **205:**875-885.
28. **Takeda Y, Katano Y, Hayashi K, Honda T, Yokozaki S, Nakano I, Yano M, Yoshioka K, Toyoda H, Kumada T, Goto H.** 2006. Difference of HBV genotype distribution between acute hepatitis and chronic hepatitis in Japan. Infection **34:**201-207.
29. **Yamada N, Yotsuyanagi H, Okuse C**. 2010. Duration of HBs antigenemia in patients with acute hepatitis B. Kanzo **51**: 534-535.

30. **Ishii K, Kogame M, Shiratori M**. 2010. Clinical characteristics of patients with acute hepatitis B genotype A. Kanzo **51**:397-399.

31. **Khudyakov Y**. 2010. Coevolution and HBV drug resistance. Antivir Ther **15**: 505-515.

32. **Osiowy C, Giles E, Tanaka Y, Mizokami M, Minuk GY**. 2006. Molecular evolution of hepatitis B virus over 25 years. J Virol; 80:10307-14.

33. **Mast EE, Weinbaum CM, Fiore AE, Alter MJ, Bell BP, Finelli L, Rodewald LE, Douglas JM, Jr., Janssen RS, Ward JW, Advisory Committee on Immunization Practices Centers for Disease C, Prevention.** 2006. A comprehensive immunization strategy to eliminate transmission of hepatitis B virus infection in the United States: recommendations of the Advisory Committee on Immunization Practices (ACIP) Part II: immunization of adults. MMWR. Recommendations and reports : Morbidity and mortality weekly report. Recommendations and reports / Centers for Disease Control **55:**1-33; quiz CE31-34.

34. **Matthews JE, Stephenson R, Sullivan PS.** 2012. Factors associated with self-reported HBV vaccination among HIV-negative MSM participating in an online sexual health survey: a cross-sectional study. PLoS One **7:**e30609.

35. **Centers for Disease Control and Prevention**. 2012. Adult vaccination coverage - United States 2010. MMWR Recomm Rep **61**:66-72.

36. **Rijckevorsel GV, Whelan J, Kretzschmar M, Siedenburg E, Sonder G, Geskus R, Coutinho R, van den Hoek A.** 2013. Targeted vaccination programme successful in reducing acute Hepatitis B in men having sex with men in Amsterdam, The Netherlands. J Hepatol **59(6) :**1177-1183.

445 **FIGURE LEGENDS**

446 **Fig. 1** Phylogenetic maximum likelihood tree constructed using (a) HBV *S gene* sequences (378
447 bp) and (b) WG sequences (3.2kb) from genotype A and D. Frequency of *S gene* sequences is
448 color-coded, with color scale shown in vertical bar for each phylogenetic tree. Red circles in
449 panel (b) represent sequences that have the most common variant of the *S* gene found in 76 WGs.
450 Each white circle represents a WG sequence in which *S* gene is unique.

451 **Fig. 2** Phylogenetic analysis of WG sequences belonging to genotypes A and D, color-coded
452 according to county and year of sampling. Sentinel Counties are: Jefferson County, Alabama;
453 Pinellas County, Florida; Pierce County, Washington; Multnomah County, Oregon; San
454 Francisco County, California and Denver County, Colorado.

455 **Fig. 3** Bayesian skyline plot showing HBV epidemic history in the US. Middle line represents
456 mean estimate of effective population size, plotted as log $N_e\tau$, and blue lines shows 95% highest
457 posterior density (HPD) intervals for this estimate. Time scale encompasses lower limit of 95%
458 HPD for time of most recent common ancestor (tMRCA) to time of collection for most recent
459 specimens.

460 **Fig. 4(a)** Phylogenetic analysis of WG (genotype A2) color-coded according to year of
461 diagnosis. The bootstrap values are marked on figure. Cluster 1 sequences (highlighted in
462 yellow) represent most heterogeneous lineage of sequences compared to Cluster 2 (not
463 highlighted). (**b**) Skyline plot for the A2 Cluster 1. Middle black line represents estimated mean
464 of effective population size (logarithmic scale). Blue lines show limits of the 95% HPD for this
465 estimate. Vertical dotted line represents lower 95% HPD for tMRCA. (**c**) Skyline plot for A2
466 Cluster 2.

467 **Fig. 5**(**a**) Nuncleotide (nt) diversity of entire HBV genome color-coded based according to year
468 of diagnosis. Each point is an average over window of 301 nt, with step of 1. The cartoon at the
469 bottom of the figure depicts the HBV genetic map in alignment with the nt numbering on the X-
470 axis. (**b**) Frequency distribution of nt distances among WG from A2 and D3 isolates color-coded
471 according to year of diagnosis.

472 **Table 1.** Demographic and risk characteristics of A2-infected cases stratified according to
473 infection by Cluster 1 and 2 strains. 'Referrant' is the group to which another group is compared
474 to.

475 **Table 2.** Distribution of mutations - (**a**) any clinical significant nt changes (in pre-core/core, S
476 and polymerase regions) in genotypes A2 and D3; p value based on comparison between the %
477 of two subtypes, (**b**) drug-resistance mutations rtA194 (to adefovir/tenofovir; n=17), rtM204 and
478 rtV207 (to lamivudine, n=16) and rtS202 and rtM250 (to entecavir, n=3). p value based on
479 comparison between % of the two A2 clusters; (**c**) *S gene* mutations (S120 to S165); and (**d**)
480 precore (A1896, stop codon 28) mutation.
481

Table 1

Genotype A2	Cluster 1		Cluster 2		OR (95% CI)*	Total
	n	%	n	%		
Race/ethnicity						
White	21	31.3	46	68.7	Referent	67
Black	25	59.5	17	40.5	3.22 (1.44, 7.20)	42
Hispanic	4	30.8	9	69.2	0.97(0.27, 3.52)	13
County						
Denver, CO	3	20.0	12	80.0	Referent	15
Jefferson, AL	19	59.4	13	40.6	5.85 (1.37, 24.89)	32
Multnomah. OR	5	31.3	11	68.7	1.82 (0.35, 9.46)	16
Pinellas, FL	24	40.7	35	59.3	2.74 (0.70, 10.77)	59
Pierce, WA	2	50.0	2	50.0	4.00 (0.39, 41.23)	4
San Francisco, CA	0	0.0	3	100.0	Not applicable	3
Year of diagnosis						
1998 - 2001	11	21.2	41	78.8	Referent	52
2002 - 2006	42	54.5	35	45.5	4.47 (2.00, 9.98)	77
HBV infection risk						
MSM - Yes	25	75.8	8	24.2	Referent	33
MSM - No	33	52.4	30	47.6	2.84 (1.11, 7.25)	63
IDU - Yes	6	54.5	5	45.5	Referent	11
IDU - No	45	39.8	68	60.2	0.55(0.16, 1.92)	113

* Exact odds ratio 95% confidence intervals

Table 2

(a)

Genotype	Clinically significant mutations (RT, S, Pre-core/Core)		Total cases
	No (%)	Yes (%)	
A2	112(86.2)	18(13.8)	130
D3	13(44.8)	16(55.2)	29
Total	125(78.6)	34(21.4)	159
	P<0.0001	P<0.0001	

(b)

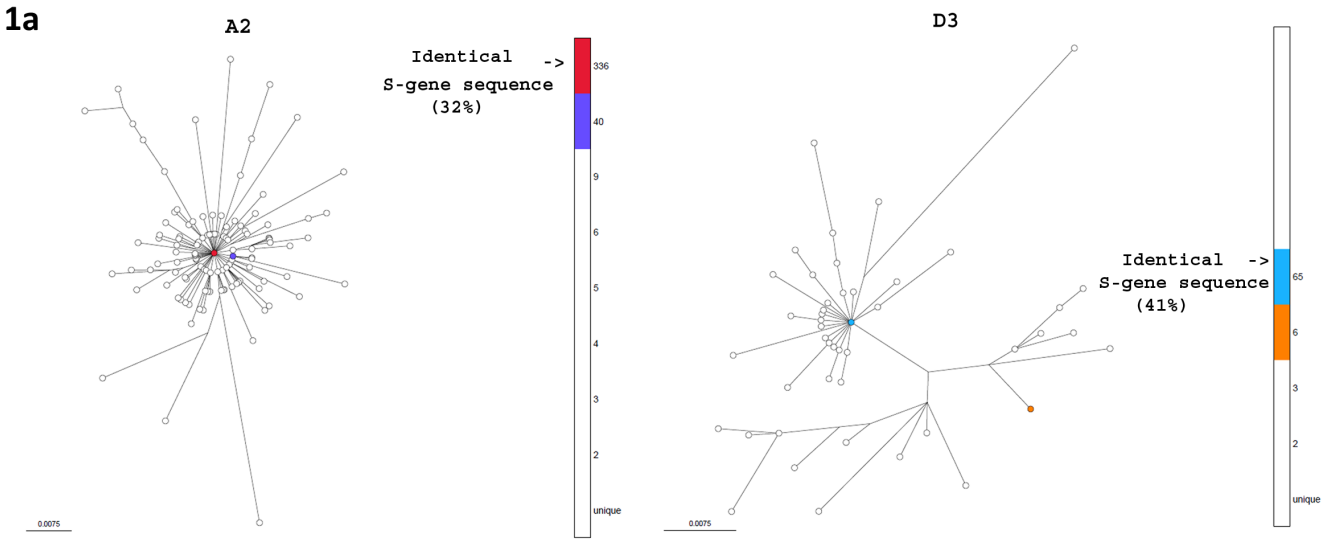
Genotype (Year)	HBsAg mutation (RT)		Total cases
	No (%)	Yes (%)	
A2 (Cluster 2)	46(71.9)	18(28.1)	64
A2 (Cluster 1)	66(100)	0(0)	66
D3 (1998-2006)	18(62.1)	11(37.9)	29
Total	101(77.7)	29(22.3)	130
		P<0.0001	

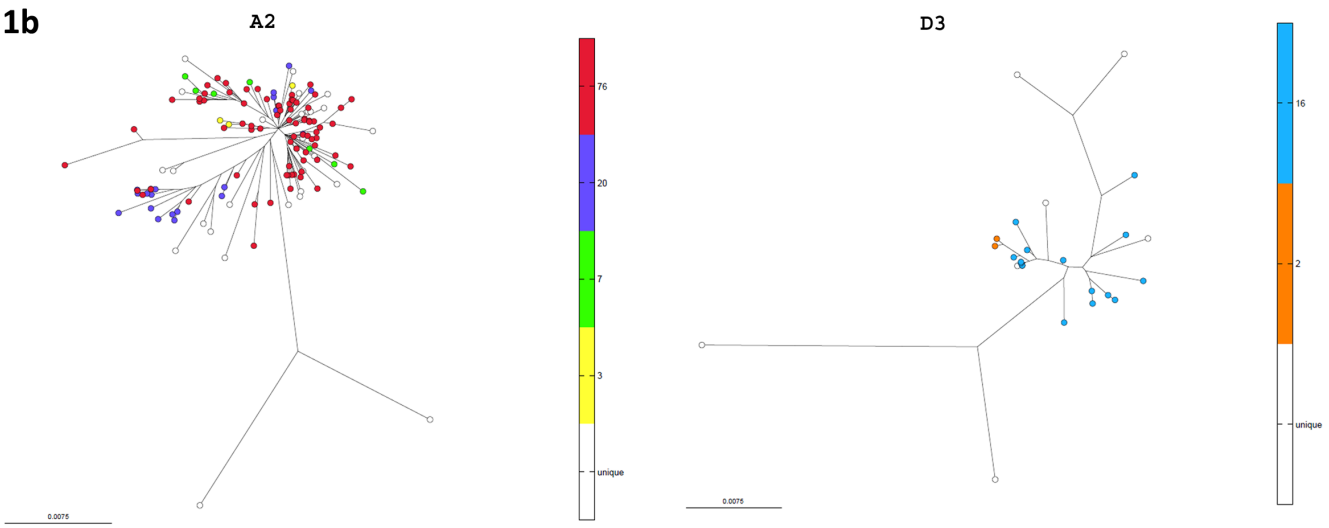
(c)

Genotype (Year)	Vaccine escape mutation (S)		Total cases
	No (%)	Yes (%)	
A2 (Cluster 2)	64(100)	0(0)	64
A2 (Cluster 1)	66(100)	0(0)	66
D3 (1998-2006)	28(96.6)	1(3.4)	29
Total	129(99.2)	1(0.8)	130

(d)

Genotype (Year)	eAg mutation (Pre-core/Core)		Total cases
	No (%)	Yes (%)	
A2 (Cluster 2)	64(100)	0(0)	64
A2 (Cluster 1)	66(100)	0(0)	66
D3 (1998-2006)	26(89.7)	3(10.3)	29
Total	127(97.7)	3(2.3)	130





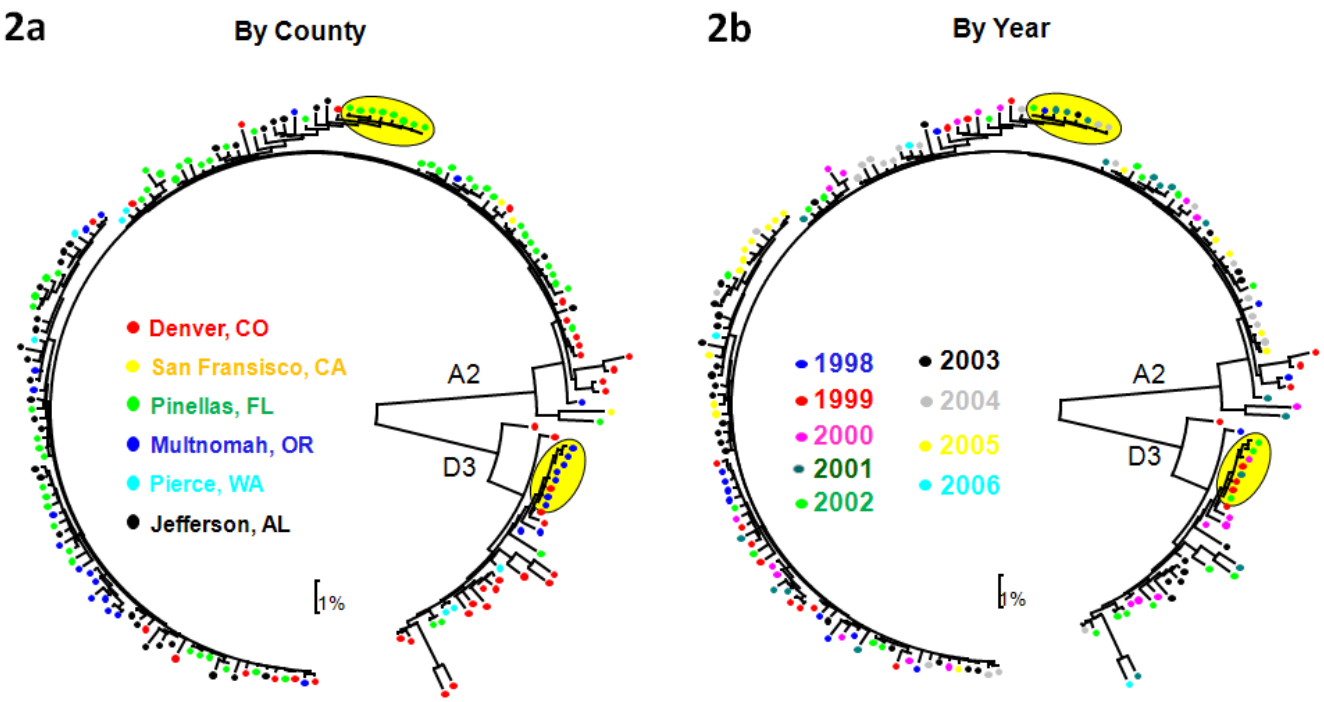


Fig 3

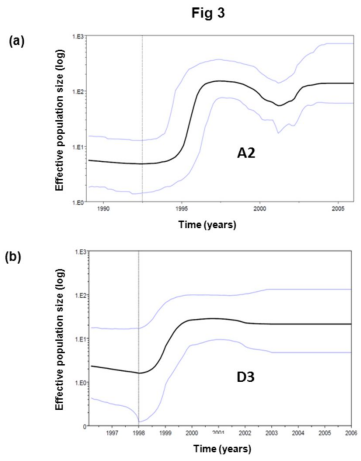


Fig 4

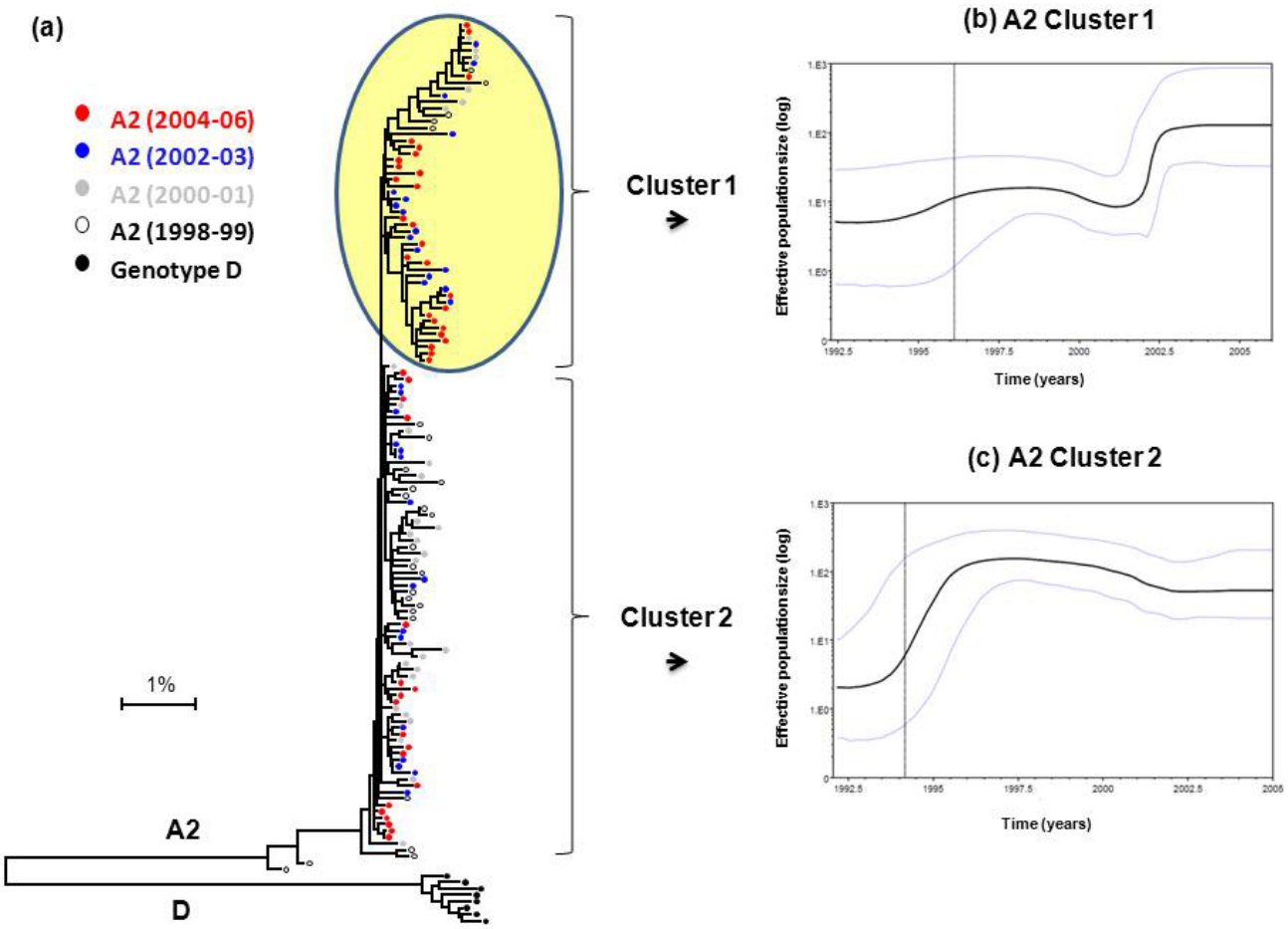


Fig 5