Selecting a Genetic Region for Molecular Analysis of Hepatitis B Virus Transmission

The optimal genetic regions for the phylogenetic analysis of hepatitis B virus (HBV) transmission continue to be a matter of debate, with different investigators preferring different regions. However, full-length HBV sequence analysis is the gold standard for the purpose. But in developing countries, such as ours, where HBV infections are endemic, the added cost of full-length sequencing becomes a limiting factor for studying large numbers of samples. Thus, the search for a suitable genetic region of HBV is important.

Initially clustering of HBV seromarkers (7, 12) and subsequently clustering of well-established mutations (e.g., HBsAgG145R or HBcAgW28Stop) were used to demonstrate intrafamilial transmission (1, 9, 11, 13, 15, 18, 19). Recently, to increase the confidence level of detection of true transmission events, phylogenetic analysis with the bootstrap resampling/maximum-likelihood test of surface/precore (preC)/core region sequences was carried out (6, 8, 22). However, clustering of sequences from epidemiologically unrelated families was suggested to be due to a high degree of conservation of surface (S) gene sequence (8), which led to the region being considered unsuitable for transmission studies.

Thus, analysis of nonoverlapping, fast-evolving regions was recommended (3, 5). Interestingly, 67% of the HBV genome is overlapping (16), leaving distal X/preC/partial core regions nonoverlapped. These regions encode important RNA structural elements, such as the epsilon signal (10), that are essential for HBV replication. Thus, one can assume that high variability in these regions might have negative selection pressure; on the other hand, variability in the HBsAg is positively selected to evade host immune pressure. Recently, using statistical models, Szmaragd and colleagues (17) found that overlapping sites have slightly higher substitution rates than nonoverlapping regions, which supports the above assumption.

Further, variability of a genetic region or prevalence of certain mutations varies with the study population, infecting genotype (10), immune status, chronicity (20), clinical outcome (21), duration (associated time frame), and mode of infection (nosocomial, vertical, or intrafamilial horizontal). Thus, as recommended by Bracho and colleagues (3) for investigating the chain of recent/nosocomial fulminant cases, analysis of highly variable preC/core sequence associated with fulminant hepatitis B (14) should be preferred.

Our population shows a low preC mutation prevalence (2), and S gene sequence analysis provided more phylogenetic signal, which is more appropriate for tracing horizontal transmission patterns among HBsAg-negative family members (6). High S-gene variability has been documented in previous studies among HBsAg-negative subjects (20) or among chronic virus carriers or their families (18). In fact, different specific variability levels for the S gene (genotype, subgenotype, and subtype), in addition to mutations, can provide enough confidence to prove transmission events. Actually, in one of our earlier studies, where the preC/core region of all the isolates was identical, genotype, subtype, and mutation analysis of S gene sequences proved two different sources and evidence of horizontal transmission of HBV infection in a family (4).

Thus, before selecting a genetic region for investigations of transmission, it is more reasonable to consider the genetic variability of HBV among the study population, their serological profile, and the mode of probable transmission rather than to adhere to a specific genetic region, found to be more variable in a study of different population groups with different HBV genotypes, disease severity, or serological patterns.

S.D., A.B., and P.K.C. are supported by research scholarships from UGC, ICMM, Government of India, and WBSAPCS, Government of West Bengal, India, respectively.

REFERENCES


Sibnarayan Datta
Arup Banerjee
Partha Kumar Chandra
Runu Chakravarty
ICMR Virus Unit, Kolkata GBA, 1st Floor
ID & BG Hospital Campus
Kolkata 700010, India

*Phone: (91 33) 2353 7425
Fax: (91 33) 2353 7424
E-mail: runugc@yahoo.co.uk

Author’s Reply

The molecular epidemiology of HBV infection is indeed complex, and phylogenetic analysis is complicated by the compact genomic structure comprising four partially overlapping open reading frames with no noncoding regions in the genome—all regulatory signals also encode proteins (2, 10). Many of the products of these genes are important to viral structure and function and may be subject to significant selection pressure, such as host immune responses acting on the HBsAg and HBcAg products or antiviral medications acting directly on the viral polymerase (with corresponding changes in overlapping S gene sequences) (3, 8, 11–13).

Much care must therefore be taken in analyzing sequence heterogeneity in putatively related HBV strains. As noted by Datta and colleagues in the preceding letter, particular regard must be given to the influence of host immune pressure on the infecting virus, ranging from little or no selection pressure being exerted during the immune-tolerant phase of infection, with a correspondingly low rate of mutations in genes encoding immunodominant epitopes, to much greater pressure during and after immune clearance, with subsequent selection of mutations providing escape from this pressure. Examples of these mutations include the classic precore mutation resulting in loss of HBcAg and the range of S-gene polymorphisms that have been implicated in immune escape (1, 7–10, 14, 15).

In the interesting study by these authors published earlier this year (5), contacts of known HBsAg-positive patients were screened to detect those with occult HBV infection—detectable HBV DNA in the serum in the absence of HBsAg. The S gene of the infecting HBV was sequenced to determine potentially significant mutations and was also used for phylogenetic analysis of transmission within families—in their letter, they state that it is a more appropriate target for tracing transmission patterns among HBsAg-negative family members, with more “phylogenetic signal” in the S gene than in other areas. However, in this study population, this signal may represent the polymorphisms evolved within individual hosts in response to immune pressure on HBsAg (8), and as stated in the introduction to their study, “ occult HBV infection is often explained by low levels of HBV DNA and a significant increase in genetic variability in the ‘a’ determinant” (5, 16). Such variability, arising within the host in response to immune pressure, could complicate attempts to define phylogenetic relatedness at the time of infection years before (15), and conversely similar patterns of mutations in unrelated viral strains could possibly explain some of the clustering observed between epidemiologically unconnected individuals in this population. This idea is supported by the observation made by the authors that the phylogenetic tree shows clustering by whether the sequenced viruses were wild-type or variant strains (5).

There is evidence that the S gene can be a relatively conserved region (outside the context of immune selection as discussed above), impairing the ability to distinguish transmission patterns (4, 6, 14). The assumption that distal X/precore/core region variability is deleterious to the virus and therefore will not be observed is not supported by other transmission investigations (2, 6, 14, 17), and in the Australian and Gambian studies previously cited (4), these regions revealed transmission patterns where S gene analysis did not, due to a lack of sufficient variability (6, 14).

Full genome sequencing is the gold standard for phylogenetic analysis of HBV transmission, but the increased cost is certainly a consideration, particularly in many regions where the burden of HBV infection is highest. The search for a suitable genetic target for transmission analysis is important. However, before selecting a convenient single gene target for such analysis, I would concur with Bracho and colleagues (2) that consideration of the heterogeneity of variability across the HBV genome in the study population is important and that a systematic way to determine this is to undertake full genome sequencing. Finally, regard must be given to the stage of HBV infection in the patients being studied, with the presence or absence of significant immune pressure being critical to the rate of mutation in genomic regions encoding immunologically important structures, such as the S gene.

B.C.C. is financially supported by the Centre for Clinical Research Excellence in Infectious Diseases and by a Public Health Postgraduate Scholarship from the National Health and Medical Research Council, Australia.

There is no conflict of interest.

REFERENCES


Benjamin C. Cowie*
Victorian Infectious Diseases Reference Laboratory
10 Wreckyn Street, North Melbourne
Victoria 3195, Australia

*Phone: (61 3) 9342 2606
Fax: (61 3) 9342 2666
E-mail: benjamin.cowie@mh.org.au