Journal of Infectious Diseases Advance Access published January 31, 2013

MAJOR ARTICLE

Detection of Hepatitis C Virus Transmission by Use of DNA Mass Spectrometry

Lilia M. Ganova-Raeva,[1] Zoya E. Dimitrova,[1] David S. Campo,[1] Yulin Lin,[1] Sumathi Ramachandran,[1] Guo-liang Xia,[1] Christiane Honisch,[2] Charles R. Cantor,[2] and Yury E. Khudyakov[1]

[1]Molecular Epidemiology and Bioinformatics Laboratory, Division of Viral Hepatitis, Centers for Disease Control and Prevention, Atlanta, Georgia; and [2]Sequenom, San Diego, California

The molecular detection of transmission of rapidly mutating pathogens such as hepatitis C virus (HCV) is commonly achieved by assessing the genetic relatedness of strains among infected patients. We describe the development of a novel mass spectrometry (MS)–based approach to identify HCV transmission. MS was used to detect products of base-specific cleavage of RNA molecules obtained from HCV polymerase chain reaction fragments. The MS-peak profiles were found to reflect variation in the HCV genomic sequence and the intrahost composition of the HCV population. Serum specimens originating from 60 case patients from 14 epidemiologically confirmed outbreaks and 25 unrelated controls were tested. Neighbor-joining trees constructed using MS-peak profile–based Hamming distances showed 100% accuracy, and linkage networks constructed using a threshold established from the Hamming distances between epidemiologically unrelated cases showed 100% sensitivity and 99.93% specificity in transmission detection. This MS-based approach is rapid, robust, reproducible, cost-effective, and applicable to investigating transmissions of other pathogens.

Keywords. viral transmission; genetic distance; RNA Mass-spectrometry; relatedness; network.

Hepatitis C virus (HCV) is a human pathogen belonging to the Flaviviridae family [1]. Its prevalence is about 2%–3% globally but can be >20% in regions of Africa and Asia where HCV is highly endemic [2–6]. In the United States, 3.2 million people are estimated to be infected with HCV [7, 8]. Infection is associated with the development of cirrhosis and hepatocellular carcinoma and is a major indication for liver transplantation [9]. Widespread use of injectable agents, whether medically approved or illicit, is a leading risk factor for HCV transmission [10, 11]; patients in healthcare settings are of special concern [12, 13]. The current standard-of-care therapy, which involves interferon in combination with ribavirin, is efficacious in approximately 50% of chronically infected patients [14]. As HCV infection is often asymptomatic, many infected people remain unidentified and, therefore, are not offered treatment [15]. There is no vaccine against hepatitis C.

Genetic heterogeneity is a hallmark of HCV. The virus is classified into 6 major genotypes and >50 subtypes [16]. In each infected individual, HCV exists as multiple variants or quasi-species [17–20]. The HCV genome, which encodes a single polyprotein, contains 2 highly heterogeneous regions: the hypervariable region 1 (HVR1), located at the 5′ end of the envelope gene E2, and a region located in the NS5a gene [21–24]. Consensus sequencing of these 2 regions, and also of NS5b [25, 26], is commonly used to determine genetic relatedness among HCV strains and to identify HCV transmission [27]. However, a consensus sequence cannot adequately represent the entire HCV population present in the host, particularly in chronically infected patients, in whom the viral genetic heterogeneity can be extensive [28, 29]. Additionally, the extensive intrahost HCV evolution results in considerable changes in the structure of the HCV population in

Received 23 July 2012; accepted 23 October 2012.

Correspondence: Lilia M. Ganova-Raeva, PhD, 1600 Clifton Rd NE, MS A-33, Atlanta, GA 30329 (lkg7@cdc.gov).

The Journal of Infectious Diseases

DOI: 10.1093/infdis/jis938

patients over time, thus impeding genetic identification of HCV transmission that occurred in the distant past [30]. In an outbreak setting, consensus sequencing may prevent the identification of HCV variants directly involved in transmission among genetically related variants that are unassociated with the outbreak [31].

Accurate identification of HCV strains involved in transmission can be achieved by matching the genetic compositions of viral populations sampled from infected hosts. Since HVR1 is one of the most variable regions of the HCV genome, analysis of intrahost HVR1 variants is frequently used for identification of and tracking HCV transmissions [32–37]. Such analysis involves (1) separation of individual HVR1 variants either by genetic cloning of PCR amplicons [38] or by polymerase chain reaction (PCR) cloning, using end-point limiting-dilution of complementary DNA [32, 39]; and (2) sequencing of these variants. Next-generation sequencing technologies couple the separation of genetic variants with sequencing, thus simplifying significantly the assessment of viral heterogeneity. However, the high rate of sequencing errors generated per DNA read [40] and the problematic representation of intrahost viral heterogeneity [41] are potential hindrances to adopting next-generation sequencing approaches to the detection of transmissions.

DNA mass spectrometry (MS) provides a different platform for genetic analysis of viral strains. It has successfully been adopted for resequencing, microbial typing, and single-nucleotide polymorphism discovery [42, 43]. Here, we describe the development of an MS approach for identification of genetic equivalence among HCV strains involved in outbreaks. It does not entail separation and sequencing of viral variants; rather, it uses patterns of MS peaks generated from the entire intrahost viral population. We found the MS approach to be as accurate as previously established molecular technologies [33, 36] in the detection of HCV transmission but significantly more cost-effective and less time-consuming.

MATERIALS AND METHODS

Specimens

Sixty serum specimens were collected from 14 epidemiologically confirmed HCV outbreaks that occurred in the United States during 2003–2008 [35, 39–43]. A total of 25 serum specimens from patients infected with epidemiologically unrelated HCV subtypes 1a and 1b, identified from the National Health and Nutrition Examination Survey III, were geographically matched to the outbreak specimens. The collections are listed and described under SD1 in the Supplementary Data file. Genotypes and subtypes of each case were previously determined by phylogenetic analysis of the NS5B sequences.

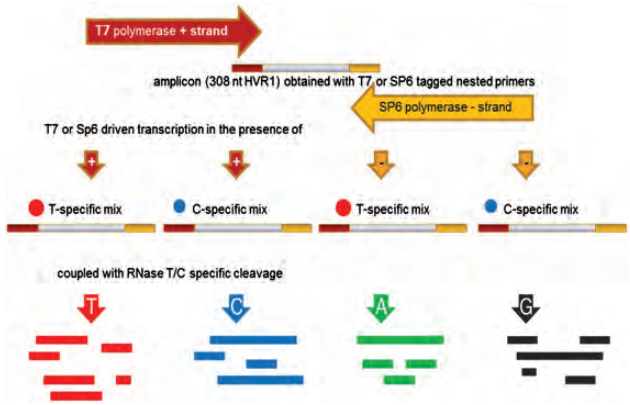

Figure 1. Principle of transcription coupled RNaseA cleavage. Cleavage products are printed on 384-spot SpectroCHIPs and analyzed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry.

Nucleic Acid Isolation, Reverse Transcription PCR (RT-PCR), Nested PCR, and MassCLEAVE™

Nucleic acids were isolated with MagnaPure LC (Roche, SD2). The extracted nucleic acid underwent RT-PCR. HVR1 was amplified by PCR by using previously described protocol [30] (SD2). To assess heterogeneity, we repeatedly (8–12 times) sampled the first-round PCR products of HVR1 without limiting dilution to generate templates for the second-round amplification (SD2), using tagged nested primers. The second-round PCR product, a 308-nucleotide fragment encompassing the HCV genomic region at nucleotide positions 1302–1610, was then subject to the MassCLEAVE™ protocol [42] (SD4). The process is illustrated in Figure 1. All primers used for the first- and second-round PCRs are listed in SD3. All samples were processed in parallel in 384-well microtiter plates. PCR setup, SAP, post-PCR base-specific cleavage reactions (MassCLEAVE™, Sequenom, San Diego, CA), and pre- and postsequencing treatments were performed using an automated liquid handler Biomek 3000 (Beckman Coulter, Fullerton, CA).

MS-Peak Profiles

Data on MS peaks were extracted from runs for all 4 nucleotides after evaluation using an internal-system quality cutoff. The data were collected as lists of detected peaks for every single nucleotide–specific reaction, along with their corresponding mass and normalized peak intensity, which reflects the number of fragments found with the same mass [43]. To allow data collection, we used a small reference set compiled from GenBank entries representing HVR1 sequences of genotypes 1a and 1b (SD5). Peak lists were exported from iSEQ and merged according to an existing algorithm provided by Sequenom (San Diego, CA).

The MS-peak list was concatenated in the order T(TF), A (TR), C(CF), and G(CR) to generate 1 MS-peak profile per sample. Each MS-peak profile was presented as a numeric vector of the detected masses and their corresponding signal intensity, which was used to calculate the distances between the MS-peak profiles (SD6). MS-peak profile distances between all subsamples with 4 eligible reactions were measured using the Hamming distance (presence or absence of a peak), Euclidean distance (taking into consideration peak intensity), and modified Euclidian distance (with a limit in the maximum intensity). Since all distances showed similar results, the Hamming distance was used throughout the analysis.

The MS-peak profile- and sequence-based distance matrices were used to construct the neighbor-joining trees [44]. The sequence-based distances were calculated using the Kimura 2-parameter model [45]. All calculations were performed using Matlab (Mathworks, Natick, MA).

Statistical Analyses

Analysis of molecular variance (AMOVA) was used to quantify the level of differentiation between cases. AMOVA partitions the variance of the MS-peak profile– or sequence-based distances into 3 sources: among the replicas for each HCV strain, among HCV variants from a single outbreak, and among epidemiologically unrelated HCV strains. AMOVA was calculated using Arlequin [46]. Significance levels of the genetic variance components were estimated using a permutation test (n = 10 000). In this test, the group label of each population sample is randomly permutated to simulate the null hypothesis that there is no difference between the groups. It must be noted that the pairwise comparisons among unrelated strains include the following: (1) pairwise distances among epidemiologically unrelated strains, (2) pairwise distances between epidemiologically unrelated and outbreak strains, and (3) pairwise distances between strains from different outbreaks. Pearson correlation was used to establish the degree to which sample heterogeneity as identified by MS-peak profile is related to the heterogeneity observed by sequencing data. The Mantel test was used to find correlation between the MS-peak profile– and sequence-based distance matrices.

RESULTS

MS-Peak Profile Variability of HCV Amplicons With Known Genetic Complexity

PCR is a stochastic process and, when applied to a complex viral population, can produce DNA amplicons that differentially represent intrahost viral variants. Experimental conditions may additionally bias representation of viral populations in PCR fragments [43]. These considerations imply that repeat amplification from heterogeneous viral populations can result in variable representation of intrahost viral variants in PCR amplicons, thus potentially reducing reproducibility of the genetic evaluation based on a single PCR amplification.

To establish the reproducibility of MS-peak profiles, 2 groups of samples were tested: (1) 16 amplicons from HCV HVR1 clones that have been separated by PCR cloning (S) and (2) amplicons from 23 artificially generated mixtures (M), each of which comprised 8 HVR1 clones (SD1). Each of those 39 samples was PCR amplified 6–12 times, and MS-peak profiles from the generated replicas were acquired after MassCLEAVE™ (see Material and Methods). The mean difference (±standard error of the mean [SEM]) in MS-peak profiles generated from the S sample replicas was 1.11%±0.06%, indicating reproducibility under the experimental conditions used for detecting MS-peak profiles. However, the mean difference (±SEM) among the M sample replicas was 3.85%±0.35%, which was significantly greater than that among S sample replicas ($P < .0001$). The lower MS-peak profile reproducibility for complex mixtures of viral variants suggests that MS-peak profiles generated from repeated rounds of PCR can more accurately represent the intrahost HCV HVR1 complexity.

Each base-specific MassCLEAVE™ reaction generates a set of short RNA fragments composed of k nucleotides (k-mers) originating from transcripts generated from PCR amplicons. The k-mers are detected as mass peaks by MS (49). The number of the mass peaks reflects the number of different k-mers in the amplicon mixture. There are 2 sources of k-mer heterogeneity: (1) k-mers derived from a single PCR target-molecule and (2) k-mers created by nucleotide variations from different molecules. Genetic diversity is associated with the latter source. Thus, the number of k-mers can be used as a measure of the intrahost viral diversity represented in the PCR amplicons. Hence, the number of mass peaks derived from PCR fragments obtained from S and M samples should be different. Indeed, the mean number of peaks was significantly higher ($P = .0001$, by the multiresponse permutation procedure) in the M samples (mean, 124.31) than in the S samples (mean, 86.36), indicating that viral genetic diversity is reflected in the MS-peak profiles (Figure 2).

Intersample MS-Peak Profile Distances

To assess the usefulness of MS-peak profiles for estimating genetic relatedness among HCV strains, we tested 60 serum samples (OB samples) obtained from 14 HCV outbreaks that occurred in the United States during 2003–2008 [32–37]. Additionally, 25 serum samples (NR samples) were selected from cases of infection with HCV subtypes 1a and 1b that were epidemiologically not related to the outbreaks (SD1). On average, 9 replicas (range, 6–12 replicas) of HVR1 were amplified by PCR from each of the specimens (SD2).

The Hamming distance between MS-peak profiles from the OB and NR samples was used to measure genetic relatedness among HCV strains. The matrix of distances among all 793 PCR replicas obtained from samples of the 85 cases allowed

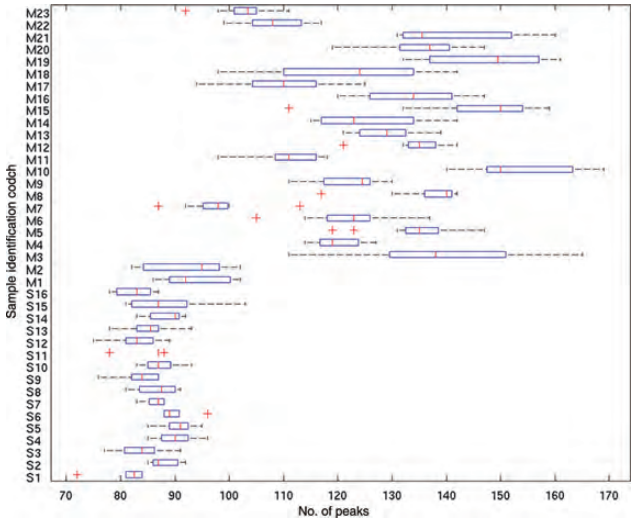

Figure 2. Number of mass peaks in polymerase chain reaction (PCR) replicas (n = 6–12) of samples with known complexity. The box encompasses the interquartile range of the data, with a red bar at the median value. S indicates samples from a single PCR clone, and M indicates a mix of 8 clones. Red crosses represent outliers, larger than q3 + [*w*(q3 − q1)] or smaller than q1 − [*w*(q3 − q1)], where q1 and q3 are the 25th and 75th percentiles, respectively. The maximum whisker length is *w*, with a default value of 1.5.

for a qualitative identification of each outbreak (Figure 3). In the OB group, interspecimen MS-peak profile distances obtained for cases in the same outbreak were similar to distances among PCR replicas and were smaller than among the NR group (Figure 4). The frequency distribution of pairwise distances among samples (Figure 4) shows that the NR MS-peak profiles are bimodal, reflecting distances within the subtypes and between different subtypes. The use of the minimum distances between PCR replicas from any pair of specimens allowed for a greater, albeit incomplete, separation between distributions of intraoutbreak distances in the OB group and distances in the NR group (Figure 4*B*).

Comparison With Sequence-Based Analysis

The agreement between sequence- and MS-peak profile-based measurements of HVR1 nucleotide-sequence diversity was low but significant (Pearson correlation, $r = 0.2637$; $P = .0188$). In contrast, the correlation between the sequence-based and MS-peak profile distances was very high (Mantel test, $r = 0.8361$; $P = .0001$). Scatterplots of sequence-based [45] and MS-peak profile–based distances (Figure 5) show clear differentiation among pairwise comparisons of OB samples from each outbreak, NR samples that carried HCV from the same subtype, and NR samples carrying HCV from different subtypes.

Neighbor-joining trees [44] were built using either sequence- or MS-peak profile–based distances. Both MS-peak profile– and sequence-based trees clustered correctly all specimens that were epidemiologically linked (Figure 6). The MS-peak profile–based clustering completely agreed with the sequence-based clustering and entirely resolved the overlap in the distance frequencies observed in Figure 4*B*, as none of the OB samples were found to group with NR cases.

AMOVA was used to quantify the level of differentiation between the OB and NR samples [46]. AMOVA partitioned the variance of the MS-peak profile– or sequence-based distances into 3 sources: (1) among the replicas of each specimen, (2) among OB specimens within an outbreak, and (3) among NR specimens (Figure 6*C*). It was found that 60.8% of all variance in the MS-peak profile distances was due to significant differences among the nonrelated samples from the OB groups and samples from the NR groups ($P = .0001$). The sequence-based method showed higher differences among nonrelated cases (73.73%), as can be observed in the smaller distances between outbreaks in the MS-peak profile similarity tree (Figure 6*B*). It also showed lower differences among the cases belonging to the same outbreak, which is reflected in the tight clusters in the sequence-based tree (Figure 6*A*).

Graphical Linkages of Related Cases

Visualization of genetic relatedness by use of tree models allows for qualitative clustering of HCV variants. However, genetic detection of transmissions does not require assessment of the degree of genetic differences among strains beyond establishing their genetic equivalence. Genetically related and unrelated HCV strains can be distinguished using a distance threshold. Analysis of the 300 pairwise MS-peak profile–based distances among the NR samples established 0.0506 as the minimal distance (Figure 4*B* and 4*C*), which was therefore established as threshold. Figure 7 shows the links among cases with a distance below the threshold. This simple method correctly identified all the samples involved in individual outbreaks, indicating 100% sensitivity in the genetic detection of transmissions. There were 1553 pairwise distances among cases belonging to different outbreaks, all of which were below the threshold. When the 1500 pairwise distances between OB and NR cases were considered, only 1 NR case had a MS-peak profile-distance to 1 OB case below the threshold, indicating 99.93% specificity of transmission detection.

DISCUSSION

Transmission is a fundamental viral property, knowledge of which is essential for understanding dissemination of infection and disease. As such, transmission detection is key to the surveillance of infectious diseases. However, molecular detection of transmissions is a very complex task. It involves not only identification of the virus, but, for virus that is highly mutable, also assessment of the genetic association among its variants.

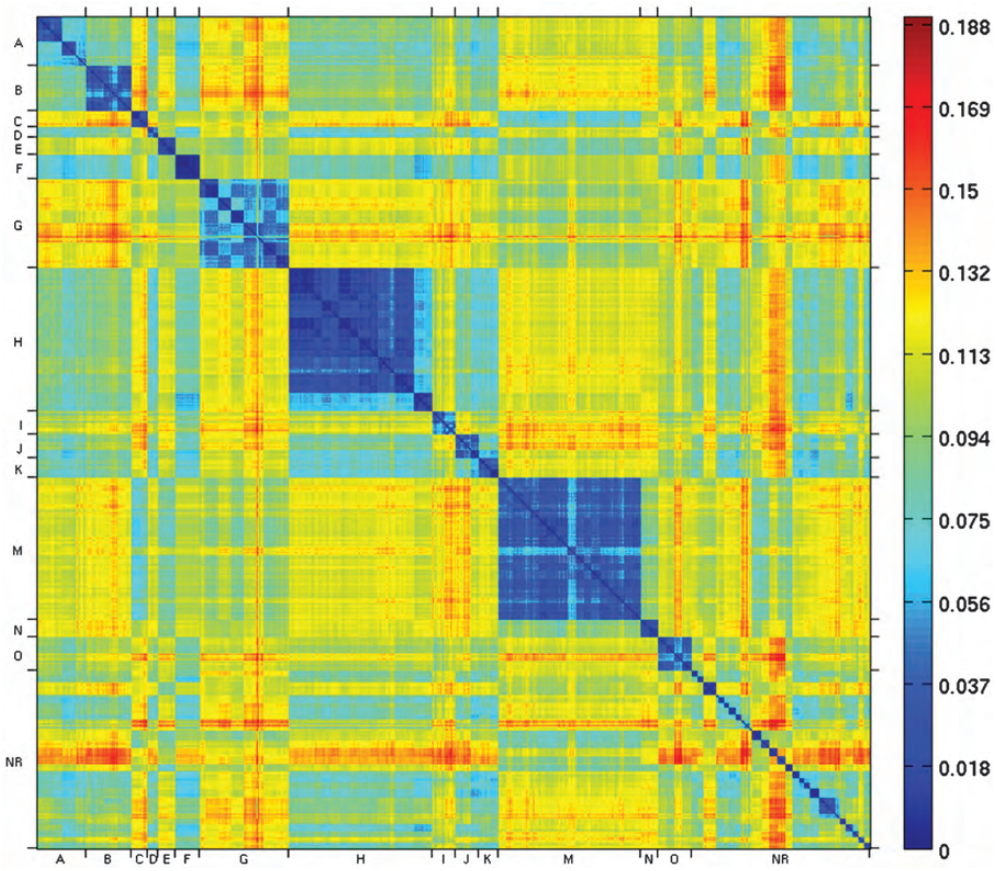

Figure 3. Heat map of the mass spectrometry–peak profile distance matrix. Distances between all subsamples (n = 793; average, 9.3 per patient) from OB (n = 60) and NR patients (n = 25) were plotted. The diagonal represents the comparison of each sample to itself. The color scale represents observed range of distances, for which colder colors signify shorter distances and warmer colors connote longer distances. Samples are grouped by specific outbreaks, denoted as capital letters (SD1) on both axes. Detailed information about the exact number of repeats per individual patient can be found in SD7. See Materials and Methods for descriptions of the OB and NR groups.

The major assumption of genome-based detection of transmissions is that the genetic composition of the viral strain that has been passed from one patient to another remains approximately "the same." Thus, the task of genome-based detection of transmission can be reduced to the evaluation of the genetic identity of viral strains without consideration of genetic relatedness beyond genetic equivalence. For HCV, this limited evaluation of genetic relatedness among variants can be best achieved by analyzing a sample of sequences obtained from the intrahost viral population, using short genomic regions [31, 32]. Consensus sequences of the regions can inadequately represent viral variants involved in transmission [30], especially when minority variants from the source establish the new infections in the recipients [36].

In this study, we explored the use of MS-peak profiles for identification of HCV transmission. MS-peak profiles generated using base-specific cleavage of RNA transcripts derived from PCR fragments [42, 43] are a very rich source of information about the nucleotide sequence and structure of the intrahost viral population. The approach used here detects a pattern of short RNA fragments, or k-mers. The k-mer structure of the MS data closely reflects the heterogeneity of viral populations. The number and variety of k-mers that can be derived from a short amplicon obtained from a single sequence variant is limited. Hence, mutations within the amplified region should generate additional k-mers that will likely result in independent MS peaks. This supposition is supported by the observation that the mean number of peaks was significantly higher in the M samples than in the S samples (Figure 2). Therefore, the genetic composition of the intrahost HCV population significantly contributes to MS-peak profile complexity. The genetic distances estimated using MS-peak profiles should be distinctly affected by sequence heterogeneity and composition of intrahost HCV populations and by their representation in the PCR products. Indeed, the sequence-based method showed average differences that were lower than MS-peak profile–based distances among cases belonging to the same outbreak, as shown by AMOVA and as evident from

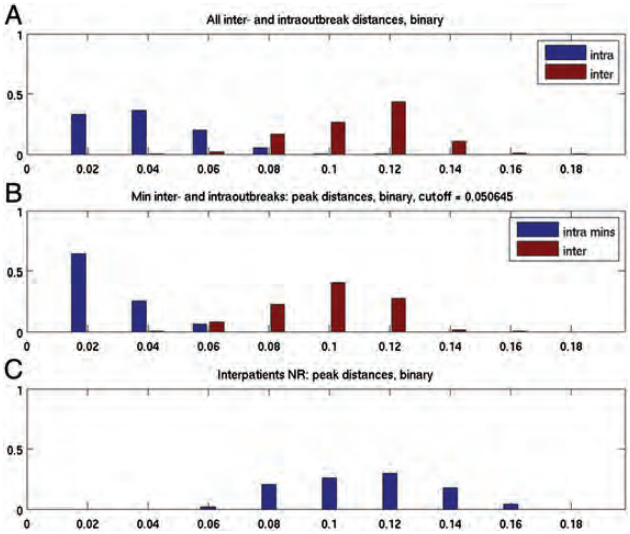

Figure 4. Histograms of pairwise mass spectrometry–peak profile (MSP) distances among samples. In all panes, the *x*-axis displays the distance, and the *y*-axis displays percentage of pairs found to have that distance. Distances within a patient and within related OB cases are denoted by blue, and distances between NR cases and denoted by red. *A*, All distances within OB and NR groups. *B*, Only minimum distances found within OB and NR groups. *C*, Distances between NR cases alone to establish the relatedness cutoff. See Materials and Methods for descriptions of the OB and NR groups.

comparing the sequence-based tree with the MS-based tree (Figure 6*A* and 6*B*).

Since the k-mers are derived from numerous RNA molecules representing the heterogeneous intrahost HCV

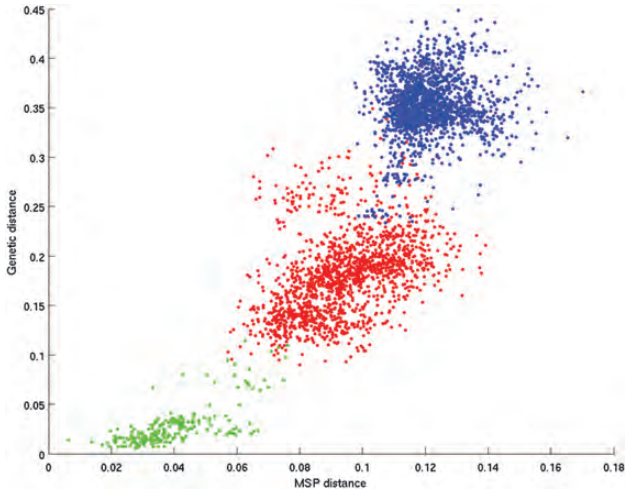

Figure 5. Scatterplot of sequence-based and mass spectrometry–peak profile (MSP)–based intersample distances. Pairwise comparisons of O samples within an outbreak are shown in green, comparisons among C samples belonging to the same subtype are in red, and comparisons of C samples of different subtypes are in blue.

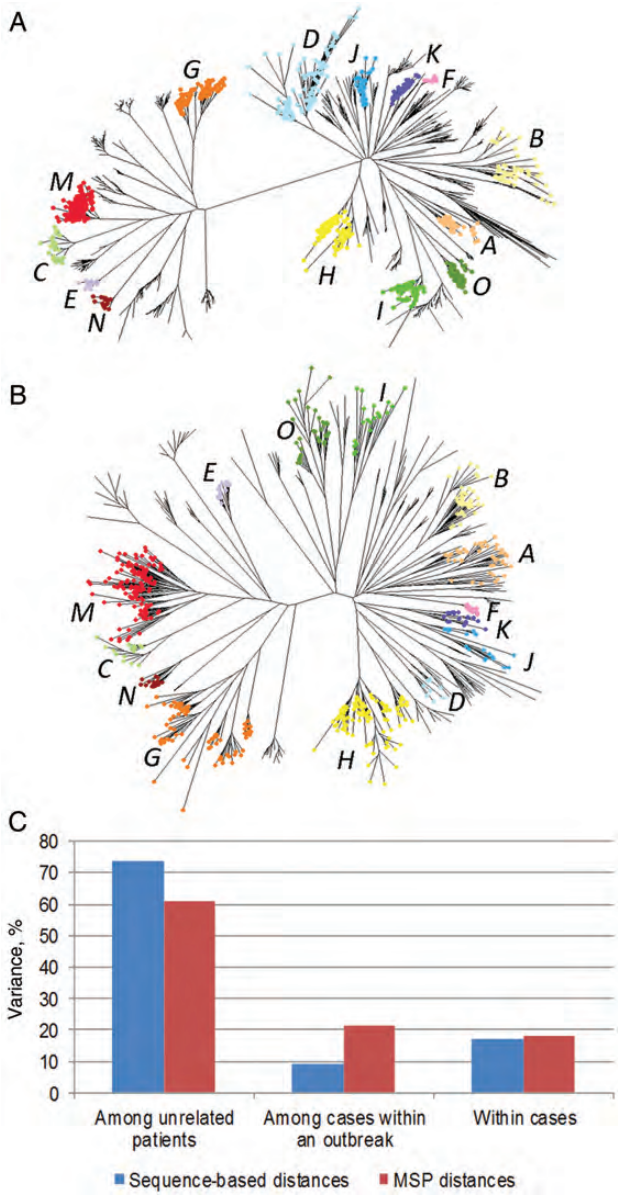

Figure 6. Similarity trees of hepatitis C virus (HCV) variants obtained from 14 outbreaks. *A*, Sequence-based neighbor-joining tree. *B*, Mass spectrometry–peak profile (MSP)–based neighbor-joining tree. Each outbreak is shown with a different color and identified by the same letter code as in Figure 3. Epidemiologically unrelated HCV variants are not colored. *C*, The variance in MSP- and sequence-based distances among samples is partitioned into 3 sources and expressed as a percentage among replicas of each sample, among OB HCV variants, and among NR HCV variants. See Materials and Methods for descriptions of the OB and NR groups.

population, an MS-peak profile reflects the primary structure of the genomic region and nucleotide variations among HCV variants in this region, thus combining all necessary genetic information for accurate molecular detection of HCV transmissions. Indeed, the MS-peak profiles obtained using the

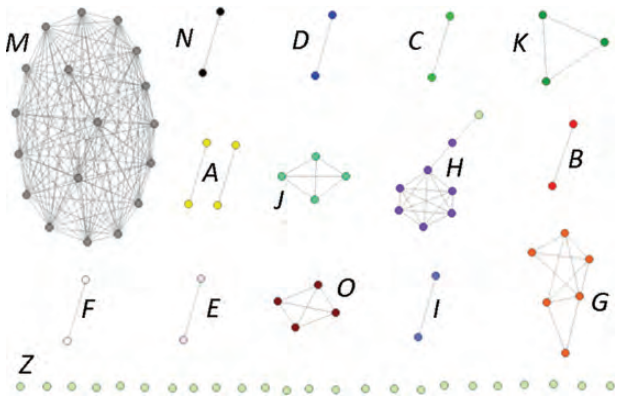

Figure 7. Linkage graphs of hepatitis C virus (HCV) variants. Links join HCV variants with a mass spectrometry–peak profile distance smaller than the relatedness cutoff (0.0506). Each outbreak is identified with a different color. Unlinked light-green dots at the bottom and 1 light-green dot linked to an outbreak (H) represent epidemiologically unrelated HCV variants.

HCV HVR1-containing PCR fragments were shown here to be efficient in establishing genetic equivalence among HCV variants involved in outbreaks. A small variation among PCR replicas from a single clone template (Figure 2) indicates significant reproducibility of the MS-peak profiles produced using the applied MS technology.

Owing to the stochastic nature of PCR, complex mixtures of the intrahost HCV variants may be inadequately represented in a single PCR product. The observed variation of peak numbers in the MS-peak profile among repeat PCR products obtained from mixtures of HVR1 clones (Figure 2) strongly supports this suggestion. The disproportionate representation of certain sequences in amplified fragments should result in a measurable variation of the k-mer composition in the MS-peak profile. Such deviations may potentially contribute to false-negative results, thus reducing sensitivity of the transmission detection. The data obtained in this study show that a more adequate representation of the intrahost HCV variants can be achieved using several PCR replicas derived from each test specimen.

Analyses conducted here reveal a fundamental similarity between MS-peak profile– and sequence-based distances for the genetic detection of transmissions (Figure 6). The separation between MS-peak profile-distances among genetically related and unrelated cases is clear and permits HCV variants involved in outbreaks to be discriminated from all nonoutbreak variants (Figure 4). Such genome-based detection of transmission does not require complete assessment of phylogenetic relationships among HCV variants and, therefore, can be achieved using a threshold and visualized using simple linkage graphs (Figure 7).

In conclusion, MS-peak profile–based detection of HCV transmissions as described here has been found to match the accuracy of sequence-based approaches. Moreover, whereas next-generation sequencing technologies require an extensive data preprocessing for identification and repair of sequence errors [40], the data generated from MS-peak profiles can be analyzed immediately upon acquisition. Because the approach developed in this study is more facile, less time-consuming, and considerably less costly than conventional sequencing, it is applicable to other pathogens and suitable for use in laboratories involved in routine surveillance of infectious diseases.

Notes

Financial support. This work was supported by intramural funding from the Centers for Disease Control and Prevention.

Potential conflicts of interest. C. C. is an officer and holds equity in Sequenom, a commercial entity that manufactures the platform used in this study. All other authors report no potential conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. Alter H. Discovery of non-A, non-B hepatitis and identification of its etiology. Am J Med **1999**; 107:16S–20.
2. Ray SC, Arthur RR, Carella A, Bukh J, Thomas DL. Genetic Epidemiology of Hepatitis C Virus throughout Egypt. J Infect Dis **2000**; 182:698–707.
3. Wasley A, Alter MJ. Epidemiology of hepatitis C: geographic differences and temporal trends. Semin Liver Dis **2000**; 20:1–16.
4. Hladik W, Kataaha P, Mermin J, et al. Prevalence and screening costs of hepatitis C virus among Ugandan blood donors. Trop Med Int Health **2006**; 11:951–4.
5. Ahmad N, Asgher M, Shafique M, Qureshi JA. An evidence of high prevalence of Hepatitis C virus in Faisalabad, Pakistan. Saudi Med J **2007**; 28:390–5.
6. Frank C, Mohamed MK, Strickland GT, et al. The role of parenteral antischistosomal therapy in the spread of hepatitis C virus in Egypt. Lancet **2000**; 355:887–91.
7. Armstrong GL, Wasley A, Simard EP, McQuillan GM, Kuhnert WL, Alter MJ. The prevalence of hepatitis C virus infection in the United States, 1999 through 2002. Ann Intern Med **2006**; 144:705–14.
8. Alter MJ, Kruszon-Moran D, Nainan OV, et al. The prevalence of hepatitis C virus infection in the United States, 1988 through 1994. N Engl J Med **1999**; 341:556–62.
9. Hepatocellular carcinoma—United States, 2001–2006. MMWR Morb Mortal Wkly Rep **2010**; 59:517–20.
10. Alter MJ. Transmission of hepatitis C virus—route, dose, and titer. N Engl J Med **1994**; 330:784–6.
11. Alter MJ. Epidemiology of hepatitis C virus infection. World J Gastroenterol **2007**; 13:2436–41.
12. Alter MJ. Healthcare should not be a vehicle for transmission of hepatitis C virus. J Hepatol **2008**; 48:2–4.
13. Patel PR, Thompson ND, Kallen AJ, Arduino MJ. Epidemiology, surveillance, and prevention of hepatitis C virus infections in hemodialysis patients. Am J Kidney Dis **2010**; 56:371–8.
14. National Institutes of Health Consensus Development Conference Statement: Management of hepatitis C 2002 (June 10–12, 2002). Gastroenterology **2002**; 123:2082–99.
15. Blackard JT, Shata MT, Shire NJ, Sherman KE. Acute hepatitis C virus infection: a chronic problem. Hepatology **2008**; 47:321–31.
16. Tellinghuisen TL, Evans MJ, von HT, You S, Rice CM. Studying hepatitis C virus: making the best of a bad virus. J Virol **2007**; 81:8853–67.
17. Argentini C, Genovese D, Dettori S, Rapicetta M. HCV genetic variability: from quasispecies evolution to genotype classification. Future Microbiol **2009**; 4:359–73.

18. Martell M, Esteban JI, Quer J, et al. Hepatitis C virus (HCV) circulates as a population of different but closely related genomes: quasispecies nature of HCV genome distribution. J Virol **1992**; 66:3225–9.
19. Okamoto H, Mishiro S. Genetic heterogeneity of hepatitis C virus. Intervirology **1994**; 37:68–76.
20. Forns X, Bukh J. The molecular biology of hepatitis C virus. Genotypes and quasispecies. Clin Liver Dis **1999**; 3:693–716, vii.
21. Sugitani M, Shikata T. Comparison of amino acid sequences in hypervariable region-1 of hepatitis C virus clones between human inocula and the infected chimpanzee sera. Virus Res **1998**; 56:177–82.
22. Vizmanos JL, Gonzalez-Navarro CJ, Novo FJ, et al. Degree and distribution of variability in the 5′ untranslated, E1, E2/NS1 and NS5 regions of the hepatitis C virus (HCV). J Viral Hepat **1998**; 5: 227–40.
23. Purcell R. The hepatitis C virus: overview. Hepatology **1997**; 26:11S–14.
24. Pawlotsky JM. Genetic heterogeneity and properties of hepatitis C virus. Acta Gastroenterol Belg **1998**; 61:189–91.
25. Massari M, Petrosillo N, Ippolito G, et al. Transmission of hepatitis C virus in a gynecological surgery setting. J Clin Microbiol **2001**; 39: 2860–3.
26. Bruguera M, Saiz JC, Franco S, et al. Outbreak of nosocomial hepatitis c virus infection resolved by genetic analysis of HCV RNA. J Clin Microbiol **2002**; 40:4363–6.
27. Widell A, Christensson B, Wiebe T, et al. Epidemiologic and molecular investigation of outbreaks of hepatitis C virus infection on a pediatric oncology service. Ann Intern Med **1999**; 130:130–4.
28. Brechot C. Hepatitis C virus: molecular biology and genetic variability. Dig Dis Sci **1996**; 41:6S–21.
29. Zeuzem S. Hepatitis C virus: kinetics and quasispecies evolution during anti-viral therapy. Forum (Genova) **2000**; 10:32–42.
30. Ramachandran S, Xia GL, Ganova-Raeva LM, Nainan OV, Khudyakov Y. End-point limiting-dilution real-time PCR assay for evaluation of hepatitis C virus quasispecies in serum: performance under optimal and suboptimal conditions. J Virol Methods **2008**; 151:217–24.
31. Bracho MA, Gosalbes MJ, Blasco D, Moya A, Gonzalez-Candelas F. Molecular epidemiology of a hepatitis C virus outbreak in a hemodialysis unit. J Clin Microbiol **2005**; 43:2750–5.
32. Patel PR, Larson AK, Castel AD, et al. Hepatitis C virus infections from a contaminated radiopharmaceutical used in myocardial perfusion studies. JAMA **2006**; 296:2005–11.
33. Thompson ND, Hellinger WC, Kay RS, et al. Healthcare-associated hepatitis C virus transmission among patients in an abdominal organ transplant center. Transpl Infect Dis **2009**; 11:324–9.
34. Thompson ND, Novak RT, Datta D, et al. Hepatitis C virus transmission in hemodialysis units: importance of infection control practices and aseptic technique. Infect Control Hosp Epidemiol **2009**; 30:900–3.
35. Thompson ND, Perz JF, Moorman AC, Holmberg SD. Nonhospital health care-associated hepatitis B and C virus transmission: United States, 1998–2008. Ann Intern Med **2009**; 150:33–9.
36. Fischer GE, Schaefer MK, Labus BJ, et al. Hepatitis C virus infections from unsafe injection practices at an endoscopy clinic in Las Vegas, Nevada, 2007–2008. Clin Infect Dis **2010**; 51:267–73.
37. Holmberg SD. Molecular epidemiology of health care-associated transmission of hepatitis B and C viruses. Clin Liver Dis **2010**; 14: 37–48.
38. Cody SH, Nainan OV, Garfein RS, et al. Hepatitis C virus transmission from an anesthesiologist to a patient. Arch Intern Med **2002**; 162:345–50.
39. Nainan OV, Lu L, Gao FX, Meeks E, Robertson BH, Margolis HS. Selective transmission of hepatitis C virus genotypes and quasispecies in humans and experimentally infected chimpanzees. J Gen Virol **2006**; 87:83–91.
40. Wijaya E, Frith MC, Suzuki Y, Horton P. Recount: expectation maximization based error correction tool for next generation sequencing data. Genome Inform **2009**; 23:189–201.
41. Dimitrova Z, Campo DS, Ramachandran S, Vaughan G, Ganova-Raeva L, Lin Y, Forbi JC, Xia G, Skums P, Pearlman B, Khudyakov Y. Evaluation of viral heterogeneity using next-generation sequencing, end-point limiting-dilution and mass spectrometry. In silico biology **2012**; 11:183–192.
42. Bocker S. SNP and mutation discovery using base-specific cleavage and MALDI-TOF mass spectrometry. Bioinformatics **2003**; 19(Suppl 1): i44–53.
43. Stanssens P, Zabeau M, Meersseman G, et al. High-throughput MALDI-TOF discovery of genomic sequence polymorphisms. Genome Res **2004**; 14:126–33.
44. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol **1987**; 4:406–25.
45. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol **1980**; 16:111–20.
46. Excoffier L, Smouse P, and Quattro J. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. Genetics **1992**; 131:479–491.