Molecular Epidemiology of Hepatitis C Virus in a Social Network of Injection Drug Users

Campbell K. Aitken, Rhonda F. McCaw, D. Scott Bowden, Samantha L. Tracy, Jenny G. Kelsall, Peter G. Higgs, Michael J. Kerger, Hoang Nguyen, and J. Nick Crofts

Epidemiology and Social Research Program and Centre for Harm Reduction, Macfarlane Burnet Institute for Medical Research and Public Health, Melbourne, and Victorian Infectious Diseases Reference Laboratory, North Melbourne, Victoria, Australia

Background. We aimed to measure the overlap between the social networks of injection drug users (IDUs) and the patterns of related hepatitis C virus (HCV) infections among IDUs.

Methods. A cohort of 199 IDUs (138 of whom were HCV RNA–positive IDUs) was recruited from a local drug scene in Melbourne, Australia, and was studied using social network analysis and molecular phylogenetic analysis of 2 regions of the HCV genome.

Results. Eighteen clusters of related infections involving 51 IDUs (37.0% of HCV RNA–positive IDUs) were detected; these clusters could be separated into 66 discrete pairs. Twelve (18.2%) of the 66 IDU pairs with related infections reported having previously injected drugs together; conversely, only 12 (3.8%) of the 313 pairs of HCV RNA–positive IDUs who were injection partners had strong molecular evidence of related infections. The social and genetic distances that separated IDUs with identical genotypes were weakly associated. Significant clusters of phylogenetically related sequences identified from core region analysis persisted in the analysis of the nonstructural 5a protein region. Genotyping and sequence analysis revealed 2 mixed-genotype infections.

Conclusions. Static social network methods are likely to gather information about a minority of patterns of HCV transmission, because of the difficulty of determining historical infection pathways in an established social network of IDUs. Nevertheless, molecular epidemiological methods identified clusters of IDUs with related viruses and provided information about mixed-genotype infection status.

Molecular epidemiological techniques have brought a new dimension to the study of the transmission and evolution of disease. With respect to HIV and hepatitis C virus (HCV), different combinations of genotyping, sequencing, and phylogenetic methods have been used to characterize historic epidemics [1, 2], to examine virus evolution [3, 4], to establish links between populations around the globe that have a high prevalence of HIV and HCV infection [5, 6], to aid in investigations of outbreaks [7, 8], and to evaluate postulated cases of transmission [9, 10]. Our understanding of the epidemiology of HIV has also benefited, in recent years, from the application of social network analysis, which has generated new insights about transmission and associated behavior. Social network methods have been used to demonstrate an association between the size and the density of the risk networks of injection drug users (IDUs) and the frequency of injection among IDUs [11] and to explain the differences in patterns of HIV infection between racial groups, on the basis of network measurements [12]. The network study of Morris et al. [13] showed that younger gay men with older partners gave significant impetus to the epidemic of HIV infection among gay men in the United States. Network methods have also been used to provide evidence that relatively small subnetworks of HIV-negative individuals may limit outbreaks in parent networks, allowing the prevalence of HIV to remain high but stable [14], and to show how specific network structures may facilitate HIV transmission [15].

Molecular epidemiological and social network methods, although stemming from vastly different scientific disciplines, have in common an ability to identify and
elucidate patterns in complex arrangements of humans (the former with respect to transmission of infection and the latter with respect to relationships and behaviors that permit transmission of infection). In the present article, we report a study in which social network methods were used to recruit IDUs in a suburb of Melbourne, Australia, and we compare the risk network data for these IDUs with the patterns of HCV infection delineated by molecular epidemiological methods. The primary objective of the present study was to establish the degree to which the genetic relatedness of HCV infection corresponded to the epidemiological map of potential routes of transmission represented by the risk networks of the IDUs. If the risk networks of the IDUs correspond well with patterns of infection, then the social networks should be capable of being exploited as loci for interventions aimed at reducing HCV transmission.

### Table 1. Genotype distribution among blood samples obtained from 138 hepatitis C virus (HCV) RNA-positive injection drug users.

<table>
<thead>
<tr>
<th>HCV genotype</th>
<th>No. (%) of samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>19 (13.8)</td>
</tr>
<tr>
<td>1a</td>
<td>40 (29.0)</td>
</tr>
<tr>
<td>1b</td>
<td>13 (9.4)</td>
</tr>
<tr>
<td>2b</td>
<td>3 (2.2)</td>
</tr>
<tr>
<td>3a</td>
<td>50 (36.2)</td>
</tr>
<tr>
<td>6a</td>
<td>7 (5.1)</td>
</tr>
<tr>
<td>7a</td>
<td>4 (2.9)</td>
</tr>
<tr>
<td>1a/3a</td>
<td>1 (0.7)</td>
</tr>
<tr>
<td>3a/7a</td>
<td>1 (0.7)</td>
</tr>
</tbody>
</table>

* Could not be further differentiated as genotype 1a or genotype 1b.

b Mixture of genotypes 1a and 3a.

### Table 2. Significant clusters identified by hepatitis C virus core and nonstructural 5a protein (NS5a) region phylogenetic analysis (bootstrapped ×1000).

<table>
<thead>
<tr>
<th>Genotype</th>
<th>No. of sequences</th>
<th>Groupingsa</th>
<th>No. of trees in the core region</th>
<th>No. of trees in the NS5a region</th>
</tr>
</thead>
<tbody>
<tr>
<td>1ab</td>
<td>61</td>
<td>MK08, JK01, MK23, and JK04</td>
<td>825</td>
<td>788</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HN26 and HN28</td>
<td>718</td>
<td>971</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MK31 and JK39</td>
<td>908</td>
<td>1000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MK58 and MK56</td>
<td>979</td>
<td>933</td>
</tr>
<tr>
<td></td>
<td></td>
<td>JK32 and JK07</td>
<td>975</td>
<td>980</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HN26 and MK23</td>
<td>692</td>
<td>728</td>
</tr>
<tr>
<td>1bb</td>
<td>31</td>
<td>MK31 and JK39</td>
<td>876</td>
<td>971</td>
</tr>
<tr>
<td></td>
<td></td>
<td>JK43 and JK14</td>
<td>719</td>
<td>761</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MK10, MK16, JK50, and JK25</td>
<td>1000</td>
<td>999</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MK52, JK09, MK05, MK13, HN38, HN03, and HN35</td>
<td>912</td>
<td>340c</td>
</tr>
<tr>
<td>2b</td>
<td>3</td>
<td>None</td>
<td>995</td>
<td>993</td>
</tr>
<tr>
<td>3a</td>
<td>51</td>
<td>JK54 and PH17</td>
<td>774</td>
<td>711</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MK54, MK52, JK58, MK04, JK35, and MK27</td>
<td>908</td>
<td>999</td>
</tr>
<tr>
<td></td>
<td></td>
<td>JK55 and PH25</td>
<td>973</td>
<td>761</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MK57 and MK55</td>
<td>994</td>
<td>583c</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HN06 and MK37</td>
<td>888</td>
<td>340c</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MK59 and MK65</td>
<td>852</td>
<td>999</td>
</tr>
<tr>
<td>7a</td>
<td>7</td>
<td>None</td>
<td>852</td>
<td>...</td>
</tr>
<tr>
<td>7a</td>
<td>5</td>
<td>HN22, HN12, and HN10</td>
<td>999</td>
<td>...</td>
</tr>
</tbody>
</table>

* Groupings shown are identification codes given to the blood samples/study participants and are based on interviewers’ initials and interview numbers.

b Includes samples genotyped as genotype 1 or genotype 1a/b.

c Not statistically significant.
allowed for the identification of individuals who were judged to be well connected in the local drug scene and for the eventual invitation extended to these individuals to become “seed cases” for the recruitment phase of the networks.

During the next 11 months (September 2001–July 2002), our researchers recruited 199 IDUs by following the social networks initially described by seed cases (recruitment ended in July 2002, to ensure that sufficient funds remained for data analysis). The criterion for the nomination of IDUs as network members was that they had injected drugs with the interviewee (at the same time and in the same location) at least once during the previous 6 months. Participants were reimbursed in Australian dollars (Au$30 [≈US$18 in early 2002]) for the time (typically, 30–40 min) and effort involved in completing the interview, and they were reimbursed an additional Au$10 for each network member who they introduced to our outreach workers. Informed consent was obtained from all research participants, and the human experimentation guidelines of the authors’ institutions were followed.

Venous blood samples or blood samples obtained by fin-

Figure 1. A, Hepatitis C virus (HCV) core gene tree for genotype 1a. B, HCV nonstructural 5a protein region gene tree for genotype 1a.
Figure 1. (Continued.)

Lab results for anti-HCV antibody by use of Murex anti-HCV antibody (version 4.0; Murex scientifi...
Biotech), for confirmation of the results. Irrespective of anti-HCV antibody status, all samples were tested for the presence of HCV RNA by use of the Cobas Amplicor HCV test (Roche). Fingerprick blood spots were obtained, in lieu of venous specimens, from 15 subjects who had badly damaged veins. The blood spots were eluted using Qiagen lysis buffer (QIAamp Viral RNA Mini Kit; Qiagen), and they were extracted using the semiautomated MagNA Pure LC instrument and dedicated Total NA Isolation Kit (Roche). Eluted material was diluted in the Roche Cobas Amplicor HCV test specimen diluent and was subjected to amplification as performed for the other blood samples.

PCR-positive blood samples were genotyped using a reverse-phase hybridization line-probe assay (LiPA [Versant HCV genotype assay; Bayer]), as described elsewhere [17]. For molecular studies, PCR amplification was performed using a nested in-house PCR with primers specific to the core region [18], and sequencing was performed on the PCR products by use of the ABI Prism Dye Terminator Cycle Sequencing Ready Re-
Figure 2. (Continued.)
action Kit (PE Applied Biosystems). Sequences were compared with sequences in the National Center for Biotechnology Information nucleotide sequence database, and the genotype obtained was compared with the LiPA result for each sample. Core sequences (307–328 bp, depending on genotype) were grouped according to genotype, and phylogenetic analysis was performed using the PHYLIP package [19]. For analysis of genotypes 2, 6, and 7, the core sequences were supplemented with genotype-specific control sequences sourced from the HCV sequence database of the Victorian Infectious Diseases Reference Laboratory (VIDRL; North Melbourne, Australia). For each pair of virus samples of the same genotype, genetic distances were calculated using the DNADIST program, for comparison with the network data and for the creation of HCV core gene trees by use of the Kimura 2-parameter method and the tree construction program NEIGHBOR. The PHYLIP program SEQBOOT was used to bootstrap (×1000) each set of genotype-specific sequence data, to estimate the significance of the clusters. HCV infections were identified as being significantly related if their bootstrap value was ≥70% [20].

Additional sequencing in the nonstructural protein 5a (NS5a) region (a more variable region) was performed on significant clusters for samples genotyped as genotype 1 and 3, for which specific primers were available [21]. Phylogenetic analysis of these NS5a sequences and of 20 genotype-specific control sequences from the VIDRL HCV sequence database (219 bp for genotype 1a sequences and 227 bp for genotype 3a sequences) was performed [9, 21], and the clusters that were identified were compared with those from the core sequence analysis.

**Statistical analysis.** Data obtained from interviews and blood tests were entered into a relational FileMaker Pro 5 database (Claris), which enabled simultaneous manipulation of related files that contained information about interviewees and their network members, and the data were exported to SPSS for analysis. The principal network measure used in this analysis is social distance—that is, the number of links that separate 2 IDUs, where a “link” denotes the naming of 1 IDU as an injection partner by another IDU, and/or the reverse. A pair of IDUs separated by 1 link (i.e., a social distance of 1) is known as a “dyad.” Social network measurements were generated by exporting data on links to UCINET [22] and by merging the resulting matrices into SPSS data files. Relationships between the social and genetic distances that separated the study participants were investigated using regression analysis.

**RESULTS**

**Network characteristics.** Of the 199 IDUs interviewed, 197 were members of 1 large, connected network component that included 577 dyadic relationships (therefore, a path that consisted of a finite number of injection-associated links could be drawn between any 2 IDUs). Connections between initially distinct network subcomponents surfaced as recruitment for the study progressed, despite the deliberate recruitment of seed cases from a broad spectrum of IDU groups identified during our initial fieldwork. The remaining 2 IDUs were “isolates”—that is, they did not nominate any other interviewed IDUs as network members, and they were nominated by none of the other interviewed IDUs. IDUs nominated a mean of 4.6 network members each (range, 1–10), and a mean of 3.3 network members (range, 1–10) were interviewed. IDUs were linked to a mean of 6.2 network members (range, 1–23).

**HCV test data.** Of 198 blood samples, anti-HCV antibody was detected in 172 (86.9%) of the samples, and HCV RNA was detected in 138 (69.7%) of the samples. Twenty-one blood samples were negative both for anti-HCV antibody and for HCV RNA. One blood sample was found to be anti-HCV antibody negative and HCV RNA positive, indicative of early infection. Two blood samples had indeterminate results of testing for anti-HCV antibody and were HCV RNA negative, and 2 blood samples obtained by fingerprick were antibody negative and unsuitable for PCR.

Genotype distribution among the 138 HCV RNA–positive IDUs (table 1) was similar to that reported for the Australian HCV-infected community, with HCV genotypes 1, 1a, and 1b (52.2% of 138 IDUs) and genotype 3a (36.2% of 138 IDUs) found to be the predominant types [17]. The genotype determined by the LiPA was compared with the genotype determined by core sequencing. The genotypes matched for all but 6 IDUs. For the blood samples of 4 IDUs, the genotype was correctly identified, by core sequence analysis, as HCV genotype 7a (this genotype had been mistyped as genotype 1b by LiPA) [18]. The blood sample of 1 IDU was found to have a mixture of genotypes 1a and 3a, according to LiPA, core, and NS5a region sequencing. For the blood sample of the other IDU, the genotype identified by LiPA was genotype 3a, but the dominant sequence produced by core sequencing was genotype 7a. The underlying genotype 3a sequence was apparent, and the presence of genotype 3a was confirmed by sequence analysis of the NS5a region, by use of type-specific primers. Type-specific primers for genotype 7a were not available for sequence analysis of the NS5a region.

**Phylogenetically related infections.** Phylogenetic analysis of 138 core region sequences from 6 genotypes identified 18 clusters of related infections (12 pairs, 2 triplets, 2 quadruplets, 1 sextuplet, and 1 septuplet) that involved 51 individual IDUs (table 2). Thus, of the 138 HCV RNA–positive IDUs, nearly 40% had an HCV infection that was genetically related to that of at least 1 other IDU in our cohort. Figures 1 and 2 show the core and NS5a region gene trees, respectively, with bootstrap values indicating the degree of relatedness for genotypes.
1a and 3a, which are the 2 most common genotypes. For genotype 1a, all 5 significant clusters in the core region were confirmed in the NS5a region. For genotype 3a, the 7 significant clusters in the core region had identical groupings in the NS5a region; however, 2 of the 7 clusters deemed to be significant on the basis of core region sequences did not reach the 70% bootstrap value (the threshold for significance) when NS5a region sequences were analyzed.

To examine the social links between the related infections, the 18 clusters of related infections were split into their 66 constituent pairs (i.e., 12 + [2*3!/2!1!] + [2*4!/2!2!] + [1*6!/4!2!]). In social network terms, only 12 (18.2%) of the 66 pairs with genetically related infections were dyads; therefore, only 12 (3.8%) of the 313 HCV RNA–positive social dyads in the network had related infections. Social distance was infinite for 9 related infection pairs that included 1 of the isolates described earlier; distances of 2–3 links (mean distance, 5.6 links) separated IDUs who comprised the remaining 45 (i.e., 66 – 12 – 9) pairs. Figure 3 depicts 4 clusters of related infections and the social relationships of the IDUs involved.

In the first triplet (genotype 3a), there is 1 dyadic relationship, but the shortest distance between either member of that pair and the third IDU is 6 links; in the second triplet (genotype 7a), there is 1 dyadic relationship, and the third IDU is 2 links distant from both members of that dyad. In the first quadruplet (genotype 1b), 2 IDUs are directly linked to a third IDU but not to each other, and all IDUs are at a distance of 2–5 links from the fourth IDU; the second quadruplet (genotype 1a) consists of a “3-clique” (i.e., 3 IDUs with all 3 possible links present) and an isolate. In the sextuplet (genotype 3a) and septuplet (genotype 1b), among 15 and 21 discrete pairs of IDUs with related infections, there were only 2 and 1 dyads, respectively. The sole dyad in the sextuplet included the only IDU with seroconversion in the present study (with seroconversion determined by a positive PCR result and a negative anti-HCV antibody test result), who reported (1) injecting with his dyadic partner 2–3 times/day over a period of 1 month and (2) injecting with the partner’s used needle 5 times, thereby strongly implicating that IDU as the source of infection.

**Genetic distance and social distance.** To examine the correspondence between virus relatedness and social distance, the social distances between all pairs of IDUs with known genotypes were compared with the genetic relatedness of their viral sequences, within each genotype group, by use of regression analysis. (Social distances for each genotype group were normally distributed, but genetic distances were not; log-normal and other transformations were evaluated, but none improved on the linear model). The 2 isolates (who neither named nor were named by other cohort members as injection partners) were removed from the analyses, because they were at an infinite social distance from all other IDUs in the data set, and no regression was performed for genotype 2b, because only 3 IDUs with that genotype were recruited for the study. IDUs with a genotype that could not be definitively identified as genotype 1a or 1b were included in both analyses, as were the 2 IDUs with mixed-genotype infections. Results of regression analysis are presented in table 3. Regressions for 4 of 5 genotypes produced statistically significant associations, all of which had low \( r^2 \) values; 2 associations are inverted, predicting decreasing social distance with increasing genetic distance. Confidence intervals are tight around the regression constants, which vary from 3.78 to 7.39 for the 4 statistically significant associations; therefore, for genetic distances of zero (which denote identical or nearly identical viral sequences), the regression constants accurately represent estimated social distances, thereby confirming the weakness of the associations. If correspondence between social and molecular distance had been very high, the regression constants would be close to 1, so that a genetic distance of zero would produce an estimate of social distance approaching 1 (denoting a dyadic relationship).

**Genetic distance and duration of injection drug use history.** We expected that related infections would be more often detected among IDUs in the early stages of their injection drug use history, because their social networks are more likely to include the source of their HCV infections, and because pairs of infections are more likely to remain phylogenetically similar. To test this idea, we compared the mean duration of the injection drug use history of IDUs who were \( n = 51 \) or were not \( n = 87 \) members of related infection pairs. The mean durations of injection drug use histories for these groups were virtually identical (10.65 years [for IDUs who were members of

![Figure 3. Four clusters of related infections involving multiple injection drug users, according to genotype. Solid arrows, dyadic relationships (social distance, 1); broken arrows, finite paths of >2 links (social distance, >1).](image-url)
related infection pairs] vs. 10.63 years [for IDUs who were not members of such pairs]; \( p = .988 \), refuting our hypothesis.

**DISCUSSION**

The present study is the first application of combined social network and molecular phylogenetic techniques to the epidemiology of HCV. The results of the study confirm that molecular epidemiological and social network methods can overlap in their descriptions of pathways of HCV transmission. Variability in the core region proved adequate for the identification of significant clusters of related infections. Further analysis in the more variable NS5a region confirmed the presence of clusters of infections of genotypes 1a and 3a, the major genotypes present in the cohort. Of IDUs who had detectable HCV RNA levels, nearly 40% had an HCV infection that was phylogenetically related to that of at least 1 other IDU in the cohort. More than 18% of the IDU pairs with related HCV infections were identified in the social network data as injection partners (i.e., as dyads). Of interest, all but 2 of the IDUs interviewed were members of 1 large network of IDUs, a scenario that has obvious implications for our ability to track patterns of HCV infection through the social network and molecular phylogenetic techniques to the epidemiology of HCV. The present study is the first application of combined social network and molecular phylogenetic techniques to the epidemiology of HCV. The results of the study confirm that molecular epidemiological and social network methods can overlap in their descriptions of pathways of HCV transmission. Variability in the core region proved adequate for the identification of significant clusters of related infections. Further analysis in the more variable NS5a region confirmed the presence of clusters of infections of genotypes 1a and 3a, the major genotypes present in the cohort. Of IDUs who had detectable HCV RNA levels, nearly 40% had an HCV infection that was phylogenetically related to that of at least 1 other IDU in the cohort. More than 18% of the IDU pairs with related HCV infections were identified in the social network data as injection partners (i.e., as dyads). Of interest, all but 2 of the IDUs interviewed were members of 1 large network of IDUs, a scenario that has obvious implications for transmission of disease and that has not been observed in previous studies of IDU networks [23–25].

The low level of correlation between social network and phylogenetic data may be ascribed to a variety of reasons. Our ability to track patterns of HCV infection through the social networks of IDUs was diminished by the practical constraints of finding and interviewing specific IDUs in a busy and rapidly shifting street-drug scene. As other authors have noted [15], the situation is one of incomplete capture of a complex picture. In addition, the epidemic nature of HCV in a high-risk population, such as IDUs, hinders reconstruction of infection patterns. There is a high prevalence of exposure to HCV among Australian IDUs (64% of participants in the 2001 national Needle and Syringe Program Survey tested positive for anti-HCV antibody), and exposure frequently occurs early in the injection drug use history of IDUs [26]. The IDUs investigated in the present study had been injecting drugs for just >10 years, on average (and 70% had been injecting drugs for ≥5 years), and the prevalence of anti-HCV antibody was already >85%; thus, for the majority of IDUs, their first exposure to HCV is likely to have occurred years before the interview was conducted, which is well beyond any reasonable period for recall of complex social network information. Meanwhile, we were unavoidably reliant on a much more recent period in the injection drug use history of our study participants, to have a good chance of recruiting their network members. The fact that a substantial minority of participants were living in transient accommodation when they were interviewed, having recently arrived from other parts of Victoria or Australia, further reduced our ability to make social network connections and to identify related infections. In contrast, molecular studies were able to definitively identify related HCV infections in clusters of IDUs, some of which had corresponding social connections.

Two blood samples were identified as showing evidence of mixed-genotype infection after comparison of the genotype results of LiPA with those of sequencing. It is apparent from these results that not all genotype mixtures may be detected using standard assays in which only the dominant genotype is detected [27, 28]—a phenomenon with obvious implications for the provision of treatment for HCV infection to current or former IDUs. This phenomenon may also be at least part of the explanation for the absence, in the present study, of any difference between the mean durations of injection drug use history of IDUs with or without related infections.

Investigation of the injection drug use histories of the 21 individuals who showed no clinical evidence of HCV infection revealed instances of needle-sharing with known HCV-positive individuals on multiple occasions. The absence of infection could be due to several factors, including innate immunity of the host [29]. Further study of this phenomenon may shed light on the mechanisms leading to initial development of infection, chronicity, and clearance.

In conclusion, the present study involved the collection and analysis of data regarding a single large social network of IDUs. The low correlation between social network data and phylogenetic data highlights the practical difficulties involved in map-
ping networks of (frequently) highly mobile IDUs in whom HCV infection typically occurs very early in their injection drug use history and among whom HCV infection is invariably highly prevalent. Another possible confounder is that infections with multiple HCV genotypes are not always identified using the LiPA. Although our data show that the risk networks of IDUs therefore are not a reliable guide to actual patterns of HCV infection, the identification of a subgroup of anti-HCV–negative IDUs with high-risk injection behavior warrants further investigation. Molecular epidemiology was shown to be a powerful tool with which to identify related infections in IDUs. It can provide insights into the spread of HCV infection in a limited population of IDUs; however, because of the rapid evolution of the virus and because of the difficulties associated with studying this population, the use of molecular epidemiological methods is probably best restricted to the tracking of recent infections.

Acknowledgment

We thank Dr. Graham Byrnes (Department of Mathematics and Statistics, University of Melbourne, Parkville, Victoria, Australia) for his advice on the interpretation of the phylogenetic data in the present study.

References

16. Australian Bureau of Statistics. 2001 census of population and housing. Available at: http://www.abs.gov.au/Ausstats/abs@.nsf/ddc/9b4b92657325ca2563e000dbafa0f0a5166ebcb6b2ca2566b0f0005433/