Colonial History and Contemporary Transmission Shape the Genetic Diversity of Hepatitis C Virus Genotype 2 in Amsterdam

Peter V. Markov,1 Thjis J. van de Laar,2,4 Xiomara V. Thomas,4 Sem J. Aronson,4 Christine J. Weegink,4 Guido E. van den Berk,4 Maria Prins,4,9 Oliver G. Pybus,2 and Janke Schinkel2

Department of Zoology, University of Oxford, Oxford, United Kingdom;2 Cluster of Infectious Diseases, Public Health Service, Amsterdam, The Netherlands;3 Department of Medical Microbiology and Infection Control, VUMC University Medical Center, Amsterdam, The Netherlands;4 Department of Medical Microbiology, Academic Medical Center/University of Amsterdam, Amsterdam, the Netherlands;5 Department of Gastroenterology and Hepatology, Academic Medical Centre/University of Amsterdam, Amsterdam, The Netherlands;5 Department of Internal Medicine, Onze Lieve Vrouwe Gasthuis, Amsterdam, The Netherlands;5 and Department of Internal Medicine, Academic Medical Centre/University of Amsterdam, CINMA, Amsterdam, The Netherlands

Evolutionary analysis of hepatitis C virus (HCV) genome sequences has provided insights into the epidemic history and transmission of this widespread human pathogen. Here we report an exceptionally diverse set of 178 HCV genotype 2 (HCV-2) isolates from 189 patients in Amsterdam, comprising 8 distinct HCV subtypes and 10 previously not recognized, unclassified lineages. By combining study subjects’ demographic information with phylogeographic and molecular clock analyses, we demonstrate for the first time that the trans-Atlantic slave trade and colonial history were the driving forces behind the global dissemination of HCV-2. We detect multiple HCV-2 movements from present-day Ghana/Benin to the Caribbean during the peak years of the slave trade (1700 to 1850) and extensive transfer of HCV-2 among the Netherlands and its former colonies Indonesia and Surinam over the last 150 years. The latter coincides with the bidirectional migration of Javanese workers between Indonesia and Surinam and subsequent immigration to the Netherlands. In addition, our study sheds light on contemporary trends in HCV transmission within the Netherlands. We observe multiple lineages of the epidemic subtypes 2a, 2b, and 2c (together 67% of HCV-2 infections in Amsterdam), which cluster according to their suspected routes of transmission, specifically, injecting drug use (IDU) and contaminated blood/blood products. Understanding the epidemiological processes that generated the global pattern of HCV diversity seen today is critical for exposing associations between populations, risk factors, and specific HCV subtypes and might help HCV screening and prevention campaigns to minimize the future burden of HCV-related liver disease.
side Africa, including the Caribbean (21, 39), Indonesia (41), and Vietnam (26). These lineages fall within the greater African diversity, but the extent of their divergence suggests dissemination from Africa before the rise of global travel and modern medicine (20). During the 20th century, transmission via contaminated blood, blood products, invasive procedures, and IDU eventually resulted in the global dispersal of HCV-2, particularly subtypes 2a and 2b, which cause the vast majority of HCV-2 infections in developed countries (9, 11, 44). Interestingly, two recent studies from France revealed a wide variety of HCV-2 subtypes, including among the native French population (5, 40).

Here we report the first study of the diversity and molecular epidemiology of HCV-2 in the Netherlands. Using phylogenographic and molecular clock approaches, we investigate the history and global dissemination of HCV-2. Our results indicate that the contemporary distribution of HCV-2 in the Netherlands is a result of both current transmission trends and historical events inextricably bound up with Dutch colonial history. Further, our findings help to interpret the complex patterns of HCV-2 diversity observed elsewhere and directly address the previously proposed hypothesis that the trans-Atlantic slave trade was instrumental to the introduction of HCV to the Americas (20). Increased understanding of past and present HCV transmission events helps to better define risk factors and improve the detection of undiagnosed infected individuals, thereby preventing or ameliorating future HCV-related liver disease (2).

### MATERIALS AND METHODS

#### Study population.
The study population included all patients diagnosed with a HCV-2 infection at the Academic Medical Center (AMC) or the Amsterdam Public Health Service (APHIS) from 2000 through 2009. AMC participants included patients who attended the hematology outpatient clinics of either the AMC or a large inner city hospital for which HCV genotyping is performed at the AMC. APHIS participants were members of surveys performed to monitor the general population or groups at risk for sexually transmitted and blood-borne infections; all signed an informed consent. For each HCV-2-infected participant, data on age, country of birth, HIV status, and plausible route of infection were collected, if available. HCV risk groups were classified into four ranked categories: (i) injecting drug users (IDUs), (ii) recipients of blood or blood products, (iii) migrants from non-Western countries (countries in Asia, Africa, and Latin America), (iv) other exposures (including, e.g., nosocomial, perinatal, and sexual/household exposure) (Table 1), and (v) unknown. Individuals belonging to more than one risk group were classified into the highest-ranked risk category.

<table>
<thead>
<tr>
<th>RNA isolation, reverse transcription, PCR, and sequencing.</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA isolation, reverse transcription, PCR, and sequencing were performed as described previously (6). Transcribed cDNA was used as the input for two separate HCV PCR assays targeting the NS5B region, resulting in the amplification of two overlapping NS5B fragments. Fragment 1 (340 nucleotides [nt]; nt 8001 to 8340) and fragment 2 (421 nt; nt 8288 to 8709) were concatenated, resulting in an NS5B sequence of a total length of 709 nt. Sera and RNA isolates were stored at −80°C.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phylogenetic analyses.</td>
<td>Ancestral relationships among HCV isolates were inferred by reconstructing phylogenetic trees of our NS5B sequences, together with confirmed and provisionally assigned HCV-2 reference sequences from the HCV sequence database (15). Maximum likelihood (ML) phylogenies were estimated using the Hasegawa-Kisho-Yano nucleotide substitution model plus a gamma-distribution model of among-site rate heterogeneity (HKY-Γ) (13) as implemented in GARLI v0.95 (50). ML phylogenies were midpoint rooted, and the statistical robustness of phylogenetic branches was tested by ML bootstrapping with 100 replicates.</td>
<td></td>
</tr>
</tbody>
</table>

### Table 1. Epidemiological characteristics of the study population

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>No. of patients (n = 189)</th>
<th>% of patients</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>65</td>
<td>34</td>
</tr>
<tr>
<td>Male</td>
<td>124</td>
<td>66</td>
</tr>
<tr>
<td><strong>Country or region of birth</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Western</td>
<td>111</td>
<td>59</td>
</tr>
<tr>
<td>The Netherlands</td>
<td>88</td>
<td>46</td>
</tr>
<tr>
<td>Western Europe, USA, Australia</td>
<td>20</td>
<td>11</td>
</tr>
<tr>
<td>Eastern Europe</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Non-Western</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surinam</td>
<td>45</td>
<td>24</td>
</tr>
<tr>
<td>Morocco</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>Indonesia</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>Other</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>Unknown</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td><strong>Transmission risk factor</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IDU</td>
<td>76</td>
<td>40</td>
</tr>
<tr>
<td>Blood transfusion</td>
<td>33</td>
<td>18</td>
</tr>
<tr>
<td>Non-Western country of birth</td>
<td>48</td>
<td>25</td>
</tr>
<tr>
<td>Other</td>
<td>14</td>
<td>7</td>
</tr>
<tr>
<td>Unknown</td>
<td>18</td>
<td>10</td>
</tr>
<tr>
<td><strong>HIV status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>12</td>
<td>6</td>
</tr>
<tr>
<td>Negative</td>
<td>110</td>
<td>58</td>
</tr>
<tr>
<td>Unknown</td>
<td>67</td>
<td>35</td>
</tr>
</tbody>
</table>

### Notes

- Sixteen of 33 participants who reported to have received blood or blood products received transfusion in a non-Western country: 11 in Surinam, 2 in Morocco, 1 in Algeria, 1 in the Dominican Republic, and 1 in the United Arab Emirates.
- Other reported routes of transmission included sexual transmission (6 participants), nosocomial transmission (3), unhygienic tattooing (2), perinatal transmission (2), and needle-stick injury (1).

### Molecular clock and phylogeographic analyses.
The long-term evolutionary history and global dissemination of HCV-2 were assessed by reconstructing a relaxed molecular clock phylogeny using the Bayesian Markov chain Monte Carlo (MCMC) approach and the HKY-Γ substitution model implemented in BEAST v1.4.6 (8). The relaxed molecular clock model was preferred over a strict clock after a formal model test resulting in an ln Bayes factor of 96.5 (38). Reference sequences with known sampling locations and dates were chosen such that the full genetic diversity of HCV-2 was evenly represented; only four representatives of each epidemic subtype (2a, 2b, and 2c) were retained. To maximize sample size, we used all reference sequences that were homologous to any part of the concatenated NS5B Amsterdam sequences and which spanned at least 290 nt of the concatenated alignment. Based on previous estimates, the inferred rate of nucleotide evolution for the NS5B sequences was 0.0005 substitutions per site per year (23, 30). Multiple MCMC chains were concatenated, and the resulting alignments were used to infer the molecular clock phylogeny.
were calculated and output files were combined using LogCombiner 1.5.4 with an appropriate burn-in removed. MCMC convergence and effective sample sizes were checked using Tracer v1.5 (http://beast.bio.ed.ac.uk). A summary phylogeny (maximum clade credibility [MCC] tree) was generated using TreeAnnotator v1.5.4 (http://beast.bio.ed.ac.uk). The geographic states of ancestral lineages were reconstructed using a parsimony approach (37), and the MCC phylogeny was annotated using FigTree (http://tree.bio.ed.ac.uk). The parsimony analysis suggested that there were insufficient lineage migration events to justify the use of parametric phylogeographic methods (16). The following criteria were used to define geographic clusters: (i) phylogeographic (lineage movement) event at the root of the clade, (ii) clade support of ≥80, (iii) at least two-thirds of the sequences forming the clade being from one location (excluding sequences that form a subcluster resulting from a subsequent phylogeographic event).

**Dating of lineage migration events.** Viral lineage migration events were estimated using the two nodes of the phylogenetic branch on which the change in location state occurs: the descendant (X) and ancestral (Y) nodes. The estimated dates of nodes X and Y (together with their 95% highest posterior density [HPD] credible regions) were obtained from the MCC tree and define a time window during which lineage movement took place. The most conservative (i.e., widest) estimate of the lineage migration date range is given by the more recent 95% HPD limit of node X and the older 95% HPD of node Y. All subtrees defined by node X were highly supported; however, many of the subtrees defined by node Y were not. When the support of node Y was <50%, the preceding node was used to define the older time limit of the migration interval. The older 95% HPD of this node was then used as the most ancient plausible date of lineage movement.

**Testing for association between geography and phylogeny.** To test for an association between phylogeographic clustering and geographic location, BaTS (28) was used to perform a test of phylogeographic structure. The robustness of phylogeographic clustering was assessed by comparing the calculated posterior distribution of the parsimony score (PS) and association index (AI) statistics of tip-trait association (37, 49) with their null distributions under the hypothesis of no phylogenetic structure. BaTS explicitly includes phylogenetic uncertainty, and therefore any significant result obtained is robust to the sequence length used (28).

**Trans-Atlantic slave trade historic data.** The number of individuals transported across the Atlantic Ocean during the entire period of trans-Atlantic slave trade was downloaded from the Trans-Atlantic Slave Trade Database (47). The adjusted number of captives, as well as the adjusted number of captives disembarked at the Dutch Guianas and in total in the New World, were downloaded in quarter-century periods.

**Nucleotide sequence accession numbers and reference sequences used to determine viral genotype.** All 178 HCV NSSB sequences obtained in this study have been submitted to GenBank. This includes 167 concatenated NSSB sequences (709 nt) with accession numbers JF722454 through JF722620, plus 11 partial sequences (NSSB fragment 1 or 2 only; see above) with accession numbers DQ238634, DQ238635, DQ238659, DQ238673, and JQ746501 through JQ746507. The subtype of each sequence was determined in a preliminary analysis using confirmed and provisionally assigned HCV-2 reference sequences from the HCV sequence database (15, 36). These included D009944 (2a), D10988 (2b), D50409 (2c), D49761 (2e), D49777 (2f), DQ155565 (2i), D86530 (2j), AB031663 (2k), and D86529 (2q), spanning the complete 709 nucleotides of our concatenated NSSB fragment; and L29634 (2d), L48494 (2l), AY434117 (2m), L44602 (2n), L38373 (2o), L46601 (2p), and EF116040 (2r), for the first 341 nucleotides (fragment 1) only.

**RESULTS**

**Study population.** In total, 189 HCV-2 infected individuals were identified: 129 were diagnosed at the AMC and 60 at the APHS. Study subjects had a median age of 47 years (interquartile range: 41 to 57 years), 66% were male, and only 59% were born in Europe, North America, or Australia (Table 1). The predominant underlying risk factors for acquiring HCV-2 infection were IDU (40%), being born in a non-Western country (mostly Surinam; 25%), and a history of blood transfusion (17%). Interestingly, 16/33 (48%) of the HCV-2 infected individuals with a history of blood transfusion were transfused in a non-Western country. Only 122/189 (65%) of the study subjects were tested for HIV. Of the 12 HIV-positive individuals, 10 were current or previous IDUs and 2 were men who have sex with men (MSM).

**RNA isolation, reverse transcription, PCR, and sequencing.** Amplification and sequencing of the concatenated 709-bp NSSB fragment were successful for 167/189 (88%) patients. For an additional 11 patients, NSSB sequences were obtained for either NSSB fragment 1 only (n = 2) or NSSB fragment 2 only (n = 9). Hence, HCV genotyping and phylogenetic analysis could be performed for 178/189 (94%) of patients based on the availability of (partial) NSSB sequence data. For the remaining 11 patients, no NSSB sequence data were obtained (Table 2). Of these, 7 patients had no serum or plasma available, and for 4 patients, all born in Surinam, amplification of both NSSB fragment 1 and NSSB fragment 2 failed.

**Phylogenetic analyses.** A phylogenetic tree was constructed for our study sequences plus established reference sequences from the HCV sequence database (Fig. 1). Amsterdam sequences fall into 13 distinct monophyletic clades, 1 homologous pair and 5 singleton sequences, representing 8 known HCV-2 subtypes and 10 phylogenetically distinct and yet unclassified HCV-2 lineages (Table 2). All monophyletic clades were supported by high bootstrap scores (>70%; blue circles).

The majority of Amsterdam sequences (61%) belonged to the three epidemic subtypes 2a, 2b, and 2c. Individuals infected with these subtypes were mostly born in the Netherlands or other Western countries and reported a history of either IDU or blood transfusion (Table 2). We estimated separate phylogenies of subtypes 2a, 2b, and 2c that included all isolates from Amsterdam plus all available reference sequences. Figure 2 depicts the ML phylogeny of HCV-2a. In addition to numerous isolates from Asia and an HCV-2a cluster from countries of the former Soviet Union, this phylogeny features two clades dominated by isolates of European origin and suggests two separate transmission networks for HCV-2a in the Netherlands. The large “IDU cluster” (n = 45) exhibits low genetic diversity, reflecting the recent and rapid transmission of this HCV-2a lineage among IDUs in Europe (29). The smaller clade (n = 5) exclusively consists of individuals from Amsterdam who received a blood transfusion in the Netherlands. Interestingly, only 7 of 35 HCV-2a isolates obtained from Amsterdam were part of the two European clusters. None of these seven participants were born in western Europe; the high genetic similarity of their infections with HCV-2a strains from the individuals’ country of birth suggests that all seven had acquired HCV before moving to the Netherlands. Figure 2 contains only 34 of the 35 HCV-2a infected study participants because amplification of NSSB fragment 1 failed in one patient. Additional phylogenetic analysis based on NSSB fragment 2 suggests that this patient was part of the European IDU cluster (data not shown). The phylogenies of HCV-2b and HCV-2c showed similar patterns, but epidemiologically meaningful groupings of sequences from Dutch individuals were not so clear-cut (data not shown). We were able to confirm one heterosexual (HCV-2c) and one homosexual (HCV-2b) transmission event.
In addition to the epidemic subtypes (2a, 2b, and 2c), phylogenetic analysis revealed 15 “nonepidemic” HCV-2 subtypes and lineages, which infected 33% (62/189) of our study population (Table 2). Only 23/62 (37%) of nonepidemic isolates could be classified as recognized HCV-2 subtypes (specifically, 2e, 2f, 2i, 2j, and 2r). The remaining 39 isolates grouped into 10 distinct unclassified HCV-2 lineages, each affecting 1 to 12 individuals. Nonepidemic isolates predominantly represent Amsterdam residents of Surinamese ancestry plus isolates from individuals of Indonesian, African, or Caribbean ancestry (Table 2). Only 5 of the 62 individuals infected with a nonepidemic strain were born in the Netherlands; 2 of these had Surinamese ethnicity and were most likely infected perinatally, 1 had a needle-stick injury (source unknown), 1 acquired HCV sexually (ethnicity of partner unknown), and for 1 no information on transmission risk was available. All blood transfusions associated with HCV-2 infections other than HCV-2a, 2b, or 2c were performed in non-Western countries.

Table 2: HCV-2 genetic diversity, country of birth, and transmission risk factor

<table>
<thead>
<tr>
<th>HCV subtype</th>
<th>No. of patients (n = 189)</th>
<th>% of patients</th>
<th>Transmission risk</th>
<th>Patient origin</th>
<th>Country of birth of majority of patients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>IDU</td>
<td>Transfusion</td>
<td>Endemic</td>
</tr>
<tr>
<td><strong>Epidemic subtypes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>116</td>
<td>61</td>
<td>73</td>
<td>16</td>
<td>4</td>
</tr>
<tr>
<td>2a</td>
<td>35</td>
<td>18</td>
<td>19</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>2b</td>
<td>72</td>
<td>38</td>
<td>53</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>2c</td>
<td>9</td>
<td>5</td>
<td>1</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td><strong>Recognized endemic subtypes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>23</td>
<td>12</td>
<td>0</td>
<td>7</td>
<td>14</td>
</tr>
<tr>
<td>2e</td>
<td>6</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>2f</td>
<td>5</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>2i</td>
<td>9</td>
<td>5</td>
<td>0</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>2j</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2r</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><strong>Unclassified endemic lineages</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>39</td>
<td>21</td>
<td>0</td>
<td>7</td>
<td>26</td>
</tr>
<tr>
<td>2(I)</td>
<td>12</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>2(II)</td>
<td>11</td>
<td>6</td>
<td>0</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>2(III)</td>
<td>5</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>2(IV)</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2(V to X)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>7</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td><strong>No PCR fragment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>11</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

<sup>a</sup>Six distinct unclassified endemic lineages each comprising one or two isolates.

<sup>b</sup>Western origin means all countries in Europe plus the United States, Canada, Australia, and New Zealand.

<sup>c</sup>All transfusions which resulted in chronic infections with endemic HCV-2 strains were performed in non-Western countries.

Molecular clock and phylogeographic analyses. To investigate the longer-term ancestry of HCV genotype 2 isolates from the Netherlands, we reconstructed a relaxed molecular clock tree from an alignment enriched for endemic HCV-2 strains. This phylogeny spans a period of nearly 600 years (Fig. 3). Branches are colored according to each individual’s country of birth, and isolates from Amsterdam are noted by black circles. We identified several major clades associated with endemic persistence in West Africa: (i) the lower 2 clades (green) mainly consist of isolates from Guinea-Bissau and include an early lineage migration to Madagascar (pink); (ii) a clade in the middle (red) that consists of sequences sampled in Cameroon; and (iii) the large upper clade (blue) represents HCV strains from the Ghana/Benin area and contains numerous strongly supported geographic clusters of isolates sampled from various parts of the world, each satisfying our cluster definition criteria. Five of these subclusters are dominated by strains from Surinam-born individuals (clades G, H, I, K, and L), two clades are predominantly associated with people from Indonesia (E and F) and others with individuals from Morocco (C), Vietnam (D), Martinique (N), and Hispaniola (M). The two aforementioned Surinamese clades (H and I) group into a single and older migrant clade (J), statistical support for which is weaker (posterior probability = 0.66).

Figure 3 also provides evidence for secondary migration events after HCV has been exported from Africa. We identified a subtype 2f Indonesian clade (F) that originates from a Surinamese lineage, as well as three Surinamese isolates that group inside a larger subtype 2e Indonesian clade (E). Further, a pair of subtype 2i Vietnamese isolates (orange) cluster amid a clade of predominantly Moroccan strains (C). All these secondary migration events are well supported, although the direction of the second Indonesian/Surinamese movement is not certain. There are three possibilities: (i) movement from Surinam to Indonesia, (ii) movement from Indonesia to Surinam, and (iii) movement from Indonesia to Surinam and then back to Indonesia.

Dating of lineage migration events. From the molecular clock
FIG 1: Maximum likelihood phylogeny including the 167 sequences originating from our study spanning both fragments 1 and 2. Taxon labels are colored by individual's risk factor: red, IDU; blue, blood transfusion/nosocomial; green, sexual/MSM; brown, birth in a non-Western country and also include the individual’s country of birth using International Organization for Standardisation (ISO) codes. Blue dots indicate well-supported clades (bootstrap value ≥ 70).
FIG 2 Maximum likelihood phylogeny of subtype 2a sequences from our study and all available homologous reference sequences in the HCV sequence database. Dots next to taxon names indicate sequences originating from our study. Taxon labels are color coded by risk factor: red, IDU; blue, blood transfusion. Tip labels denote the country of birth for sequences originating from our study and the country of isolation for sequences obtained from the HCV sequence database using International Organization for Standardization (ISO) codes.
FIG 3 Molecular clock phylogeny of endemic HCV genotype 2 strains. Black dots next to phylogeny tips indicate sequences resulting from our study. Tree branches are colored according to inferred lineage locations: green, West Africa (Senegal, Gambia, Guinea Bissau, and Guinea); blue, Benin-Ghana area (Benin, Burkina Faso, and Ghana); red, Cameroon, Central African Republic; pink, Madagascar; brown, Caribbean; orange, Asia; light blue, North Africa; black, Western countries. Orange dots on nodes indicate well-supported clades (posterior probability ≥ 80); red dots on nodes indicate clades with excellent support (posterior probability ≥ 90). The capital letters with arrows label the root nodes of the major migrant clades.
Molecular clock-based intervals for lineage migration dates and their credible intervals. The lower box of Fig. 4a shows only those HCV lineages that moved across the Atlantic Ocean; the upper box shows HCV-2 lineages that moved to other areas. Gray circles in the middle of the intervals indicate our best estimate of time of migration. In some estimates a vertical bar is shown, representing an ancestral node with weak support. In these cases, the left end of the interval represents a conservative estimate of the upper date of migration (see Materials and Methods). Whiskers represent 95% highest posterior density (HPD) credible regions of the most conservative estimates on both sides. Color coding as in Fig. 3. Capital letters in brackets next to arrival destinations correspond to the clade labels in Fig. 3. CAR, Central African Republic. (b) Number of individuals transported as part of the trans-Atlantic slave trade over time: black line, numbers of individuals disembarked in Dutch Guianas; gray line, total numbers of individuals disembarked in the New World (in quarter-century periods). Brown marks indicate the best estimates for times of movement of the trans-Atlantic infection lineages.

FIG 4 Estimated time periods of HCV genotype 2 global lineage migrations and the numbers of individuals transported as part of the trans-Atlantic slave trade over time. (a) Molecular clock-based intervals for lineage migration dates and their credible intervals: lower box, trans-Atlantic migration events; upper box, migration events to other areas. Gray circles in the middle of the intervals indicate our best estimate of time of migration. In some estimates a vertical bar is shown, representing an ancestral node with weak support. In these cases, the left end of the interval represents a conservative estimate of the upper date of migration (see Materials and Methods). Whiskers represent 95% highest posterior density (HPD) credible regions of the most conservative estimates on both sides. Color coding as in Fig. 3. Capital letters in brackets next to arrival destinations correspond to the clade labels in Fig. 3. CAR, Central African Republic. (b) Number of individuals transported as part of the trans-Atlantic slave trade over time: black line, numbers of individuals disembarked in Dutch Guianas; gray line, total numbers of individuals disembarked in the New World (in quarter-century periods). Brown marks indicate the best estimates for times of movement of the trans-Atlantic infection lineages.

tree, we obtained estimated dates of phylogenetic nodes that define conservative lower and upper limits for the timing of lineage migration events. The lower box of Fig. 4a shows only those HCV lineages that moved across the Atlantic Ocean; the upper box shows HCV-2 lineages that moved to other destinations. These estimated migration dates are displayed above a graph that shows the numbers of individuals transported from West Africa during the transatlantic slave trade (Fig. 4b). Both the total number of transported slaves (gray line) and the number transported to Dutch Guianas (a historical colonial territory that includes present-day Surinam; black line) are shown. All transatlantic movements are estimated to have occurred within the period of trans-Atlantic slave trade (1650 to 1900). The majority of the dating intervals end before 1850. The substantial uncertainty of the upper bound for the date of movement of clade M to Hispaniola is due to (i) the long internal branch that precedes this clade and (ii) the low statistical support for the pre-migration node ancestral to clade M (Fig. 3).

Migration of HCV-2 lineages from West Africa to Central Africa (A) and Madagascar (B) is dated to the 17th century. HCV lineage migration to Morocco is dated to between the early 17th and early 19th centuries, reflecting a considerable uncertainty in the older time limit (for the same reasons as those applying to clade M). Movement of HCV-2 to Vietnam most likely occurred between the second halves of the 18th and 19th centuries, with a lower 95% HPD limit that stretches into the late 17th century. The two lineage migrations to Indonesia happened at distinctly different times. The earlier migration most likely occurred sometime between the mid-18th and the late 19th centuries, while the more recent migration, which represents the movement of HCV-2f to Indonesia from Surinam, is dated to the first few decades of the 20th century.

Testing for association between geography and phylogeny. Table 3 shows the results of the statistical test for significant phylogeographic structure in the phylogeny. The association index (AI) is significant for all geographical locations, except for Martinique, while the parsimony score (PS) is statistically significant for all locations. This is a strong evidence for geographical grouping in the phylogeny, and the conclusion is robust in terms of phylogenetic estimation uncertainty.

DISCUSSION

Our analysis of HCV-2 from the Amsterdam area uncovered substantial viral genetic diversity, including 8 known HCV-2 subtypes and 10 previously unclassified subtype-like lineages, and revealed a complex phylogenetic pattern associated with geography and risk groups. As discussed below, these associations are most likely driven by a combination of social, demographic, and economic factors over both recent and historical timescales. The phylogenetic distribution of the previously unclassified lineages suggests that they represent separate subtypes; however, our data did not satisfy the complete criteria for assigning new subtypes (36). More sampling and (complete) genome sequences will be required before these lineages can be definitively classified. To obtain a broad sample of HCV infections, we included patients from an academic hospital, an inner-city hospital, and the public health service, of which the latter specifically deals with groups at risk for sexually transmitted and blood-borne infections. Although this study includes all patients diagnosed with HCV-2 at these three large centers in Amsterdam during a 10-year time period, selection bias cannot entirely be excluded.

The majority of Amsterdam isolates belonged to the epidemic subtypes 2a, 2b, and 2c commonly found in western Europe (9, 11, 43). During the 20th century, the growth of parenteral routes of transmission resulted in the worldwide exponential growth of HCV, including HCV subtypes 2a, 2b, and 2c (30). Although the phylogenies of these subtypes revealed similar epidemiological profiles, the existence of "epidemic" transmission networks for HCV-2 is best illustrated by the HCV-2a phylogeny (Fig. 2). The
reflects the global nature of human movement during the 500
While incomplete sampling may contribute to this pattern, it also
individual location significantly group together on the phylogeny.
structure. Statistical tests demonstrated that isolates from each
spread geographic distribution, with a pronounced geographical
lonial rule there. Nevertheless, HCV-2 reveals a remarkably wide-
Two countries is unsurprising given the long history of Dutch co-
Indonesia. The high frequency of endemic strains from the last
In contrast to the French studies, the vast majority of Amsterdam
HCV-2 infections in Canada to 67% in southeast France [24,40].
found at varying frequencies in other Western countries (10% of
one-third of our isolates. Endemic HCV-2 lineages have been
explained HCV-2 lineages to Surinam. Strikingly, multiple separate
migrations of HCV occurred from Africa to various countries in
the Caribbean. Even a relatively small territory such as Surinam
previously proposed hypothesis (20) that HCV was introduced to the Americas
likely low prevalence of endemic HCV in Africa and its low infec-
tions fell within the 150 years between 1700 and 1850. This coin-
introduction. Most of the estimated time windows for viral migra-
tions occurred from Africa to various countries in the past from West Africa (where HCV-2 originated [20]) to
other areas. Among others, these lineages include HCV-2e and 2f to Indonesia; HCV-2i to Morocco, France, Vietnam, and Quebec (24, 26, 40); HCV-2j to Venezuela (39); HCV-2k to Martinique and France (21); HCV-2m to Vietnam (26); HCV-2r to Haiti and the Dominican Republic (24); and numerous nonclassified di-
verse HCV-2 lineages to Surinam. Strikingly, multiple separate migration
occurred from Africa to various countries in the Caribbean. Even a relatively small territory such as Surinam
received multiple independent viral introductions, each drawn from the broader pool of HCV-2 diversity in West Africa. The likely low prevalence of endemic HCV in Africa and its low infect-
ivity prior to the age of modern medicine suggest that such pat-
ttern was likely generated by a process that involved the movement
of substantial numbers of Africans.
Crucially, in this study, we used a combined phylogeographic
and molecular clock approach to directly test the previously pro-
posed hypothesis (20) that HCV was introduced to the Americas
by the trans-Atlantic slave trade. If this hypothesis is correct, then
the timing of viral introductions should coincide with the period
of slave transportation as estimated from historical records. The
transatlantic slave trade started gradually at the beginning of 16th
century and ended in the latter half of the 19th century, during
which it was responsible for the forced movement of an estimated
10.7 million people from Africa to the Americas [47]. We used
evolutionary analysis to estimate the date of movement of each
migrant lineage with appropriate confidence limits: the midpoint
of these limits represents our best estimate of the date of viral
introduction. Most of the estimated time windows for viral migra-
tions fell within the 150 years between 1700 and 1850. This coin-

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Geographic cluster</th>
<th>Observed mean (95% CI)</th>
<th>Null mean (95% CI)</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Association index</td>
<td>Surinam</td>
<td>0.98 (0.53–1.46)</td>
<td>6.54 (5.33–7.8)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Hispaniola</td>
<td>0.004 (0.002–0.01)</td>
<td>1.25 (0.79–1.65)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Martinique</td>
<td>1.39 (1.04–1.7)</td>
<td>1.81 (1.3–2.21)</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>Indonesia</td>
<td>0.59 (0.35–0.97)</td>
<td>3.30 (2.51–3.94)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Vietnam</td>
<td>0.3 (0.13–0.47)</td>
<td>1.49 (1.05–1.91)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Morocco</td>
<td>0.85 (0.52–1.16)</td>
<td>2.42 (1.84–2.9)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Benin-Ghana</td>
<td>1.82 (1.26–2.39)</td>
<td>5.95 (4.94–6.79)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Guinea-Gambia</td>
<td>1.04 (0.6–1.53)</td>
<td>4.83 (3.97–5.71)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Cameroon/CAR</td>
<td>0.54 (0.37–0.74)</td>
<td>2.76 (2–3.43)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Parsimony score</td>
<td>Surinam</td>
<td>11.37 (10–13)</td>
<td>34.67 (32.36–36.58)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Hispaniola</td>
<td>1 (1–1)</td>
<td>5.95 (5.52–6.0)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Martinique</td>
<td>6.76 (6–7)</td>
<td>8.91 (8.14–9)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Indonesia</td>
<td>4.96 (4–6)</td>
<td>16.49 (15.5–17)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Vietnam</td>
<td>3 (3–3)</td>
<td>6.95 (6.66–7)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Morocco</td>
<td>7.06 (6–8)</td>
<td>11.83 (11.01–12)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Benin-Ghana</td>
<td>20.69 (18–23)</td>
<td>30.51 (29.03–31.81)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Guinea-Gambia</td>
<td>11.32 (10–12)</td>
<td>24.18 (22.94–24.97)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Cameroon/CAR</td>
<td>7 (7–7)</td>
<td>13.65 (12.86–14)</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

*a CI, confidence interval; CAR, Central African Republic.

TABLE 3 Phylogeographic structure: association indices and parsimony scores

Continental Africa could have resulted from land travel or the
gradual diffusion of viruses through local interaction, the only
plausible route by which HCV lineages could have moved among
continents before the 20th century is via the emergence and rise of
global marine navigation for exploration and trade.

Our results demonstrate the existence of multiple HCV-2 “mi-
grant clusters,” which represent lineages that moved at some point
in the past from West Africa (where HCV-2 originated [20]) to
other areas. Among others, these lineages include HCV-2e and 2f
to Indonesia; HCV-2i to Morocco, France, Vietnam, and Quebec
(24, 26, 40); HCV-2j to Venezuela (39); HCV-2k to Martinique and
France (21); HCV-2m to Vietnam (26); HCV-2r to Haiti and the
Dominican Republic (24); and numerous nonclassified di-
verse HCV-2 lineages to Surinam. Strikingly, multiple separate
migration occurred from Africa to various countries in the
Caribbean. Even a relatively small territory such as Surinam
received multiple independent viral introductions, each drawn from the
broader pool of HCV-2 diversity in West Africa. The likely
low prevalence of endemic HCV in Africa and its low infect-
ivity prior to the age of modern medicine suggest that such pat-
ttern was likely generated by a process that involved the movement
of substantial numbers of Africans.

In addition to the epidemic subtypes, we observed 5 known endemic HCV subtypes and 10 distinct unclassified HCV-2 line-
eages each affecting 1 to 12 individuals, together accounting for
one-third of our isolates. Endemic HCV-2 lineages have been
found at varying frequencies in other Western countries (10% of
HCV-2 infections in Canada to 67% in southeast France [24,40]).
In contrast to the French studies, the vast majority of Amsterdam
demic isolates were from individuals born outside the Nether-
lands, in Africa, Asia, or Latin America, particularly Surinam and
Indonesia. The high frequency of endemic strains from the last
two countries is unsurprising given the long history of Dutch co-

Colonial History and HCV-2 Diversity in Amsterdam
cides with the historical period during which slave shipments reached their peak. Although trans-Atlantic slave trade continued for nearly 4 centuries, more than 80% (approximately 8.8 million people) of Africans ever transported were moved across the Atlantic during these 150 years (47).

Phylogeographic analysis provides a second line of evidence supporting the trans-Atlantic slave trade hypothesis. All the HCV-2 trans-Atlantic “migrant clusters” (including those from Surinam) originated from the Benin-Ghana region of Africa (blue in Fig. 3) while none originate from Senegambia (green in Fig. 3). Present-day Ghana and Benin featured 30 European trading posts along their coast from the 15th to 19th centuries, compared to only 5 posts among the four countries of Senegambia (48). Approximately 2.8 million slaves were moved to the New World from the Gold Coast and Bight of Benin (present-day Ghana, Benin, and Burkina Faso) between the 16th and 19th centuries, whereas ~600,000 slaves were transported from Senegambia during the same period. The former’s relatively short span of coast was the embarkation point of more than a quarter of all people transported (47).

The trans-Atlantic slave trade has been proposed as the driving force of intercontinental expansion of other viruses. A recent phylogenetic study suggests that West African hepatitis B virus (HBV) lineages were introduced to Haiti during the peak years of slave movement (1). Similarly, recent studies of human T-cell leukemia virus type 1 (HTLV-1), which sometimes combine human and viral evidence, have indicated that Surinamese, Guyanan, Brazilian, Argentinian, and Peruvian HTLV-1 lineages grouped among West African lineages (3, 12, 32).

Although the trans-Atlantic slave trade provides a plausible explanation for the movement of HCV-2 to the Americas, it leaves the long-term presence of endemic HCV-2 in Asia unexplained. European nations likely played a role in the global dissemination of HCV, additionally through trade and other forms of human migration among their colonies. An example is subtype 2e, which links an endemic HCV-2 lineage from the Benin-Ghana area to Indonesia. Dutch companies established multiple trading ports in this part of Africa and traded intensively with the largest Dutch colonial territory—Indonesia. Records also suggest the significant movement of African slaves from West Africa to the Dutch Cape colony and Asia by Dutch traders (27). Our results also highlight secondary lineage movements among colonial territories sharing the same metropole. Subtypes 2e and 2f suggest viral movement in both directions between the former Dutch colonies of Surinam and Indonesia. Although the timing and direction of subtype 2e migration is uncertain, we provide evidence that HCV-2f moved from Surinam to Indonesia during the early 20th century. These more recent movements are consistent with the bidirectional migration of initially 2,600 Chinese migrant workers mainly from Java (Indonesia) between 1853 and 1874, and later 33,000 Javanese contract laborers from Indonesia to Surinam during 1890 to 1940, a direct consequence of the abolition of slavery in the latter by the Dutch government in 1862 (19, 22). About a quarter of these laborers returned to Indonesia at the end of their 5-year contract in the period between 1897 and 1956. People of Javanese descent still comprise 15 to 20% of the present-day Surinamese population (17).

By combining demographic, historical, and viral genetic data, we have reconstructed the transmission history of HCV-2, which explains its worldwide distribution as well as the present-day pattern of diversity observed in Amsterdam. It appears that the colonial activities of European countries played a decisive role in the dissemination of HCV genotype 2, predominantly to the Americas, but also to former colonial territories in Asia. Interestingly, HCV subtype 2f (cluster C in Fig. 3) falls within the Benin-Ghana region and solely contains isolates from francophone regions, including Morocco, France, Quebec, and Vietnam (former French Indochina). Although indicative of a role for the colonial activities of France in the dispersal of HCV-2f, our sample of “Dutch origin” is not sufficiently representative to address this hypothesis. The Benin-Ghana region of Africa, both a major embarkation area for vessels carrying slaves to the Americas and an important trading post along European-Asian trading routes, emerges as an important historical hub of global HCV-2 distribution. Since the global epidemic HCV subtypes 1a and 1b were estimated to originated from the United States (18), we hypothesize here that trans-Atlantic slave trade was also responsible for the movement to the Americas of these genotype 1 endemic African lineages. Molecular clock dating of transatlantic movements of HCV from West Africa strongly supports our historical slave trade hypothesis. In addition to information on the global nature of human movement before 1900, our phylogenetic analysis of epidemic subtypes 2a, 2b, and 2c, together 67% of HCV-2 infections in our study, exposed both national and international transmission networks responsible for the emergence of HCV-2 in most Western countries during the 20th century, specifically injecting drug use and contaminated blood/blood products. Undiagnosed chronically HCV-infected individuals might benefit from an increased understanding of the epidemiological processes that generated the global pattern of HCV diversity seen today. Better understanding of the associations between risk factors, populations, and HCV subtypes might help target HCV prevention and screening campaigns and decrease the future burden of HCV-related liver disease.

REFERENCES