Spatiotemporal reconstruction of the introduction of hepatitis C virus into Scotland and its subsequent regional transmission

Anna L. McNaughton[a], Iain Dugald Cameron[a], Elizabeth B. Wignall-Fleming[a], Roman Biek[a,b], John McLauchlan[a], Rory N. Gunson[c], Kate Templeton[d], Harriet Mei-Lin Tan[a], E. Carol McWilliam Leitch[a]#

MRC-University of Glasgow Centre for Virus Research, Glasgow, UK[a]; IBAHCM, University of Glasgow, Glasgow, UK[b]; West of Scotland Specialist Virology Centre, Glasgow Royal Infirmary, Glasgow, UK[c]; Edinburgh Specialist Virology Centre, Edinburgh, UK[d]

Running Head: HCV transmission in Scotland

#Address correspondence to E. Carol McWilliam Leitch, carol.leitch@glasgow.ac.uk

ALM, IDC and EBW-F contributed equally to this work.

Abstract word count: 250

Text word count: 4716

Abstract

A more comprehensive understanding of hepatitis C virus (HCV) transmission dynamics could facilitate public health initiatives to reduce the prevalence of HCV in people who inject drugs. We aimed to determine how HCV sequences entered and spread throughout Scotland and to identify transmission hotspots. A Scottish dataset with embedded demographic data was created by sequencing the NS5B of 125 genotype (Gt) 1a and 166 Gt3a samples and analysed alongside sequences from public databases. Applying Bayesian inference methods, we reconstructed the global origin and local spatiotemporal dissemination of HCV in Scotland. Scottish sequences mainly formed discrete clusters interspersed between sequences from the rest of the world; the most recent common ancestors of these clusters dated to 1942-1952 (Gt1a) and 1926-1942 (Gt3a), coincident with global diversification and distribution. Extant Scottish sequences originated in Edinburgh (Gt1a) and Glasgow (Gt3a) in the 1970s but both genotypes spread from Glasgow to other regions. The dominant Gt1a strain differed between Edinburgh (cluster [C]2), Glasgow (C3) and Aberdeen (C4) whereas significant Gt3a strain-specificity only occurred in Aberdeen. Specific clusters initially formed separate transmission zones in Glasgow that subsequently overlapped occasioning city-wide co-circulation. Transmission hotspots were detected with 45% of samples from patients residing in just nine of Glasgow's 57 postcode districts. HCV was introduced into Scotland in the 1940s concomitant with its worldwide dispersal likely arising from global-scale historical events. Cluster-specific transmission hubs were identified in Glasgow, the key Scottish city implicated in HCV dissemination. This fine-scale spatiotemporal reconstruction improves understanding of HCV transmission dynamics in Scotland.

Importance

HCV is a major health burden and the leading cause of hepatocellular carcinoma. Public health needle exchange and "treatment as prevention" strategies targeting HCV are designed to reduce prevalence of the virus in people who inject drugs (PWID), potentially mitigating the future burden of HCV-associated liver disease. Understanding HCV transmission dynamics could increase the effectiveness of such public health initiatives by identifying and targeting regions playing a central role in virus dispersal. In this study we examined HCV transmission in Scotland by analysing the genetic relatedness of strains from PWID alongside data inferring the year individuals became infected and residential information at a geographically finer-scale resolution than in previous studies. Clusters of Scotland-specific strains were identified with regional specificity and mapping the spread of HCV allowed the identification of key areas central to HCV transmission in Scotland. This research provides a basis for identifying HCV transmission hotspots.

Hepatitis C virus (HCV) currently infects an estimated 180 million people throughout the world. Following a short acute phase, the virus enters an asymptomatic, chronic stage which can persist for decades before the potential onset of severe sequelae such as liver cirrhosis and hepatocellular carcinoma. HCV-associated liver disease is expected to grow exponentially in the UK over the next decade and will become a major health care burden (1). In developed countries, HCV transmission occurs predominantly through injecting drug use (IDU). Currently there is no vaccine against HCV and the new, highly effective direct acting antivirals (DAAs) are extremely costly. Prevention of new infections is therefore considered the most cost-effective policy for reducing overall HCV prevalence in PWID (2).

HCV is a significant public health risk in Scotland where the HCV incidence is twice that of the rest of the UK (3). Of the seven recognised genotypes of HCV, the most prevalent in Scotland are Gt1 (49%) and Gt3 (46%) (3). Programmes aimed at reducing transmission of the virus, such as needle exchange initiatives, have been implemented in Scotland since the early 1990s and resulted in decreases in HCV prevalence levels of approximately 16% over six years; these measures however have proven insufficient to completely control the epidemic (4). "Treatment as prevention" schemes have been successful in reducing transmission of HIV (5) and a similar strategy treating HCV-infected PWID with the DAA Telaprevir (Vertex Pharmaceuticals, Switzerland), interferon and ribavirin has recently commenced in Scotland (J. Dillon, personal communication). Mathematical models predict that DAA treatment of as little as 2% of HCV-positive PWID in Edinburgh could reduce the incidence of infection by 26% in 15 years (2). Better tools are required to monitor the effect of intervention strategies on patterns of local HCV incidence and transmission, including the potential spread of antiviral resistance.

Tremendous advances have been made in recent years inferring viral transmission dynamics from sequence data, in part through the development of powerful statistical methods (6). As an extension of this methodology, phylogeographical approaches combine viral genetic data with information on the time and place of infection in order to reconstruct viral spread and to quantify viral transmission (7, 8). Transmission hotspots, areas of high incidence of an infective agent thought to drive its spread, have been exploited to devise novel prevention and control approaches for various infectious diseases such as cholera and malaria (9). While in-depth studies of HCV transmission dynamics combining sequence data and residential information are currently limited, some have successfully used social network data as a substitute for geographical information. A recent study in Australia (10) was the first to identify a positive association between HCV strain genetic relatedness and reported injecting relationship. An earlier study in Brazil (11) showed that HCV transmission dynamics in Sao Paulo differed according to genotype and that social factors play an important role in the spread of the virus. These studies demonstrate the great potential of phylogeographic inference to quantify HCV spread from viral sequence data.

In this study we applied Bayesian inference methods to embedded phylogenetic and demographic data with a threefold purpose: (i) to reconstruct the introduction of the epidemic HCV genotypes 1a and 3a into Scotland, (ii) to infer the spatiotemporal dispersal of sequences within Scotland and (iii) to identify hotspots of transmission. Scottish sequences tended to form discrete clades within global trees with most recent common ancestors (MRCAs) suggestive of multiple concurrent introductions of HCV in the 1940s. Although a degree of regional strain-specificity has been maintained to the present time, Glasgow was the major source of strains

disseminated to neighbouring regions. Transmission zones in Glasgow expanded markedly within a single decade resulting in eventual co-circulation of genotypes and strains. A small number of districts within Glasgow were identified as key centres of transmission.

MATERIALS AND METHODS

Study design. We first created a Scottish dataset by sequencing a partial NS5B region of 125 Gt1a and 163 Gt3a anonymised samples from throughout Scotland collected between 2011 and 2014. Subsequently we assembled a dataset comprising sequences from the rest of the world (RoW) by retrieving and collating all sequences from the NCBI and Los Alamos HCV databases with information on country of origin and sample year, and covering the same genomic region as the Scottish dataset. This resulted in a RoW dataset of 381 Gt1a and 47 Gt3a sequences, three of which were from Edinburgh and were included in the Scottish dataset. The sampling locations of the RoW dataset were United States of America (N=320; US), Switzerland (N=45; CH), Germany (N=14; DE), Brazil (N=1; BR) and China (N =1; CN) for Gt1a and China (N=16), India (N=9; IN), Pakistan (N=1; PK), Japan (N=1; JP) and United Kingdom, not Scotland (N=17; UK other) for Gt3a. We reconstructed the origin and spatial distribution of Scottish sequences within a local and a global context by analysing the embedded phylogenetic, geographical and epidemiological data contained in these datasets.

Creation of the Scottish Dataset. Samples from HCV-infected individuals were obtained from two diagnostic labs, the West of Scotland Specialist Virology Centre (WoS-SVC) and the Edinburgh Specialist Virology Centre (ESVC). Ethical approval was obtained from the NHS

119 Greater Glasgow and Clyde Biorepository (application 140) and South East Scotland SAHSC
120 Human Annotated BioResource (reference No.10/S1402/33) respectively. To maintain
121 confidentiality, identifiers and clinical data other than partial postcode (PC) data, year of birth,
122 year commenced injecting and likely route of infection were delinked. Samples diagnosed at the
123 WoS-SVC were from patients attending clinics throughout Scotland between July 2013 and
124 March 2014. Extracted RNA was reverse transcribed and amplified using genotype-specific
125 primers (Table 1) in a nested PCR covering a 695bp (Gt1a) or 679bp (Gt3a) region of the NS5B.
126 The ESCV samples were collected between April and December 2011, and were treated
127 similarly but amplified using a combined reverse transcription (RT)-PCR procedure. Amplified
128 products were visualised by electrophoresis to confirm product presence and size, then Sanger-
129 sequenced. All samples were screened for the presence of Gt1a/Gt3a co-infections utilising
130 genotype-specific primers and samples from individuals with co-infections (4.1% of samples)
131 were excluded from the cohort.
132 An estimation of the average age drug use commenced (21 years) was calculated from
133 the detailed data on date of birth and year IDU commenced available for 34 Edinburgh subjects.
134 There was a wide range of ages for the commencement of IDU (14-39 years of age) although
135 the mean, median and mode (21, 19 and 19 years respectively) ages of these individuals
136 confirmed that older age groups were unusual. This was applied to all other samples in the
137 Scottish dataset to give the estimated year of infection (EYI) for these individuals and used in
138 the phylogeographical analysis within Scotland. As screening of blood samples commenced in
139 1989 in the UK it is highly likely that individuals who became infected with HCV after this time
140 acquired the infection through IDU. To enrich for PWID, only samples from subjects born in

1969 or later or those born before 1969 but who reported drug use as the main risk factor were included in the analyses on transmission within Scotland.

Study subject demographics. The mean age of individuals was similar for the Gt1a group (38.6 years [SD, 9.0]) and the Gt3a (41.0 years [SD, 9.8]) groups. Information on residential PC district was captured for all study participants and the finer spatial resolution contained in PC sector data was available for 85% of samples.

Temporal phylogenetic analyses. Scottish sequences were aligned and edited where required using SSE version 1.1 (12) and added to the imported RoW sequences to form the global dataset. Bayesian MCMC inference was implemented in BEAST v1.8.0 (13) and the output inspected by the software Tracer v1.6 (14). The HKY nucleotide substitution model (15) with gamma rate heterogeneity was used throughout with a relaxed uncorrelated lognormal molecular clock. A Bayesian skyline coalescent model (16) was used as a flexible demographic prior in all analyses with chain lengths of 200 million. The global, RoW and Scottish datasets were analysed for each genotype.

Phylogeographic analyses. Geographical regions in the UK are divided into postcodes, analogous to zip codes, and consist of six or seven alphanumeric characters (e.g. G61 1QH or EH16 3JG) normally representing a single street. The first letter/s denote the PC region (e.g. G or EH), the first grouping is the PC district (e.g. G61) and an additional digit is used for PC sectors (e.g. G61 1). There are 16 PC regions, 476 PC districts and 1274 PC sectors in Scotland. Regional PCs included in this study were Glasgow, Paisley (PA), Kilmarnock (KA), Motherwell (ML), Edinburgh (EH), and Aberdeen (AB) for both genotypes, and Dundee (DD) additionally for Gt3a. Other regions were not included in the analyses due to small sample numbers ($N \leq 5$).

Maps with PC regions (http://free-postcode-maps.co.uk/) or without (https://maps.google.co.uk/maps/) were downloaded and utilised with the Scottish dataset. Samples indicating the cluster group and EYI were located on maps according to PC information; PC districts were used on maps of Scotland and, for finer resolution, PC sectors were used for Glasgow maps.

Nucleotide sequence accession numbers. All newly generated sequences from this study were submitted to GenBank and were assigned the accession numbers KR071882 to KR072203. The HCV sequences that were included in the RoW dataset were downloaded from GenBank and are shown in the supplemental table S1. The three GenBank sequences used in the Scottish dataset were AF516368-AF516370.

RESULTS

MCMC analysis. The MRCAs of the global datasets (Table 2) was 1888 (95% highest posterior density intervals [HPD], 1857-1914) for Gt1a and 1899 for Gt3a (95% HPD, 1857-1914). Lineages expanded exponentially, with 90% of Gt1a lineages emerging between 1940 and 1965 (Fig. S1); Gt3a lineages emerged over a longer time period (1935-1975) in two distinct periods peaking in 1940 and 1960 (Fig. S1). The dates of the Scottish and RoW datasets analysed separately for Gt1a (1918 and 1907 respectively) and Gt3a (1901 and 1897 respectively) paralleled the values of the global dataset. The evolutionary rate of Gt1a was 1.88 $x10^{-3}$ substitutions/site/year (s/s/y) (95% HPD, 1.53-2.29 $x10^{-3}$ s/s/y), similar to the evolutionary rate of Gt3a (1.65 $x10^{-3}$ s/s/y; 95% HPD, 1.19 $x10^{-3}$-2.14 $x10^{-3}$).

Scottish clusters. For the purposes of this study Scottish clusters were defined as clades containing $\geq$ 5 Scottish sequences of which > 70% of sequences were from Scotland. The Gt1a global dataset contained four Scottish clusters (Table 3), designated C1 to C4, interspersed

between groups of RoW sequences (Fig. 1A). The most abundant Scottish Gt1a clusters were C2 (N=34, 89% Scottish sequences) and C3 (N=43, 100% Scottish sequences). The MRCAs of the Gt1a Scottish clusters ranged from 1942 to 1952 (Table 3). It was difficult to interpret the positioning of the Scottish Gt3a sequences in the context of the global setting as there were many more Scottish sequences (N=200) than available RoW sequences (N=44, Fig. 1B). We identified seven Scottish clusters (C1 to C7) containing between 7 and 76 sequences; 72-97% of sequences in each of these clusters originated from Scotland (Table 3). The Gt3a Scottish clusters had MRCAs ranging between 1926 and 1942.

Phylogeographical analyses. The frequency of Scottish clusters in each of the major PC regions was determined. As there was little difference in the types and frequencies of clusters isolated in the PC regions Glasgow, PA, KA and ML, this data was additionally analysed together as west central Scotland (WCS, Fig. 2). The types and frequencies of Gt1a sequence clusters differed between AB, EH and WCS (Fig. 2A). The predominant clade in AB was C4 (35%), in EH it was C2 (65%) and in WCS it was C3 (51%). 35% of AB sequences belonged to non-Scottish clades compared to 15% in EH and 12% in WCS. Although none of the EH sequences occurred in clades unique to that area, the proportion of clades differed considerably from the other regions. The predominant EH clade C2 (65%) was less frequent in WCS (23%) and AB (12%), and the minor EH clade C3 (5%) occurred more frequently in WCS (51%) and AB (12%). Geographic-specific clustering patterns were observed for Gt3a only as regards AB sequences (Fig. 2B). The predominant AB clade C1 (46%) was represented less in EH (5%) and WCS (7%). Conversely clade C6 was not observed in AB but constituted 9%, 18% and 22% in EH, DD and WCS respectively.

Gt1a transmission in Scotland. The Scottish Gt1a sequences were derived from individuals residing in 71 PC districts and 85 PC sectors. Sequences indicating strain cluster were represented on maps according to the residential PC district and EYI of the participant (Fig. S2). The earliest Gt1a sequences in this study (1970-1981) occurred predominantly in EH and mainly comprised clade C2 (75%). Glasgow became the focus of Gt1a transmission in the following 5 years, particularly clade C3, and sequences subsequently spread to the neighbouring regions PA, KA and ML. Gt1a sequences were not apparent in AB until 1994 with the AB-specific clade C4. Overall the initial dominant clades in WCS (C3), EH (C2) and AB (C4) persisted and remained the dominant clades throughout the 42 year period of the study.

Transmission of Gt1a sequences in Glasgow. The dominant Gt1a clades in Glasgow (C2 and C3) were mapped by residential PC sector to obtain a finer resolution of dispersal within the city (Fig. 3). Initially C3 sequences were mainly dispersed in the west of the city but transmission switched to the eastern suburbs in 1988 and after 2000 spread to northern and south-western districts. Clade C2 was detected later than C3 in Glasgow, with an initial transmission zone in the south-east and northern suburbs before expanding to inner western suburbs.

Gt3a transmission in Scotland. Sequences from Gt3a Scottish clusters belonged to 98 PC districts and 126 PC sectors. Samples were classified by strain cluster and represented on maps by residential PC district and EYI (Fig S3). Before 1980 Gt3a sequences were mainly from individuals residing in the Glasgow PC area and this genotype did not spread to EH until the following decade. Transmission did not expand to the other regions of Scotland until the 1990s, later than Gt1a.

Transmission of Gt3a sequences in Glasgow. In Glasgow transmission of Gt3a clade C5 occurred in northern PC sectors before 1983 (Fig. 3). In the following decade the transmission zone expanded to include more northern suburbs and areas in the east and south of the city. After 1998 C5 was mainly transmitted in central districts. Initial dispersal of clade C6 occurred in the inner northern suburbs of Glasgow and spread to the north-east and south-west of the city. Later transmission occurred mainly in semi-rural north-western regions.

Co-circulation of Gt1a and Gt3a clusters in Glasgow. The temporal dispersal of the two major Gt1a (C2 and C3) and Gt3a (C5 and C6) clusters within Glasgow was reconstructed (Fig. 4). Between 1974 and 1983, individual strains formed largely separate transmission networks within the city. These zones expanded in the next 5 years and began to overlap. Co-circulation of all 4 clades was apparent by 1994 and subsequent transmission occurred mainly within these regions, suggestive of self-sustaining networks.

Hotspots of HCV infection in Glasgow. There were a sufficient number of Gt1a (N=54) and Gt3a (N=94) samples from the Glasgow PC region to examine in more detail. Gt1a sequences were associated with 25/57 Glasgow PC districts and seven of these districts contained >50% of the sequences (Fig. 5). At a finer resolution, the predominant PC sector locations of Gt1a samples were G64-9 (7.4%) and G33-2 (5.9%) and the EYI in these sectors was 1997-2008 and 1988-2006 respectively. 51% of cluster C2 sequences were derived from PC districts G32, G42 and G73 whereas C3 was associated with PC districts G33 and G64 (12% each) (data not shown). The Gt3a Glasgow samples were from individuals residing in 35 of the 57 PC districts. Six of these districts each contained >5% of all sequences and accounted for 40% of the total sequences

(Fig. 5). PC sectors G33-9 (7.4%) and G21-2 (5.1%) contained the most samples and the EYI in these sectors were 1995-2004 and 1978-1994 respectively. Cluster C5 was associated with PC districts G21 and G51 (11% each) whereas C6 was associated with the G64 district (19%) (data not shown). Overall, 45% of samples were from patients residing in nine of Glasgow's 57 PC districts and 11% of samples were from just four of Glasgow's 241 PC sectors, representing a 7-fold increase from the expected.

DISCUSSION

Bayesian MCMC methods were used to reconstruct the introduction of HCV Gt1a and Gt3a extant sequences into Scotland from the global pool using the BSP coalescent model (16) which allows for variable population sizes over time. A Gt1a global dataset containing 506 sequences, the largest study of its type to date, was used to calculate the evolutionary rate of this genotype (1.88×10^{-3} s/s/y) which was faster than previously determined. Calculations of evolutionary rates can be adversely affected by analysing small numbers of sequences and slower evolutionary rates (1.48×10^{-3}, 1.0×10^{-3} and 9.05×10^{-4} s/s/y) have been calculated in studies with smaller datasets [N=334 (17), N=111 (18) and N=176 (19) respectively]. Our results suggest that Gt1a coalesced in 1888 (1857-1914) which confirms a previous report [1900 (1802-1957)] (18) but pre-dates another comparable study [1920, (19)].

Gt1a lineage expansion principally occurred between 1940 and 1965 and subsequent sequence clustering displayed geographical-specificity. This time period coincides with the mass population mixing of World War II, the subsequent return of individuals to their countries of

origin and the first large-scale parenteral treatments. It can be speculated that these global-scale events were instrumental in the worldwide mixing and spread of a large pool of strains which then disseminated locally as a consequence of commonplace parenteral treatment and IDU of the 1960s. Examination of the Scottish dataset within the global context also supports this theory. The corresponding coalescence dates of the Scottish and global datasets, the intercalation of Scottish Gt1a clusters amongst RoW sequences and the similarity in dates between the MRCA of the Scottish clusters and global lineage expansion suggests extant lineages were introduced into Scotland from multiple independent sources concurrently with the diversification and dissemination of HCV worldwide.

Subsequent expansion of a proportion of lineages occurred within Scotland and there is no evidence of onward transmission or new introductions with the exception of the Aberdeen-specific clade C4, which was first detected in this study in the 1990s. Interestingly the MRCA of C4 with the RoW pool dates to 1928, pre-dating the emergence of other Scottish clusters from the global pool; it may have remained unobserved earlier due to the small sample numbers from Aberdeen or, more likely, this strain originates from a country not included in the RoW dataset. The lack of onward transmission of this strain to other Scottish regions suggests that Aberdeen may operate within a different network, perhaps linked more closely to international HCV transmission networks. Although a small city, Aberdeen is a major oil-producing centre with consequent international population links, including connections to regions where information on HCV is sparse such as central Asia.

The RoW dataset for Gt3a consisted of only 44 sequences. Although more NS5B sequences with data on sampling date and geographic origin were available from international databases, these sequences were short in length (approximately 300 bp) and/or only partially overlapped the genome region used in the current study. As short sequences can adversely affect the calculation of important parameters in MCMC analyses, they were not included in this analysis. The evolutionary rate of the Gt3a sequences from the global dataset was 1.65 x 10^{-3} s/s/y which is slightly faster than previously noted [1.3 x 10^{-3} s/s/y, (20)]. These sequences coalesced in 1899 (1865-1932) similar to previous studies [1905 (1851-1932), (21); 1920s, (22)], but another study has estimated that extant global Gt3a sequences coalesced much earlier, approximately 300 years ago (20). In the latter study however three sequences from Pakistan formed a distinct cluster which was phylogenetically distant from the remainder of the Pakistani sequences and sequences from other countries; without this outlying cluster the remaining global Gt3a pool coalesces approximately 80 years ago, supporting our findings.

Although robust contextualisation of the Scottish Gt3a sequences within the global setting was not possible due to the excess of Scottish compared to RoW sequences within the dataset, the Scottish sequences did however form distinct clusters. Gt3a lineage expansion occurred in two periods of exponential growth, peaking in 1940 and 1960. The earlier time period parallels the MRCAs of the Scottish Gt3a clusters (1926 to 1942) and suggests lineage expansion during World War II was responsible for the introduction of HCV Gt3a into Scotland, similar to Gt1a. The second peak of Gt3a lineage expansion (1951-1975) suggests that migration from the Indian subcontinent may be the source of currently circulating Gt3a sequences in Scotland. Approximately 1% of the Scottish population originates from Pakistan where epidemic

312 Gt3a transmission has been shown to have occurred earlier than in any other country (22).
313 Extant Gt3a strains in the UK are considered to originate from the HCV-endemic Indian
314 subcontinent (23) via migration which peaked in the 1950s and 1960s and subsequently these
315 sequences progressed into and expanded within the PWID community (20). Resolving which of
316 the two time periods led to the extant pool of Scottish Gt3a strains would require an increase in
317 global Gt3a sequences, particularly from endemic regions of the world.
318 The transmission of HCV Gt1a and Gt3a within Scotland's PWID community was
319 reconstructed. Since the predominant strains differed between the different Scottish regions it
320 is likely that they represent separate introductory events from the global pool. An alternative
321 hypothesis for the regional specificity of HCV strains is the variability in the genetic make-up of
322 the host population in Scotland, as revealed by a recent study which shows separate genetic
323 clustering of individuals from Aberdeenshire and WCS (24). The earliest HCV sequences in the
324 study were from individuals living in Glasgow (Gt3a) or Edinburgh (Gt1a) suggesting that these
325 two major cities constituted initial hubs of infection. Although index Gt1a sequences occurred
326 in Edinburgh, they were mainly transmitted within that region alone; in contrast widespread
327 Gt1a transmission in Glasgow occurred ten years later but constituted the dominant cluster
328 that subsequently spread to surrounding regions. Taken together this suggests that Glasgow is
329 the key region driving HCV transmission in Scotland. An analysis of HCV infection hotspots in
330 Glasgow was performed at a finer geographical precision than previously attempted. A
331 disproportional amount of sequences were derived from individuals residing in a few PC sectors
332 in a cluster-specific manner suggesting that these regions represent key individual networks

that play a central role in HCV transmission within the city. Targeting intervention and treatment initiatives to these regions could aid their overall effectiveness.

It is curious that despite the co-circulation of both genotypes within Glasgow, co-infection with multiple HCV genotypes is relatively rare in Scotland (25), with an overall rate of 4.1% in the current study. This is supported by a recent modelling study (26) suggesting that HCV re-infections frequently result in spontaneous clearance and by empirical data from a number of studies showing re-infection rates in PWID of 2-9% (27-30). Other studies have suggested that re-infection is much more common in PWID [20-39% (31-33)] and differences may be due to the provision of needle exchange programmes in the regions studied or sensitivities of the assays used. If re-infection is a rare occurrence (26), it is unlikely that sequences from individuals infected before the nationwide dispersal of strains would be supplanted by a recently imported strain from a different region and newly-infected individuals would become the main source of such introduced strains. Since the 1990s only a few new lineages have been imported from the global pool that have undergone onward transmission, notably the Aberdeen-specific Gt1a cluster C4. It is interesting to speculate whether the continued geographical restriction of this clade is due to a later introduction into Scotland subsequent to the full implementation of clean needle strategies or to separate networks operating in Aberdeen.

The combination of epidemiology data and molecular phylogenetics has been previously used to construct injecting social networks for the investigation of HCV transmission dynamics (10). Researchers established an association between reported injecting relationship and HCV phylogeny but not between genetic and social distance. A disadvantage of this study type is that

355 transmission inference is based on phylogenetic data derived from strains of long-standing
356 infection combined with information on current injecting partners, similar to the
357 contemporaneous residential data used in the current study. Other investigators have taken
358 phylogenetic clustering outcome as direct evidence of individual membership in a contact
359 network (35-39).
360 Spatiotemporal analysis based on combined phylogenetic clustering data, EYI's and
361 residential PC information for each sample was used in this study. There are limitations
362 associated with utilising this data for time-correlated phylogeographic reconstruction of chronic
363 viruses such as HCV. Firstly diagnosis normally occurs many years subsequent to infection. To
364 overcome this drawback we calculated the average age individuals commenced IDU and
365 surmised HCV infection within the first year of injecting. Studies have suggested that new
366 initiates to injecting are particularly vulnerable to HCV infection, seroconverting in
367 approximately 4 months (40). As well as new infections, PWID may experience multiple
368 exposures to HCV, with the possibility that a new strain may infect and supplant the original
369 strain, thus affecting the EYI. Recent modelling studies however suggest that HCV re-infection
370 generally results in spontaneous clearance of the newly infecting strain (26) and this is
371 supported by experimental data indicating a low incidence of co-infection with more than one
372 HCV genotype (25). The second limitation arises from combining current residential with
373 historical EYI data. Studies however suggest that PWID in the UK are not a highly mobile group
374 of individuals (34, 41, 42). Nevertheless it is likely that the discernment of transmission
375 networks at a finer temporal and spatial resolution could be achieved by either the availability

of residential data from the EYI or, alternatively, obtaining samples from newly-infected individuals in concert with contemporary residential information.

Although we observed a decade of increased transmission commencing in the mid-1980s, this observation may be subject to a degree of sample bias. The average age of HCV-infected individuals was 40 years, equivalent to an EYI in the mid-1990s. To enrich for samples from PWID, sequences from subjects with an EYI before 1989, the year widespread testing for HCV in blood products commenced in the UK, were excluded unless they specifically confirmed IDU; this has likely resulted in an augmentation of samples from patients born after 1969. Individuals born in the 1950s (EYI in the 1970s) are subject to naturally higher, age-associated mortality rates whereas more recently-acquired HCV strains may not be captured in this cohort of hospital-diagnosed individuals due to the extensive asymptomatic period of HCV infection. Previous epidemiological studies have however shown a similar pattern of peak HCV transmission occurring in Glasgow in the 1980's (4, 34).

Through this research we reconstructed the dispersal of HCV into and throughout Scotland and successfully identified transmission hotspots that could enable healthcare and outreach workers to effectively target intervention and treatment as prevention strategies and monitor the effectiveness of these methods. The accuracy of this strategy could be further improved by increasing the sample size from individual cities and analysing sequences from recently-infected individuals with current residential data. Another tool that could help improve HCV transmission inference studies is next generation sequencing (NGS) which has recently been investigated for analysing small-scale transmission events and within-patient connections (43-45). Minor sequence variants in one individual may predominate in a different individual, as

illustrated by studies of liver transplants in patients with HCV (46, 47), and NGS may enable the identification of connections involving minority variants which are not apparent using traditional sequencing techniques. We now intend to utilise NGS methodology to expand this study, investigating HCV transmission on a UK-wide basis.

In summary, heterochronous HCV sequences were used in this study to determine the timeline of introduction of extant Gt1a and Gt3a into Scotland from the global pool and to infer the spatiotemporal distribution of these sequences in the regional setting. Most Scottish sequences formed discrete clusters interspersed between RoW sequences and the MRCAs of these Scottish clusters were dated to the period of global exponential lineage expansion. Although the origin of the two epidemic HCV genotypes in Scotland differed (Gt1a, Edinburgh; Gt3a, Glasgow), transmission to other regions occurred predominantly from Glasgow. Geographical specificity of transmitting clusters was apparent, particularly for Gt1a. In Glasgow individual clusters formed different transmission zones initially but these networks subsequently overlapped suggesting co-circulation of genotypes and clusters has occurred throughout the city since the 1990s. Hubs of infection were detected that likely play a key role in HCV transmission throughout the city; targeting intervention strategies to these regions could assist their effectiveness.

ACKNOWLEDGEMENTS

This study was supported by MRC grant MC_UU_12014/1 from the Medical Research Council.

REFERENCES

1. **Shephard CW, Finelli L, Alter MJ.** 2005. Global epidemiology of hepatitis C virus infection. Lancet Infect Dis **5**:558-567.
2. **Martin NK, Vickerman P, Grebely J, Hellard M, Hutchinson SJ, Lima VD, Foster GR, Dillon JF, DJ Goldberg, GJ Dore, M Hickman**. 2013. Hepatitis C virus treatment for prevention among people who inject drugs: modelling treatment scale-up in the age of direct-acting antivirals. Hepatology **58**:1598-1609.
3. **Public Health England**. 2013. Hepatitis C in the UK: 2013 Report. Available: http://www.hpa.org.uk/webc/HPAwebFile/HPAweb_C/1317139502302.
4. **Hutchinson SJ, Bird SM, Taylor A, Goldberg DJ.** 2006. Modelling the spread of hepatitis C virus infection among injecting drug users in Glasgow: implications for prevention. Int J Drug Policy **17**:211-221.
5. **Anglemyer A, Rutherford GW, Baggaley RC, Egger M, Siegfried N.** 2011. Antiretroviral therapy for prevention of HIV transmission in HIV-discordant couples. Cochrane Database Syst Rev **8**: CD009153. doi: 10.1002/14651858.CD009153.pub2.
6. **Pybus OG, Rambaut A.** 2009. Evolutionary analysis of the dynamics of viral infectious disease. Nature Reviews Genetics **10**:540-550.
7. **Lemey P, Rambaut A, Drummond AJ, Suchard MA.** 2009. Bayesian phylogeography finds its roots. PLoS Comput Biol **5**:e10000520.
8. **Morelli MJ, Thebaud G, Chadoeuf J, King DP, Haydon DT, Soubeyrand S.** 2012. A Bayesian inference framework to reconstruct transmission trees using epidemiological and genetic data. PLoS Comput Biol **8**:e1002768.

9. **Azman AS, Luquero FJ, Rodrigues A, Palma PP, Grais RF, Banga CN, Grenfell BT, Lessler J.** 2012. Urban Cholera Transmission Hotspots and Their Implications for Reactive Vaccination: Evidence from Bissau City, Guinea Bissau. PLoS Negl Trop Dis **6**:e1901.

10. **Sacks-Davis R, Daraganova G, Aitken C, Higgs P, Tracy L, Bowden S, Jenjinson R, Rolls D, Pattison P, Robins G, Grebely J, Barry A, Hellard M.** 2012. Hepatitis C virus phylogenetic clustering is associated with the social-injecting network in a cohort of people who inject drugs. PLoS ONE **7**:e47335.

11. **Romano CM, Guedes de Carvalho-Mello IMV, Jamal LF, Lucas de Melo F, Motoki M, Pinho JRR, Holmes E, de Andrade Zanotto MP, the VGDN Consortium.** 2010. Social networks shape the transmission dynamics of hepatitis C virus. PLoS ONE **5**:e11170.

12. **Simmonds P.** 2012. SSE: A nucleotide and amino acid sequence analysis platform. BMC Research Notes. **5**:50.

13. **Drummond AJ, Suchard MA, Dong X, Rambaut A.** 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol **29**:1969-1973.

14. **Rambaut A, Suchard MA, Xie D, Drummond AJ.** 2014. Tracer v1.6. Available: http://beast.bio.ed.ac.uk/Tracer

15. **Hasegawa M, Kishino H, and Yano T.** 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol **22**:160-174.

16. **Drummond AJ, Rambaut A, Shapiro B, Pybus OG**. 2005. Bayesian Coalescent Inference of Past Population Dynamics from Molecular Sequences. Mol Biol Evol **2**2:1185-1192.

17. **Gray RR, Parker J, Lemey P, Salemi M, Katzourakis A, Pybus OG**. 2011. The mode and tempo of hepatitis C virus evolution within and among hosts. BMC Evolut Biol **11**:131-140.

18. **Magiorkinis G, Magiorkinis E, Paraskevis D, Ho SYW, Shapiro B, Pybus OG, Alain J-P, Hatzakis A.** 2009**.** The global spread of hepatitis C virus 1a and 1b: a phylodynamic and phylogeographic analysis. PLoS Medicine **6**:e1000198.

19. **Yuan M, Lu T, Li C, Lu L**. 2013. The evolutionary rates of HCV estimated with subtype 1a and 1b sequences over the ORF length and in different genomic regions. PLoS ONE **8**:e64698.

20. **Zehender G, Sorrentino C, Lai A, Ebranati E, Gabanelli E, Lo Presti A, Vojosevic D, Lausevic D, Terzic D, Shkjezi R, Bino S, Vratnica Z, Mugosa B, Galli M, Ciccozzi M.** 2013. Reconstruction of the evolutionary dynamics of hepatitis C virus subtypes in Montenegro and the Balkan region. Infect Genet Evol **17**:223–230.

21. **Rehman IU, Idrees M, Ali M, Liaqat A, Butt S, Hussain A, Akbar H, Afzal S.** 2011. Hepatitis C virus genotype 3a with phylogenetically distinct origin is circulating in Pakistan. Genet Vaccines Ther **9**: 1-2.

22. **Khan A, Tanaka Y, Azam Z, Abbas Z, Kurbanov F, Saleem U, Hamid S, Jafri W, Mizokami M.** 2009. Epidemic spread of hepatitis C virus genotype 3a and relation to high incidence of hepatocellular carcinoma in Pakistan. J Med Virol **81**:1189–1197.

23. **Pybus OG, Markov PV, Wu A, Tatem AJ.** 2007. Investigating the endemic transmission of the hepatitis C virus. Int J Parasitol **37**:839–849.

24. **Leslie S, Winney B, Hellenthal G, Davison D, Boumertit A, Day T, Hutnik K, Royrvik EC, Cunliffe B, Wellcome Trust Case Control Consortium, International Multiple Sclerosis Genetics Consortium, Lawson DJ, Falush D, Freeman C, Pirinen M, Myers S, Robinson M, Donnelly P, Bodmer W.** 2015. The fine-scale genetic structure of the British population. Nature **519**:309-314.

25. **McNaughton AL, Thomson EC, Templeton K, Gunson RN, Leitch ECM**. 2014. Mixed Genotype Hepatitis C Infections and Implications for Treatment. Hepatology **59**:1209.

26. **Sacks-Davis R, McBryde E, Grebely J, Hellard M, Vickerman P.** 2014. Many hepatitis C reinfections that spontaneously clear may be undetected: Markov-chain Monte Carlo analysis of observational study data. J R Soc Interface **12**:1197-1209.

27. **Schroter M, Feucht H-H, Zollner B, Schafer P, Laufs R.** 2003. Multiple infections with different HCV genotypes: prevalence and clinical impact. J Clinical Virol **27**:200-204.

28. **Buckton AJ, Ngui S-L, Arnold C, Boast K, Kovacs J, Klapper PE, Patel B, Ibrahim I, Rangarajan S, Ramsay ME, Teo C-G.** 2006. Multitypic hepatitis C virus infection identified by real-time nucleotide sequencing of minority genotypes. J Clinical microbial **44**:2779-2784.

29. **Butt S, Idrees M, Rehman IU, Akbar H, Shahid M, Afzal S, Akbar H, ur Rehman I.** 2011. Mixed genotype infections with hepatitis C virus, Pakistan. Emerg Infect Dis **17**:1565-1566.

30. **Lee Y-M, Chen Y-J, Lee C-M, Kuo L-H, Wong W-W, Chen Y-MA.** 2011. Detection of hepatitis C virus subtypes 6a, 6n, 6w and mixed infections using a modified multiplex real-time polymerase chain reaction protocol. J Formosan Med Assoc **110**:762-767.

31. **van de Laar TJW, Molenkamp R, van den Berg C, Schinkel J, Beld MGHM, Prins M, Coutinho RA and Bruisten SM.** 2009. Frequent HCV reinfection and superinfection in a cohort of injecting drug users in Amsterdam. J Hepatol **51**:667-674.

32. **Pham ST, Bull RA, Bennett JM, Rawlinson WD, Dore GJ, Lloyd AR, White PA.** 2010. Frequent multiple hepatitis C virus infections among injection drug users in a prison setting. Hepatology **52**:1564-1572.

33. **Herring BL, Page-Shafer K, Tobler LH, Delwart EL.** 2004. Frequent hepatitis C virus superinfection in injection drug users. J Infect Dis **190**:1396-1403.

34. **Taylor A, Goldberg D, Hutchinson S, Cameron S, Gore SM, McMenamin J, Green S, Pithie A, Fox R.** 2000. Prevalence of hepatitis C virus infection among injecting drug users in Glasgow 1990-1996: are current harm reduction strategies working? J Infect **40**:176-183.

35. **Cochrane A, Searle B, Hardie A, Robertson R, Delahooke T, Cameron S, Tedder RS, Dusheiko GM, de Lamballerie X, Simmonds P.** 2002. A genetic analysis of hepatitis C virus transmission between injection drug users. J Infect Dis **186**:1212-1221.

36. **Hope VD, Hickman M, Ngui SL, Jones S, Telfer M, Bizzarri M, Ncube F, Parry JV.** 2011. Measuring the incidence, prevalence and genetic relatedness of hepatitis C infections among a community recruited sample of injecting drug users, using dried blood spots. J Viral Hepat **18**:262-270.

37. **Oliveira MdLA, Bastos FI, Telles PR, Hacker MdA, Oliveira SANd, Miguel JC, Yoshida CFT.** 2009. Epidemiological and genetic analyses of hepatitis C virus transmission among young/short- and long-term injecting drug users from Rio de Janeiro, Brazil. J Clin Virol **44**:200-206.

38. **Matthews GV, Pham ST, Hellard M, Grebely J, Zhang L, Oon A, Marks P, van Beek I, Rawlinson W, Kaldor JM, Lloyd A, Dore GJ, White PA, the ATAHC Study Group.** 2011. Patterns and characteristics of hepatitis C transmission clusters among HIV-positive and HIV-negative individuals in the Australian trial in acute hepatitis C. Clin Infect Dis **52**:803-811.

527 39. **Jacka B, Applegate T, Krajden M, Olmstead A, Harrigan PR, Marshall BDL, DeBeck K, Milloy**
528 **MJ, Lamoury F Pybus OG, Lima VD, Magiorkinis G, Montoya V, Montaner J, Joy J, Woods C,**
529 **Dobrer S, Dore GJ, Poon AFY, Grebely J.** 2014. Phylogenetic clustering of hepatitis C virus
530 among people who inject drugs in Vancouver, Canada. Hepatology **60**:1571-1580.
531 40. **Maher L, Li J, Jalaludin B, Chant KG, Kaldor JM.** 2007. High hepatitis C incidence in new
532 injecting drug users: a policy failure? Aust N Z J Public Health **31**:30-35.
533 41. **Hickman M, Hope V, Brady T, Madden P, Jones S, Holloway G, Ncube F, Parry J.** 2007.
534 Hepatitis C virus (HCV) prevalence and injecting risk behaviour in multiple sites in England in
535 2004. J Viral Hepat **14**:645-652.
536 42. **Harris M, Rhodes T.** 2013. Hepatitis C treatment access and uptake for people who inject
537 drugs: a review mapping the role of social factors. Harm Reduct J **10**:7.
538 43. **Campo DS, Dimitrova Z, Yamasaki L, Skums P, Lau DTY, Vaughan G, Forbi JC, Teo C-G,**
539 **Khudyakov Y.** 2014. Next-generation sequencing reveals large connected networks of intra-
540 host HCV variants. BMC Genomics **15**:S4.
541 **44. Kundu S, Lockwood J, Depledge DP, Chaudhry Y, Aston A, Rao K, Hartley JC, Goodfellow I, Breuer J**.
542 2013. Next-generation whole genome sequencing identifies the direction of Norovirus transmission
543 in linked patients. Clin Infect Dis **57**:407-414.
544 45. **Redd AD, Wendel SKJ, Longosz AF, Fogel JM, Dadabhai S, Kumwenda N, Sun J, Walker MP,**
545 **Bruno D, Martens C, Eshleman SH, Porcella SF, Quinn TC, Taha TE.** 2014. Evaluation of
546 postpartum HIV superinfection and mother-to-child transmission. AIDS **29**:1567-1573.

547

46. **Ramirez S, Perez-del-Pulgar S, Carrion JA, Coto-Llerena M, Mensa L, Dragun J, Garcia-Valdecasas JC, Navasa M, Forns X.** 2010. Hepatitis C superinfection of liver grafts: a detailed analysis of early exclusion of non-dominant virus strains. J Gen Virol **91**:1183-1188.

47. **Ohtsuru S, Ueda Y, Marusawa H, Inuzuka T, Nishijima N, Nasu A, Shimizu K, Koike K, Uemoto S, Chiba T.** 2013. Dynamics of defective hepatitis C virus clones in reinfected liver transplant recipients: ultradeep sequencing analysis. J Clin Microbiol **51**:3645-3652.

FIGURE LEGENDS

FIG 1 Bayesian time-scaled trees of the global Gt1a (A) and Gt3a (B) datasets with branches colour-coded according to the country of origin of the sequences. Clusters identified as containing predominantly Scottish sequences are highlighted and designated C1-C4 (Gt1a) or C1-C7 (Gt3a).

FIG 2 Geographical distribution of Gt1a (A) and Gt3a (B) clusters in Scotland. The type, frequency and number of samples of individual clusters in each geographical region is shown.

FIG 3 Geographical distribution of sequences within Glasgow. Sequences from the major Gt1a (C2 and C3) and Gt3a (C5 and C6) clusters are indicated on maps by the last two digits of the EYI and colour-coded over three transmission periods. The residential PC sectors associated with sequences are outlined in black. Likely transmission zones over each time period are highlighted.

FIG 4 Co-circulation of the major Gt1a and Gt3a clusters in Glasgow. Likely transmission zones of the predominant Gt1a (C2 and C3) and Gt3a (C5 and C6) clusters in Glasgow are highlighted over four time periods. Transmission zones are comprised of the residential PC sectors of the study participants and connecting regions.

576 **FIG 5** Map of Glasgow highlighting the predominant residential PC districts of individuals infected with
577 Gt1a (blue), Gt3a (red) or areas predominant for both genotypes (purple). The percentages of samples
578 associated with the districts are shown, coloured in blue (Gt1a) and red (Gt3a). Four PC sectors
579 representing hotspots of transmission are shown within PC districts, coloured by genotype and with the
580 associated percentage of samples derived from the sector. Insert shows a map of the complete Glasgow
581 PC region with the nine predominant PC districts in colour.

582

583

584

585

586 **TABLE 1** HCV genotype-specific primers for nested PCR

Genotype	Primer Sequence (5' to 3')	Genome Position[a]
1a	CRT ATG AYA CCC GCT GYT TTG AC	8254-8276
1a	CTC CAC AGT CAC TGA GAG CGA YAT	8276-8299
1a	AAT GCG CTR AGR CCA TGG AGT C	9016-8995
1a	CCT GGA GAG TAA CTR TGG AGT G	9040-9019
3a	CRT ATG AYA CCC GCT GYT TTG AC	8254-8276
3a	CTC NAC YGY CAC TGA RCA GGA YAT C	8276-8300
3a	CCA TGG AGT CTT TCA ATG ATT GCT G	9004-8980
3a	CTC TAC TGG AGA GTA ACT GTG GA	9044-9022

587 [a] The genome position with reference to strain H77

588

589 **TABLE 2** Measures of evolution of the HCV datasets

Genotype	Dataset	N	MRCA[a]	Rate (x 10^{-3} s/s/y)[b]
1a	Global	506	1888 (1857-1914)	1.88 (1.53-2.29)
1a	Scottish	125	1918 (1882-1951)	1.81 (1.19-2.48)
1a	RoW[c]	381	1907 (1879-1934)	1.84 (1.39-2.28)
3a	Global	244	1899 (1865-1932)	1.65 (1.19-2.14)
3a	Scottish	200	1901 (1866-1935)	1.62 (1.15-2.15)
3a	RoW	44	1897 (1860-1934)	1.53 (1.04-2.05)

590

591 [a] The most recent common ancestor (MRCA) of datasets with the 95% HPD is shown in

592 parenthesis.

593 [b] The evolutionary rate. s/s/y, number of substitutions/site/year.

594 [c] RoW, rest of the world dataset.

595

596

TABLE 3 Composition and MRCA of Scottish clusters

Gt[a]	Cluster	No. of Scottish sequences[b]	No. of RoW sequences (UK other)[c]	MRCA[d]
1a	1	14	1 (0)	1949 (1946-1953)
	2	34	4 (0)	1944 (1932-1956)
	3	43	0 (0)	1942 (1937-1951)
	4	8	0 (0)	1952 (1937-1962)
3a	1	27	6 (6)	1936 (1932-1940)
	2	5	2 (2)	1939 (1912-1956)
	3	12	1 (1)	1939 (1925-1954)
	4	6	2 (1)	1940 (1916-1960)
	5	72	4 (2)	1936 (1918-1956)
	6	33	1 (1)	1942 (1915-1963)
	7	36	5 (3)	1926 (1918-1947)

[a] HCV genotype.

[b] Number of sequences contained in Scottish clusters.

[c] Number of sequences in Scottish clusters originating from the rest of the world (RoW), with sequences from other parts of the UK excluding Scotland (UK other) shown in parenthesis.

[d] The most recent common ancestor (MRCA) of each cluster with the 95% HPD shown in parenthesis.

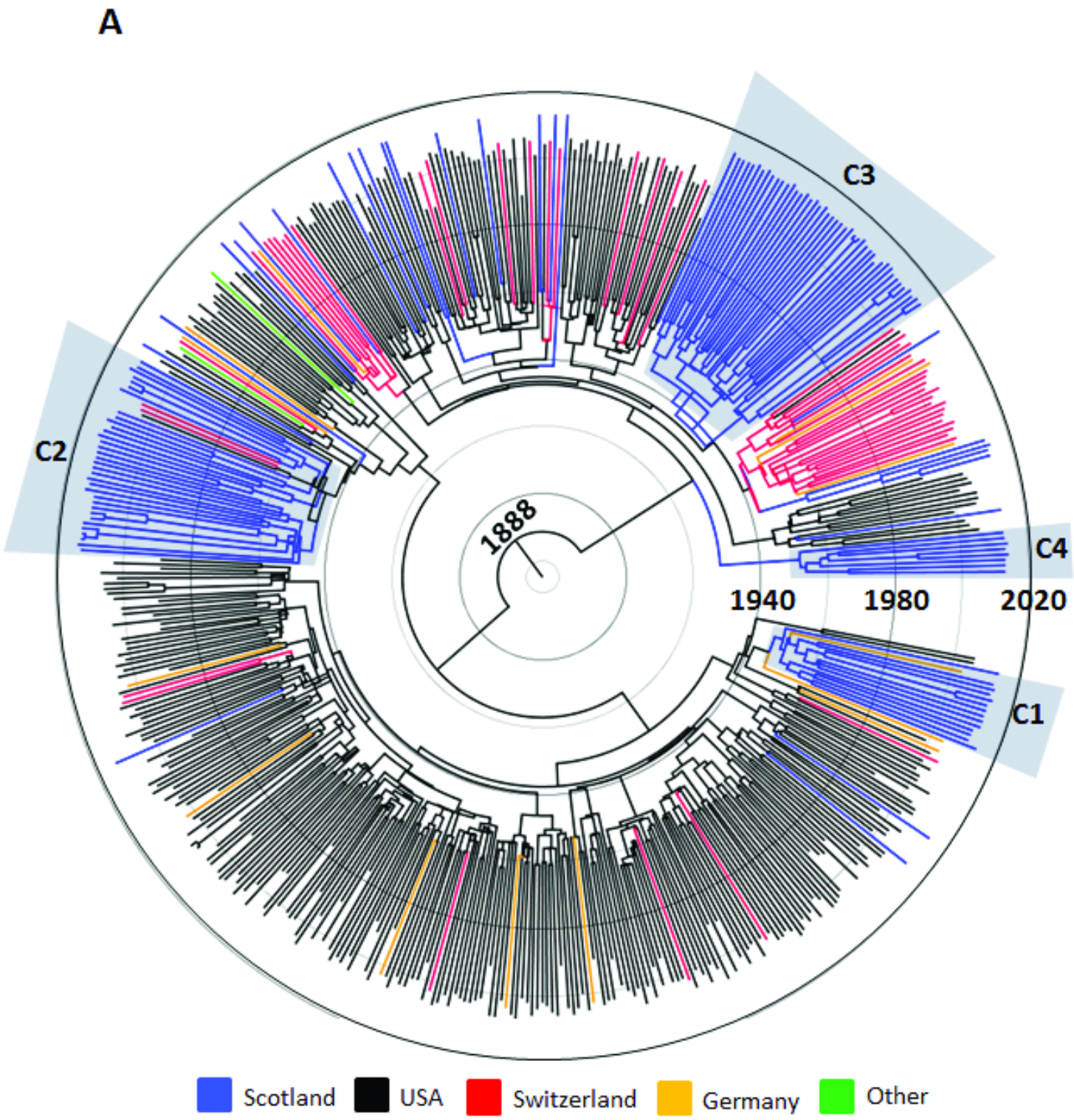
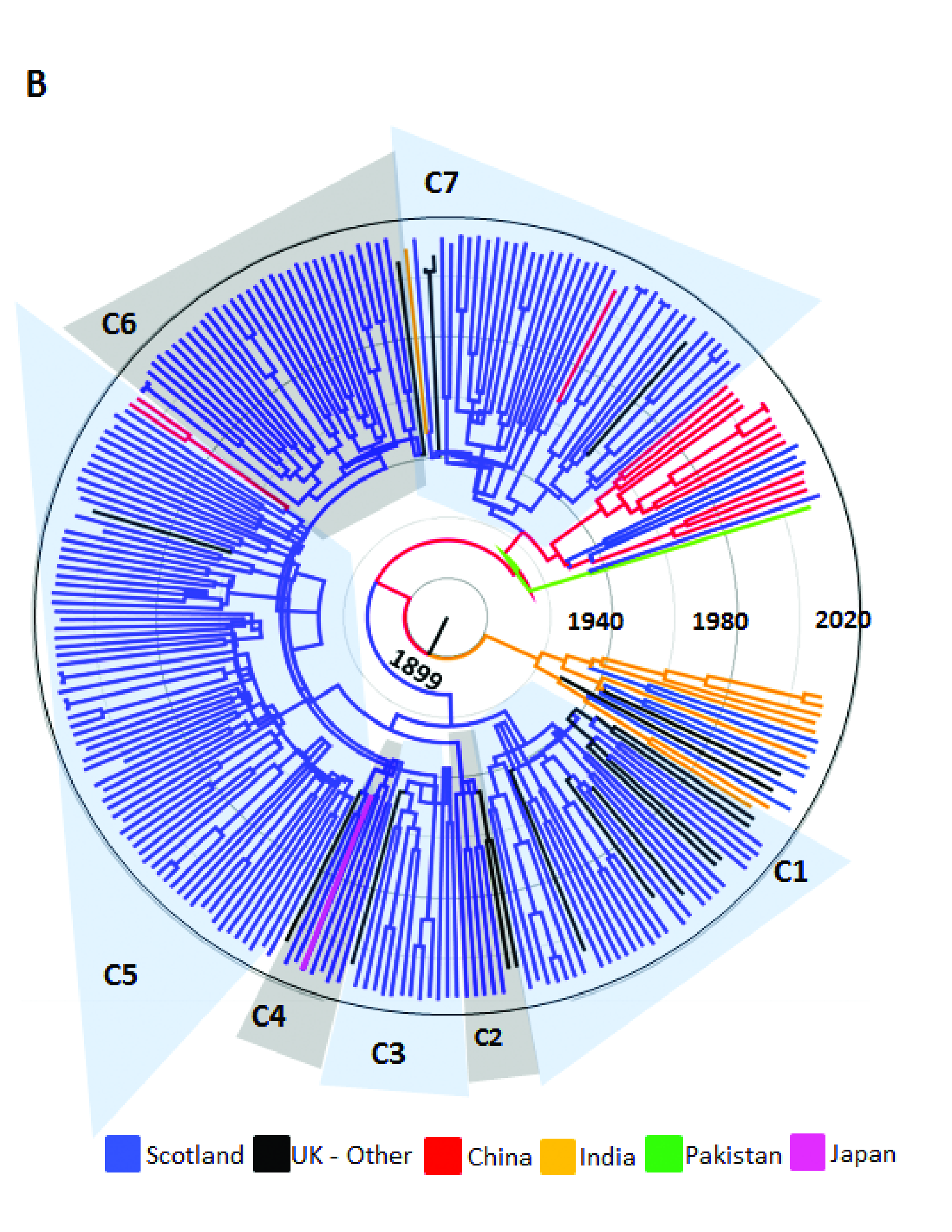


FIG 1 Bayesian time-scaled trees of the global Gt1a (A) and Gt3a (B) datasets with branches colour-coded according to the country of origin of the sequences. Clusters identified as containing predominantly Scottish sequences are highlighted and designated C1-C4 (Gt1a) or C1-C7 (Gt3a).


B
C7
C6
C5
C4
C3
C2
C1
1940
1980
2020
1899
Scotland
UK - Other
China
India
Pakistan
Japan

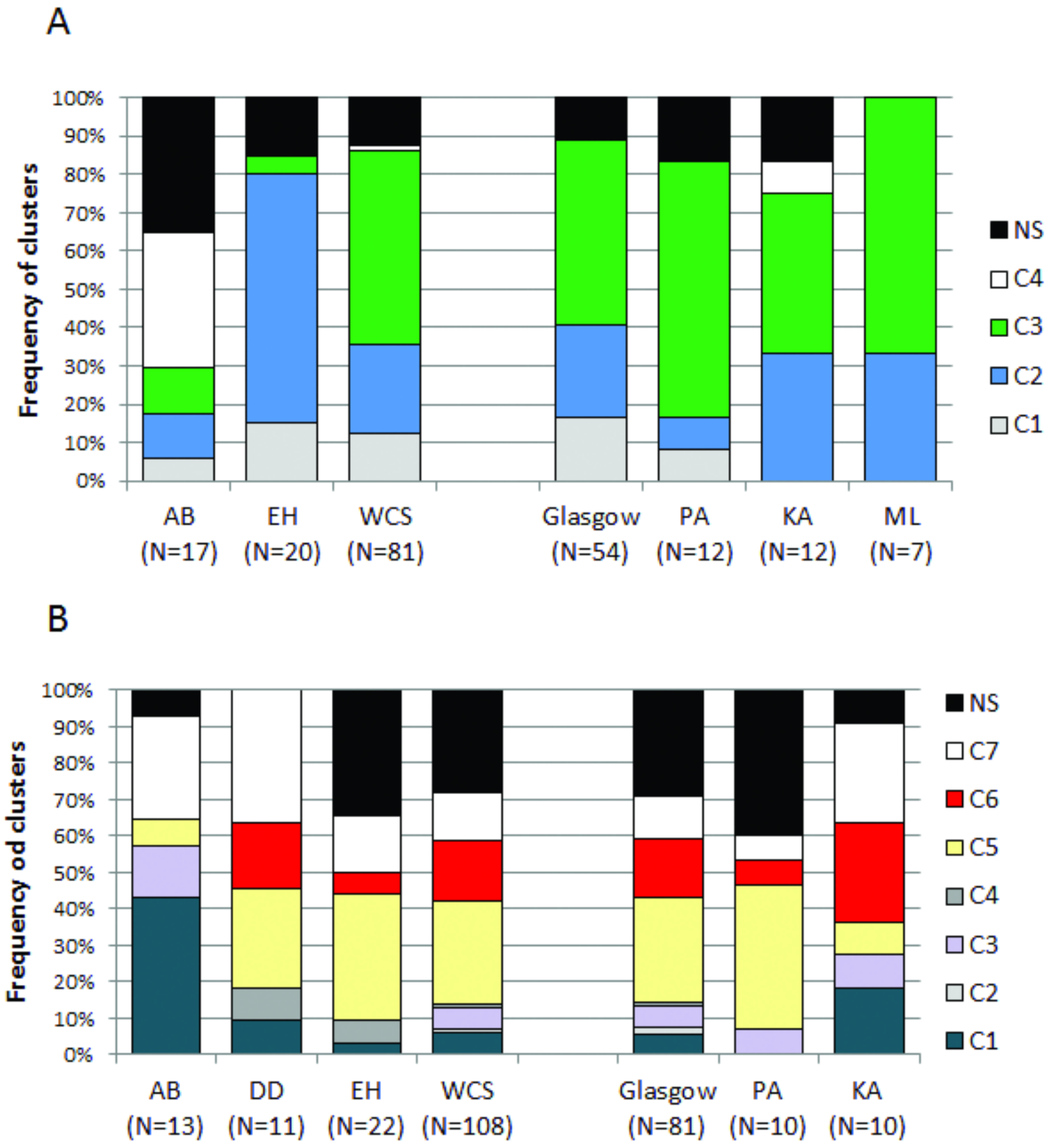


FIG 2 Geographical distribution of Gt1a (A) and Gt3a (B) clusters in Scotland. The type, frequency and number of samples of individual cluster in each geographical region is shown.

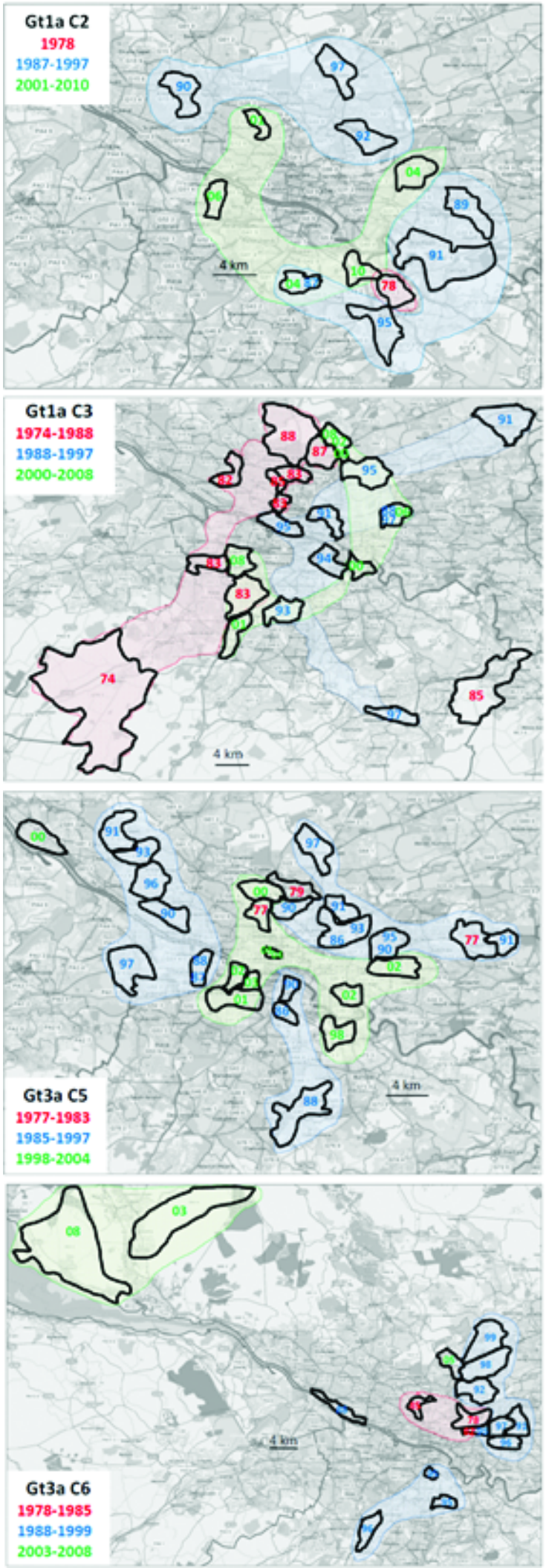


FIG 3 Geographical distribution of sequences within Glasgow. Sequences from the major Gt1a (C2 and C3) and Gt3a (C5 and C6) clusters are indicated on maps by the last two digits of the EYI and colour-coded over three transmission periods. The residential PC sectors associated with sequences are outlined in black. Likely transmission zones over each time period are highlighted.

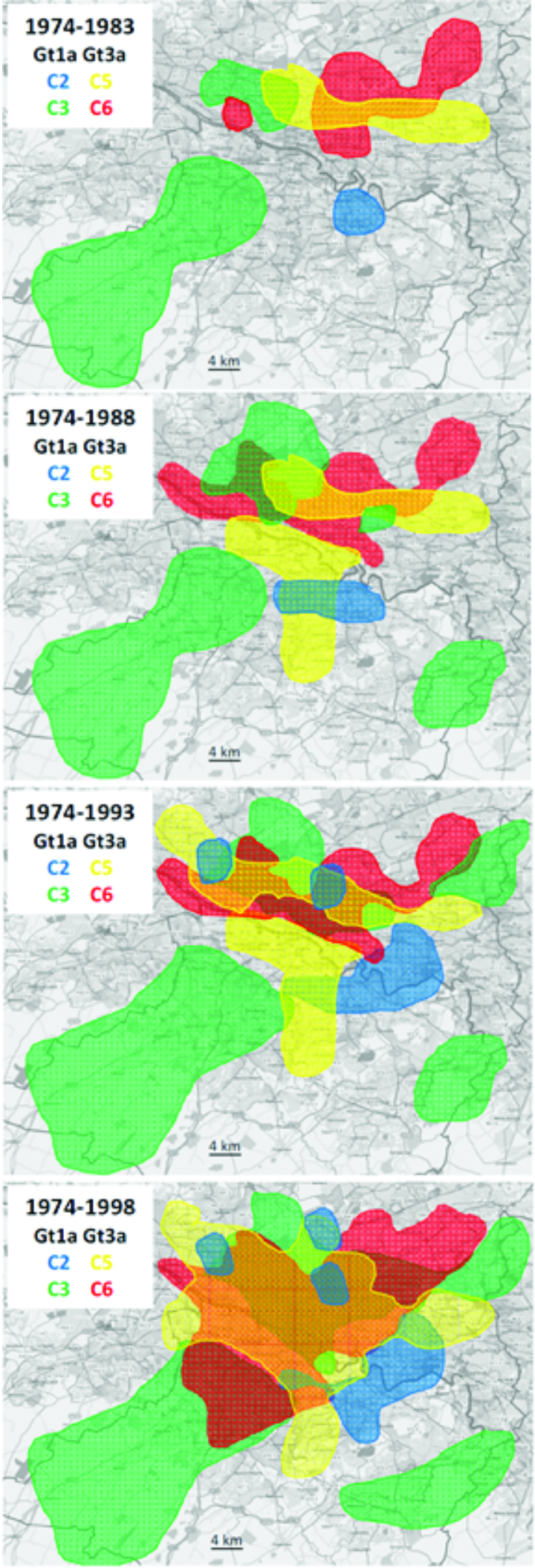


FIG 4 Co-circulation of the major Gt1a and Gt3a clades in Glasgow. Likely transmission zones of the predominant Gt1a (C2 and C3) and Gt3a (C5 and C6) clusters in Glasgow are highlighted over four time periods. Transmission zones are comprised of the residential PC sectors of the study participants and connecting regions.

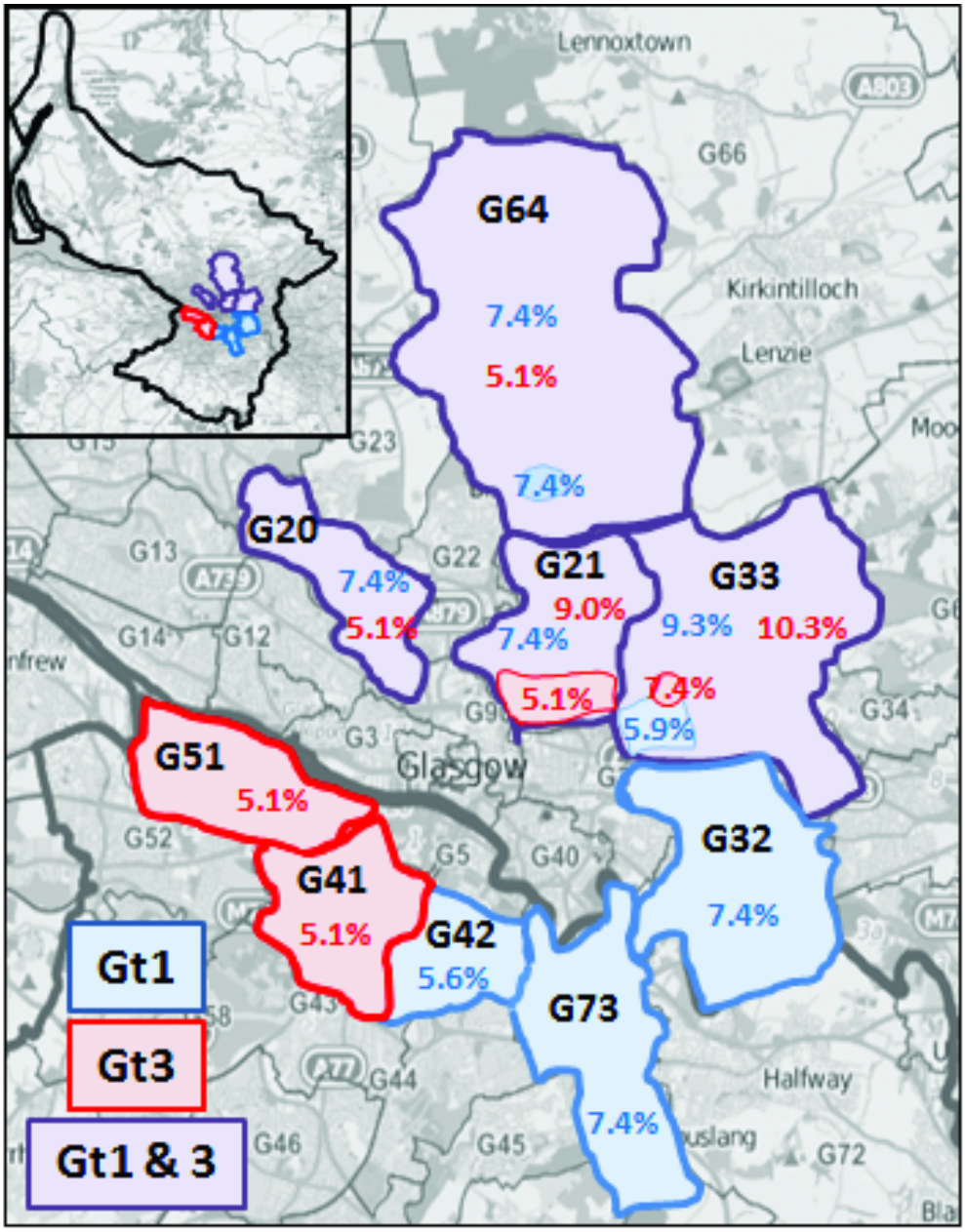


FIG 5 Map of Glasgow highlighting the predominant residential PC districts of individuals infected with Gt1a (blue), Gt3a (red) or areas predominant for both genotypes (purple). The percentages of samples associated with the districts are coloured in blue (Gt1a) and red (Gt3a). Four PC sectors representing hotspots of transmission are shown within PC districts, coloured by genotype and with the associated percentage of samples derived from the sector. Insert shows a map of the complete Glasgow PC region with the nine predominant PC districts in colour.