Phylodynamic Insights into the HIV Epidemic within Canada

Jeffrey B. Joy[1,2], Richard H. Liang[1], Bluma Brenner[3,4], Tarah Lynch[5], John Gill[5], Jared Buller[6], David Alexander[6], Zabrina Brumme[7], Ann Burchell[8], Sean Rourke[9], Mona Loutfy[9], Janet Raboud[9], Curtis Cooper[10], Deborah Kelly[11], Chris Tsoukas[4], Nima Machouf[12], Marina Klein[4], Alexander Wong[13], Paul Levett[14], Sean Hosein[15], Mark Wainberg[3,4], Paul Sandstrom[16], Julio S.G. Montaner[1,2], Robert S. Hogg[1,6], Art F.Y. Poon[17], P. Richard Harrigan[1,2], CANOC Collaboration

1. Centre for Excellence in HIV/AIDS, 2. University of British Columbia, 3. Lady Davis Institute for Medical Research, 4. McGill University, 5. Alberta Health Services, 6. University of Manitoba, 7. Simon Fraser University, 8. St. Michael's Hospital, 9. University of Toronto, 10. University of Ottawa, 11. Memorial University, 12. Université de Montreal, 13. University of Saskatchewan, 14. University of Regina, 15. CATIE, 16. Public Health Agency of Canada, 17. University of Western Ontario

Background

Despite the importance of HIV, the early history, geographic dissemination, and dynamics of the virus within populations across Canada remain unclear. Epidemiological processes stamp measurable signatures on HIV genomes sampled at different places and times. Using statistical phylodynamic approaches applied to HIV sequences sampled from across Canada (Figure 1) we test hypotheses concerning past and present HIV epidemic dynamics.

Methods

We compiled 51,493 doubly-anonymized HIV pol sequences from 14,000+ patients annotated with clinical and socio-demographic parameters. Data were available from 5 Canadian provinces: British Columbia, Alberta, Saskatchewan, Ontario, and Quebec. Analyses were restricted to the first sample collected from each patient and drug resistance codons were censored from the alignment. Phylogenetic trees were inferred using FastTree2. Phylogenetic clusters (Figure 2A) of five or more participants were identified using a tip-to-tip distance cutoff < 0.02 substitutions per site (Figure 2B). Diversification rate (Figure 3) analyses were conducted in R.

Results

We observed variation among provinces in the proportion of non-subtype B infections, with the Prairies (Alberta and Saskatchewan) displaying significantly greater numbers of non-B infections ($p<0.05$, Table 1). We recovered 285 clusters of size ≥ 5 (Figure 4). Cluster size was associated with proportion of people who inject drugs ($p<0.004$). Most provinces contain large, primarily province specific, clusters dominated by transmission through injection drug use. Some between-province clustering is observed (n=55 clusters including 3 or more provinces, Figure 5). Association of clusters with more than one province was associated with proportion MSM risk factor ($P<0.05$). Consistent with other evidence, the Prairies had the highest rates of HIV diversification (Figure 6).

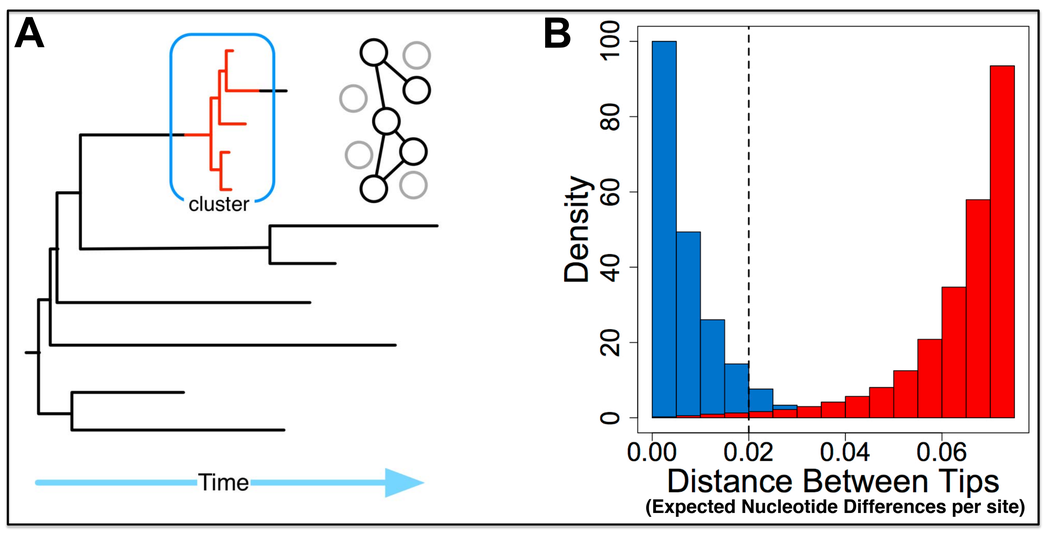

Figure 2. Inference of phylogenetic clusters. A. Since HIV evolves very rapidly, infections that retain significant genetic similarity are likely related by recent transmission events. Thus, clusters of genetically similar infections can represent localized outbreaks of HIV transmission. However, branching events on a phylogeny do not necessarily represent direct transmission because of the possibility of unsampled infections. **B.** Distribution of intra (blue) and inter (red) patient patristic (branch to branch) distances measured from a phylogenetic tree. The patristic distance cutoff is quantitatively assigned based on 95% quantile of intra-patient and 0.1% quantile of inter-patient distances.

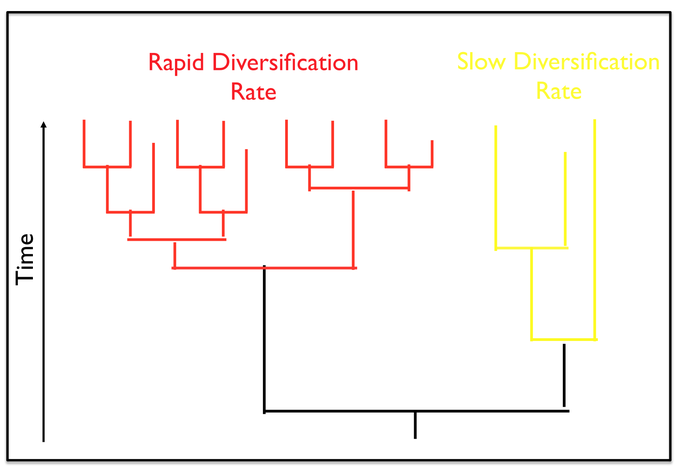

Figure 3. Diversification rate is the rate of splitting of branches in a phylogenetic tree, when combined with geographic annotations diversification rate tells us where epidemics are increasing most rapidly.

Figure 1. Map of Canadian Provinces and Territories.

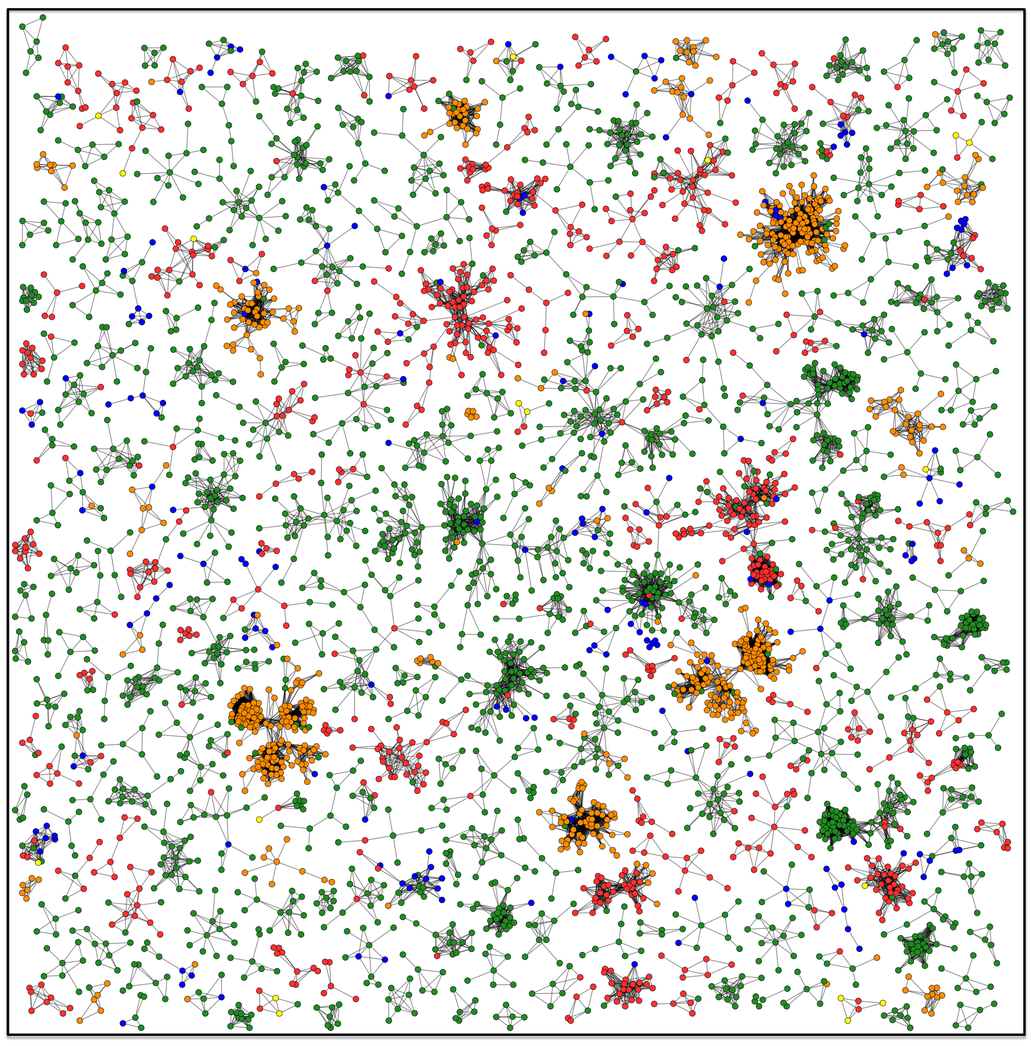

Figure 4. Phylogenetic clusters of HIV Transmission within Canada. 1 baseline sample from each patient was included in the phylogenetic tree and clusters of 5 or more inferred using a patristic distance cutoff of 0.02 substitutions/site (see Figure 2). Nodes in the graph are coloured by Canadian Province of sequence origin: green = British Columbia, blue = Alberta, orange = Saskatchewan, red = Ontario, and yellow = Quebec.

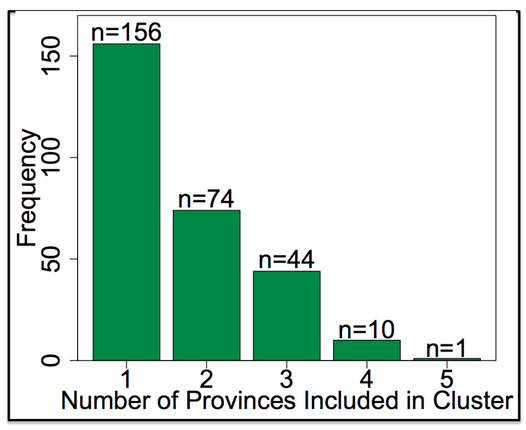

Figure 5. Summary of clustering within and between Canadian Provinces.

Province	# of Patients	Subtype B	MSM	Male
BC	7,712	0.91	0.68	0.80
AB	789	0.69	0.43	0.71
SK	987	0.84	0.02	0.60
ON	4,803	0.88	0.61	0.83
QC	401	0.66	0.36	0.69

Table 1. Dataset characteristics. Samples sizes of patients from each included Canadian Province and proportion subtype B, MSM and Male.

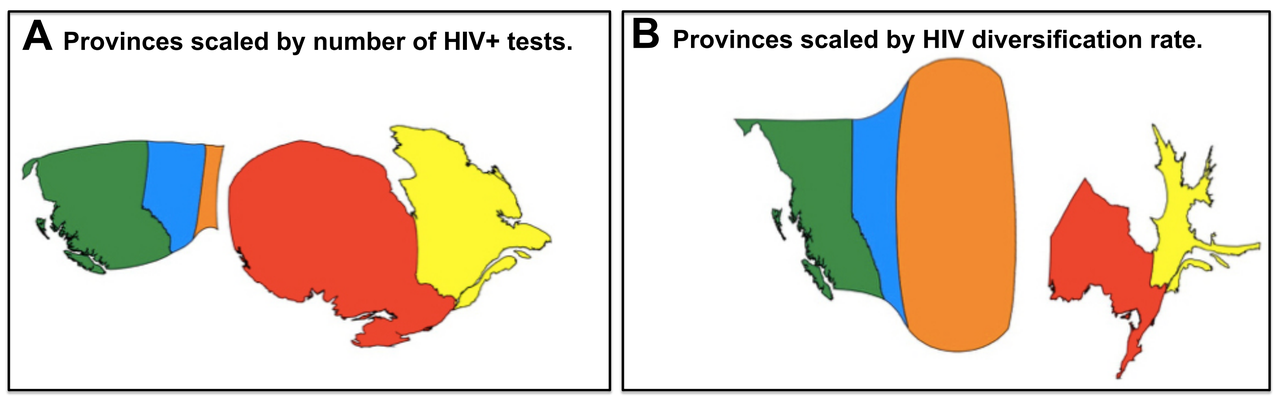

Figure 6. Comparison of HIV positive tests and HIV diversification rates across Canada. **A.** Map of 5 Canadian Provinces for which data are included transformed by number of HIV positive tests between 1985 and 2011 (data from PHAC population specific status report: people living with HIV). **B.** Map of 5 Canadian Provinces for which data are included transformed by HIV diversification rate illustrating a higher HIV diversification rate in the Province of Saskatchewan relative to other included Provinces.

Conclusion

Secondary analysis of genotypic resistance data provides useful epidemiological inferences on a national scale. Different circumstances permitted establishment, dissemination, and growth of the HIV epidemic in Canada at different times within component subpopulations. Our results emphasize the varied challenges facing different regions of Canada in controlling the HIV epidemic in the future.

Acknowledgments

Funding for this work was provided by Genome Canada, the BC Centre for Excellence in HIV/AIDS, and the Public Health Agency of Canada.