Commentary

Intra-Host Bayesian Phylogeography Linking Viral Evolution and Pathogenesis Comes to an Age

Brittany D. Rife [1,2], Marco Salemi [1,2*]

[1]Department of Pathology, Immunology, and Laboratory Medicine; [2]Emerging Pathogens Institute, University of Florida, Gainesville, FL 32610, USA.

Abstract | Bayesian phylogeographic analysis has recently been at the forefront of understanding epidemiological processes through the reconstruction of evolutionary history and spatial dissemination of pathogens. Infection of a single host can be viewed similarly as an epidemic spread on a microscale, creating a unique opportunity for the application of phylogeography at the intra-host level, which has until recently been limited by dataset quality and the lack of a rigorous statistical framework. As recently described in studies using the SIV-macaque model of neuroAIDS, the incorporation of spatial, temporal, and demographic inference into intra-host evolutionary analysis using Bayesian phylogeography allows for inference of pathogen migration between infected tissues and cell types, provides valuable insight into immunopathological processes, and ultimately paves the way for a better understanding of the interplay between viral evolution and pathogenesis.

Editor | Muhammad Munir, The Pirbright Institute, Compton Laboratory, UK.
Received | March 13, 2015; **Accepted** | March 26, 2015; **Published** | September 30, 2015
***Correspondence** | Marco Salemi, Emerging Pathogens Institute, University of Florida, Mowry Road, Gainesville, USA; **E-mail** | salemi@pathology.ufl.edu
DOI | http://dx.doi.org/10.17582/journal.bjv/2015/2.4.58.61
Citation | Rife, B. D. and M. Salemi. 2015. Intra-host Bayesian Phylogeography linking viral evolution and pathogenesis comes to an age. *British Journal of Virology,* 2(4): 58-61.

Contribution of microbial pathogen evolutionary processes both to epidemic spread within a susceptible host population and disease progression in the individual host is irrefutable, albeit often elusive. RNA viruses, such as hepatitis C virus (HCV), and members of the *Retroviridae* family, such as human and simian immunodeficiency viruses (HIV/SIV), are characterized by error-prone genome replication and elevated viral generation time resulting in a fast molecular clock within chronically infected subjects (Salemi, 2013; Gray et al., 2012a). Intra-host genetic diversity can reach as high as 0.1-10% over the span of just one year, while infection of and adaptation to new cell types/tissues, as well as viral gene flow (migration) between different anatomic compartments, often generates a complex metapopulation structure (Korber et al., 1994; Coombs et al., 1998; Gray et al., 2011; Gray et al., 2012b). Therefore, incorporation of spatial and temporal information – i.e. tissue/cell of origin and sampling time of viral sequences – into the evolutionary analysis is crucial for understanding how viral population dynamics are ultimately linked to disease. Knowledge of both the tempo and mode of viral dissemination within the host can provide valuable insight into the emergence of strains with increased virulence, the range of cell/tissue tropism, or the ability to evade host and synthetic antiviral responses (Salemi, 2013), which can pose a serious threat to the infected individual and the population at large, often resulting in augmented morbidity and mortality.

The first phylogeographic demonstration in 1876 consisted of a phylogeny superimposed on a map used to explain the distribution of lemurs across Asia

(Haeckel, 1876). Since then, the processes governing the geographical distribution of genealogical lineages has been widely studied in the context of epidemic spreads, aiding in the understanding of the spatial dynamics of viral migration and planning for the prevention of major transmission events (Faria et al., 2011). Similarly, given an alignment of viral sequences for a gene of interest sampled longitudinally from different tissues of an infected host, it is possible in principle to infer *in vivo* migration and diffusion patterns (Figure 1). During systemic infection, pathogens utilize the vascular system as means of transportation from one tissue to the next, often within diverse host cell vectors. Despite ecological differences, from a population genetic standpoint, the host anatomy resembles a demographically diverse geographical region in which viral transmission and spread occur.

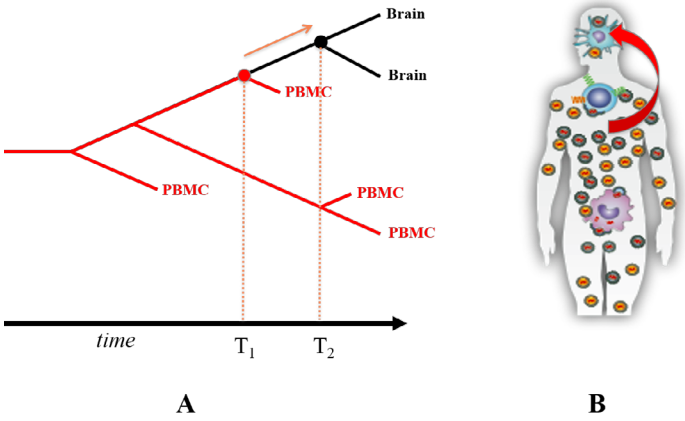

Figure 1: *Schematic representation of the intra-host phylogeography framework*
(A) *Longitudinally sampled viral sequences, amplified from various infected tissues and/or cell populations – e.g. peripheral blood mononuclear cells (PBMC) and brain – are used to infer a phylogenetic tree with branch lengths scaled in time by enforcing a molecular clock. Each sequence is assigned to a discrete category representing the tissue of origin and location of ancestral sequences, which is inferred by either a parsimony- or Bayesian-based algorithm. Viral migrations between tissues and/or cell populations are assumed to occur during the time interval defined by those branches, with parent (earlier sequence) and daughter (later sequence) nodes belonging to different anatomical compartments;* **(B)** *Viral migration patterns mapped in the host anatomy can provide insights on mechanisms of pathogen dissemination and disease progression.*

Recently, Strickland, Rife, and colleagues (2014) have shown how the Bayesian phylogeography framework (Lemey et al., 2009; Lemey et al., 2010), implemented in the BEAST software package (http://beast.bio.ed.ac.uk), can be adapted to investigate the spatiotemporal dynamics of SIV brain infection in an animal model of HIV-associated neurocognitive disorders (HAND). The Markov Chain Monte Carlo algorithm in BEAST incorporates both temporal (sampling time) and spatial (sampling location, e.g. tissue/cell type of origin) information in the phylogenetic reconstruction. Phylogenies are scaled in time by assuming either a strict or relaxed molecular clock (Drummond et al., 2006) and provide posterior probabilities for the location of the internal nodes of the tree (ancestral strains) that allow tracking SIV transmission/dissemination events between anatomic compartments over time. Thus far, intra-host Bayesian phylogeography studies have largely been limited by data set quality, i.e. the lack of longitudinal data and/or sequences derived from several tissue samples, which would require multiple biopsies rarely available from human subjects. Animal models, on the other hand, represent a viable alternative. Rhesus macaques inoculated with the SIVmac251 viral swarm develop AIDS and often display, as in HIV seropositive patients, neuropathology due to the accumulation of productively infected perivascular macrophages and microglia, which dysregulation leads to lesions in the central nervous system (CNS) (Burdo et al., 2013).

In order to define the evolutionary steps preceding and associated with the onset of neuroAIDS, it is necessary to address two main questions: when viral entry to the brain occurs and the tissue/cell type of origin of the neurotropic virions. Strickland, Rife and colleagues (2014) analyzed an extensive viral sequence dataset obtained from infected macaques, longitudinally sampled from early infection to terminal illness, including plasma, sorted peripheral T cells and monocytes, bone marrow and bronchoalveolar lavage fluid, as well as several CNS tissues. Individual sequences were assigned to discrete states corresponding to the individual tissues/cell populations from which they were sampled, and ancestral reconstruction of discrete states at internal nodes was implemented in the Bayesian statistical framework during the reconstruction of the posterior distribution of trees inferred using the non-parametric Bayesian Skyride demographic prior (Minin et al., 2008) and a relaxed clock (Drummond et al., 2006) model. Migration rates between these tissue/cell populations were assumed to follow a standard continuous-time Markov chain, in which transition rates between locations are reversible, or symmetrical, and the Bayesian stochastic search variable selection was applied to account for the large proportion of unlikely migration events over all possible events, thereby reducing variance estimates of ancestral states (Lemey et al., 2009). For phylogeographic analysis

pertaining to inter-host migration, the resulting tree file, with spatiotemporal information as to individual migration events, and a file specifying latitude and longitude information for each of the discrete states can be used in SPREAD (Bielejec et al., 2011) to create the file needed to visualize the migration events in real time using Google Earth, as in Lemey et al. (2009). However, as a program has not been developed to visualize viral migration within the body of rhesus macaques, an additional computational tool was developed for the generation of a special history log file, reporting within a specified period (migration period) the inferred transition along the branches at different time points, as drawn from their posterior distribution. This information was then translated into a matrix from which boxplots depicting median migration times from individual tissues to the brain could be generated in the R statistical package (http://www.r-project.org; Strickland and Rife et al., 2014).

In all SIV-infected macaques, phylogeographic analysis clearly showed continual viral seeding of the brain throughout infection – beginning as early as a few days post infection – originating from several peripheral tissues. Most importantly, the last viral migration event before terminal illness occurred from cells within the bone marrow. The result is particularly interesting in light of recent data demonstrating that monocyte-derived perivascular macrophages in the CNS are replenished from hematopoietic stem cells originating in the bone marrow and that increased monocyte turnover from bone marrow correlates with severity of neuropathology (Burdo et al., 2010). In other words, Bayesian phylogeography applied at the micro-scale of intra-host viral evolution is now emerging as an important tool in the realm of immunopathology, paving the way for a novel approach potentially able to link population genetics of the pathogen with disease processes in the host.

In conclusion, recent advancements in statistical methods of phylogenetic inference have provided the perfect opportunity for the merging of areas of epidemiology and molecular evolution. The result of this melding, the basis of Bayesian phylogeography, is a novel research framework that utilizes phylogenetic information and knowledge of the interplay between viral population dynamics and environmental/ecological factors in order to elucidate the spatiotemporal dynamics of viral evolution. These dynamics, as shown conclusively in Strickland and Rife et al. (2014) can also be explored at the level of the individual host, with the objective of characterizing specific viral migration events critical to pathogenesis, as is the case of progression to neuroAIDS in HIV-infected patients.

ACKNOWLEDGEMENTS

BDR and MS were supported by National Institutes of Health funding grants R01 NS063897-01A2.

REFERENCES

- Bielejec, F., Rambaut, A., Suchard, M.A., Lemey, P. (2011). SPREAD: Spatial Phylogenetic Reconstruction of Evolutionary Dynamics. *Bioinformatics* 27. 2910-12. http://dx.doi.org/10.1093/bioinformatics/btr481
- Burdo, T.H., Lackner, A., Williams, K.C. (2013). Monocyte/macrophages and their role in HIV neuropathogenesis. *Immunology Reviews* 254, 102-13. http://dx.doi.org/10.1111/imr.12068
- Burdo, T.H., Soulas, C., Orzechowski, K,. Button, J., Krishnan, A., Sugimoto, C., Alvarez, X., Kuroda, M.J., Williams, K.C. (2010) Increased monocyte turnover from bone marrow correlates with severity of SIV encephalitis and CD163 levels in plasma. *PLoS Pathogens* 6, e1000842. http://dx.doi.org/10.1371/journal.ppat.1000842
- Coombs RW, Speck CE, Hughes JP, Lee, W. Sampoleo, R., Ross, S.O., Dragavon, J., Peterson, G., Hooton, T.M., Collier, A.C., Corey, L., Koutsky, L., Krieger, J.N. (1998). Association between culturable human immunodeficiency virus type 1 (HIV-1) in semen and HIV-1 RNA levels in semen and blood: evidence for compartmentalization of HIV-1 between semen and blood. *Journal of Infectious Diseases* 177, 320-30. http://dx.doi.org/10.1086/514213
- Drummond, A.J., Ho, S.Y., Phillips, M.J. Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. *PLoS Biology* 4, e88. http://dx.doi.org/10.1371/journal.pbio.0040088
- Faria, N.R. Suchard, M.A., Rambaut, A., Lemey, P. (2011). Toward a quantitative understanding of viral phylogeography. *Current Opinions in Virology* 1, 423-9. http://dx.doi.org/10.1016/j.coviro.2011.10.003
- Gray, R.R., Salemi, M., Lowe, A., Nakamura, K.J., Decker, W.D., Sinkala, M., Mulligan, C.J., Thea, D.M., Kuhn, L., Aldrovandi, G., Goodenow, M.M. (2011). Multiple independent line-

ages of HIV-1 persist in breast milk and plasma. *AIDS* 25, 143- 52. http://dx.doi.org/10.1097/QAD.0b013e328340fdaf

- Gray, R.R., Salemi, M., Klenerman, P., Pybus, O.G. (2012a). A new evolutionary model for hepatitis C virus chronic infection. *PLoS Pathogens* 8, e1002656. http://dx.doi.org/10.1371/journal.ppat.1002656
- Gray, R.R., Strickland, S.L., Veras, N.M.C., Goodenow, M.M., Pybus, O.G., S. Lemon, M., Fried, M.W., Nelson, D.R., Salemi, M. (2012b). Unexpected Maintenance of Hepatitis C Viral Diversity Following Liver Transplantation. *Journal of Virology* 86, 8432-9. http://dx.doi.org/10.1128/JVI.00749-12
- Haeckel, E. (1876). The History of Creation: Or the Development of the Earth and Its Inhabitants by the Action of Natural Causes. *London: King*.
- Korber, B.T., Kunstman, K.J., Patterson, B.K., Furtado, M., McEvilly, M.M., Levy, R., Wolinsky, S.M. (1994). Genetic differences between blood- and brain-derived viral sequences from human immunodeficiency virus type 1-infected patients: evidence of conserved elements in the V3 region of the envelope protein of brain-derived sequences. *Journal of Virology* 68, 7467-81.
- Lemey, P., Rambaut, A., Drummond, A.J., Suchard, M.A. (2009). Bayesian phylogeography finds its roots. *PLoS Computational Biology* 5, e1000520. http://dx.doi.org/10.1371/journal.pcbi.1000520
- Lemey, P., Rambaut, A., Welch, J.J., Suchard, M.A. (2010). Phylogeography takes a relaxed random walk in continuous space and time. *Molecular Biology & Evolution* 27, 1877-85. http://dx.doi.org/10.1093/molbev/msq067
- Minin, V.N., Bloomquist, E.K., Suchard, M.A. (2008). Smooth Skyride through a rough Skyline: Bayesian coalescent-based inference of population dynamics. *Molecular Biology & Evolution* 25. 1458-71. http://dx.doi.org/10.1093/molbev/msn090
- Salemi, M. (2013). The intra-host evolutionary and population dynamics of human immunodeficiency virus type 1: a phylogenetic perspective. *Infectious Disease Reports* 5, s1e3. http://dx.doi.org/10.4081/idr.2013.s1.e3
- Strickland, S.L., Rife, B.D., Lamers, S.L., Nolan, D.J., Veras, N.M., Prosperi, M.C., Burdo, T.H., Autissier, P., Nowlin, B., Goodenow, M.M., Suchard, M.A., Williams, K.C., Salemi, M. (2014). Spatiotemporal dynamics of simian immunodeficiency virus brain infection in CD8+ lymphocyte-depleted rhesus macaques with neuroAIDS. *Journal of General Virology* 95, 2784-95. http://dx.doi.org/10.1099/vir.0.070318-0