Toward a quantitative understanding of viral phylogeography

Nuno Rodrigues Faria[1], Marc A Suchard[2,3,4], Andrew Rambaut[5,6] and Philippe Lemey[1]

Phylogeographic approaches help uncover the imprint that spatial epidemiological processes leave in the genomes of fast evolving viruses. Recent Bayesian inference methods that consider phylogenetic diffusion of discretely and continuously distributed traits offer a unique opportunity to explore genotypic and phenotypic evolution in greater detail. To provide a taste of the recent advances in viral diffusion approaches, we highlight key findings arising at the intrahost, local and global epidemiological scales. We also outline future areas of research and discuss how these may contribute to a quantitative understanding of the phylodynamics of RNA viruses.

Addresses
[1] Department of Microbiology and Immunology, Katholieke Universiteit Leuven, Leuven, Belgium
[2] Department of Biomathematics, David Geffen School of Medicine at UCLA, University of California, Los Angeles, CA 90095, United States
[3] Department of Human Genetics, David Geffen School of Medicine at UCLA, University of California, Los Angeles, CA 90095, United States
[4] Department of Biostatistics, UCLA School of Public Health, University of California, Los Angeles, CA 90095, United States
[5] Institute of Evolutionary Biology, University of Edinburgh, Ashworth Laboratories, Edinburgh, United Kingdom
[6] Fogarty International Center, National Institutes of Health, Bethesda, MD, United States

Corresponding author: Lemey, Philippe
(philippe.Lemey@rega.kuleuven.be)

Introduction

The ever-rising flood of viral genetic data and continual advances in statistical inference are fostering new opportunities for phylodynamic studies of infectious diseases. The central premise of this discipline is that epidemic processes, such as viral population growth and subdivision, leave a measurable imprint on the genome of viruses over the course of years, months or even days [1•,2•]. In these rapidly evolving pathogens, sequence evolution is occurring simultaneously with geographic dispersal; this interaction characterizes a spatial phylodynamic process that can be recovered from genomic data using phylogeographic analyses. Beyond a historical perspective on viral evolution, this discipline has the potential to help predict the emergence of infectious diseases by identifying key reservoir species and geographic areas from which new infections are likely to emerge and spread [3], and to elucidate the impact of animal movement or human mobility on viral disease spread.

Here, we review recent advances in our understanding of the complex interplay between molecular evolution and spatial dynamics of RNA viruses at different evolutionary scales. We highlight the role of state-of-the-art probabilistic methods that connect phylogenetic inference to a statistical description of trait evolution in providing novel phylodynamic insights. The studies we subsequently discuss serve as examples to illustrate emerging areas of future research.

Reconstructing spatial dynamics through time

As a primary tool for gaining insight into the origins of viral lineages, phylogenetic reconstruction has proven instrumental in molecular epidemiology [4]. Spatial processes are naturally embedded in phylogenetic trees or genealogies as a record of transitions between locations along each branch. In this sense, phylogeographic diffusion can be considered as a process of trait evolution where the particular trait — in this case, geographic location, is treated as an inherited property of the virus. The aim is then to estimate the ancestral locations in a phylogenetic tree conditional on the observed locations of viral sequences represented by the tips of the tree. Under this premise, a range of methods have been developed that can be categorized according to the criterion they use to chose between alternative hypotheses [5] and by the processes used to model how the traits change across the tree. In particular, different processes need to be assumed for discretely distributed traits, for example, for viruses sampled from different countries or cities, and continuously distributed traits, for example, when latitude and longitude coordinates are used as spatial locations for viral samples (Box 1).

Although the simplicity of parsimony reconstruction of discrete locations states at the internal nodes of the tree has been most appealing in viral epidemiology, probabilistic, and in particular Bayesian, statistical frameworks have recently been gaining ground due to their flexibility in hypothesis testing and the ready integration of ecological information [6•,7]. Recent advances in Bayesian

Box 1 Viral diffusion in discrete space

Phylogeographic inference has primarily considered diffusion of discrete traits such as spatial locations (e.g. Figure 1a) or body compartments, for example [12,48]. With analogy to the general time-reversal nucleotide substitution model, stochastic diffusion processes are modeled using continuous-time Markov chains (CTMC) (Figure 1d), which are fully characterized by infinitesimal rate matrices that contain symmetrical instantaneous rates of location exchange. A recent extension of this model accommodates asymmetrical transitions among locations in the phylogeny [49], which can provide a more realistic description of spatial dynamics in viral epidemics provided sufficient information is available to inform the more complex parameterization.

Because all rates are generally not required to adequately explain the diffusion process, the estimation procedure would gain efficiency through focusing on a limited set of well-supported migration pathways [9•]. The Bayesian phylogeographic framework accomplishes this by using a Bayesian stochastic search variable selection (BSSVS) procedure that allows rates to shrink to zero with some probability based on prior specification. This naturally leads to a Bayes factor (BF) test [50] to identify significant dispersal pathways by comparing the posterior to the prior odds that the transition rates are nonzero [9•].

Markov jumps and rewards
While none of the CTMC transitions between spatial locations in the phylogeny are directly observed, recent advances in explicit calculation [51,52] enable computation of the expected number of transitions (jumps) and waiting times in given locations (rewards) (Figure 1d). These jumps and rewards are possible on a branch-by-branch basis and can track specifically defined sets of transitions [53]. For example, it is often of interest to count the expected number of transitions out of a source-location or into a sink-location.

Continuous diffusion models
Model-based reconstructions of continuous-valued traits rely on a phylogenetic Brownian motion process [10•], the random walk model analog for discrete state CTMCs (Figure 1f). In contrast to the discrete phylogeographic approach, where estimates of ancestral locations are necessarily drawn from the set of sampled locations, Brownian diffusion models allow to fully explore two-dimensional space [6•,10•] and deliver a more realistic representation of the diffusion process, particularly for continuously distributed samples (Figure 1c). Inspired by relaxed molecular clock models [54], several extensions to relax the constant variance assumption of Brownian motion have been proposed, which permit diffusion rates to vary through time according to different underlying rate change processes.

Visualization tools
Bayesian inference of spatio-temporal dynamics incorporating phylogenetic and mapping uncertainty produces high-dimensional estimates that require dedicated visual summaries. Software to convert phylogenies annotated with divergence time and spatial estimates to keyhole markup language (KML), compatible with virtual globe software like Google Earth (earth.google.com) or GIS programs has recently been released ([55]; www.phylogeography.org). Figure 2 provides an example of such a visualization in Google Earth and Cartographica (http://www.macgis.com/).

inference methods are focused on directly linking patterns of genetic diversity to ecological processes, including changes in population size and substructure through time, allowing simultaneous insights into the spatial, temporal, and demographic dynamics of rapidly evolving pathogen populations [8,9•,10•]. We refer to Box 1 for a brief account on how Bayesian models of phylogenetic diffusion in discrete or continuous space constitute the spatial component of the integrated statistical framework in the BEAST software package [8]. In the next section, we also allude to other applications of phylogenetic diffusion models.

Viral diffusion at the intrahost level

Arguably the smallest biological scale for investigating viral evolutionary processes is provided by intrahost samples, in particular for viruses that cause chronic infections such as human immunodeficiency virus type-1 (HIV-1) [11,12] and Hepatitis C virus (HCV) [13].

In this environment, spatial structure arises from infection of different body compartments, organs, tissues, or cells characterized by specific receptors [14,15]. The extent to which viral diversity is compartmentalized [13,16–20] and how easily viruses migrate between different organs or tissues [12,21,22] are some of the evolutionary questions with important medical implications that are being addressed using phylogeographic approaches at the intrahost scale. One illustrative example explored viral migration among viral subpopulations localized in distinct brain compartments of a single patient to suggest a model for the neuropathogenesis of AIDS [12].

To our knowledge, phylogenetic diffusion of continuous viral traits within hosts has not been considered to date. However, longitudinal analysis of phenotypic data including, for example, replicative capacity [23] or antigenic information [24,25••] would be amenable to such approaches and may further advance our understanding of intrahost viral dynamics.

Current approaches to study compartmentalization frequently arrive at inconsistent findings between individual patients and sensitivity seems to decrease with increasing differences in the rates of nucleotide substitution between compartments [26]. The typically high variance of within-host evolutionary parameter estimates further complicates such studies. Similar to nucleotide substitution, molecular clock and demographic inference [27], inference of spatial diffusion may strongly benefit from hierarchical phylogenetic model (HPM) extensions that pool information across patients while still permitting patient-specific differences. Bayesian HPMs are particularly useful when data from a patient are sparse and, by introducing fixed effects, differences among patient groups can be rigorously tested [27,28].

Endemic and epidemic scales of spatial dispersal

At the interhost level, the choice for a discrete or continuous phylogeographic diffusion approach may depend

Figure 1
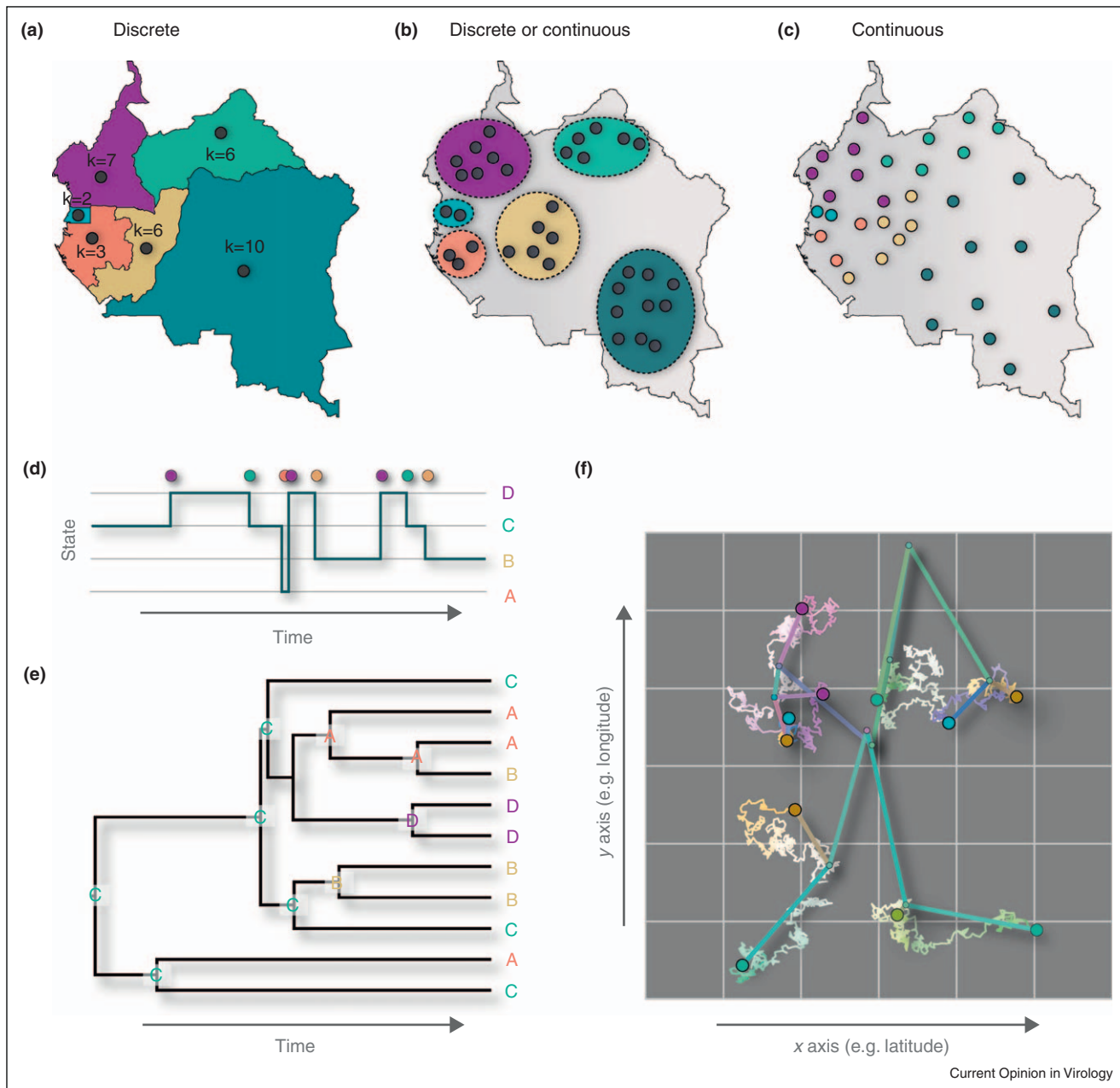
Hypothetical scenario for discretely and continuously distributed samples on the same geographical scale (top) and modeling assumptions underlying the discrete and continuous phylogeographic approaches (bottom). The choice of the phylogeographic approach depends on whether the sampling scheme is amenable to discretization or not. For example, if sequences are drawn from a single city in each country or if only the country of sampling is known (panel **a**; *k* represents the number of sequences available for each state or location), a discrete diffusion model may be preferred, although such sampling does not necessarily preclude the application of a continuous diffusion model. Intermediate scenarios may be treated either way (panel **b**). In this case, the choice may more depend on the objectives of the analysis (see Box 1). Phylogeographic inference for sequences drawn from unique locations that are continuously distributed over this geographic area and for which administrative borders do not offer a realistic discretization (panel **c**) will have to resort to continuous diffusion models. To illustrate the assumptions underlying for the discrete model, we consider a graphical representation of a four-state CTMC path (panel **d**). All possible transitions from state *i* to state *j* are color-labeled according to the end state *j* (diffusion to a location) within a time interval, although other arbitrary labels can be considered to build different counting processes [52,53]. Conditioning on the observed locations at the tips of a rooted phylogeny, CTMCs model the instantaneous locations along each branch of a tree [9•,56•,57] to infer the ancestral states at the internal nodes (panel **e**). Continuous diffusion approaches are based in Brownian diffusion models and can account for variability on the branch dispersal rates [10•]. We consider a simulation of a Brownian diffusion process, in which the lines represent branches of the tree projected on a two-dimensional arbitrary map (panel **f**). In this case, only diffusion pathways for the tips are shown.

on the sampling scheme and the question under investigation. If multiple sequences are available for a limited number of locations (Figure 1a), or if the spatial distribution of the samples is amenable to discretization into a limited set of locations (Figure 1b), a flexible continuous-time Markov chain (CTMC) model can be applied (Viral diffusion in discrete space, Box 1; Figure 1d,e). Such models are almost universally employed to model

sequence evolution and efficient algorithms exist for computing the probability of data under these. Although the discrete phylogeographic model comes at a price of making an abstraction of geography, ecological or geographical information can still be used to inform the CTMC parameters, and different parameterizations representing distinct spatial hypotheses can be tested against each other. This strategy has, for example, been

Figure 2

Current Opinion in Virology
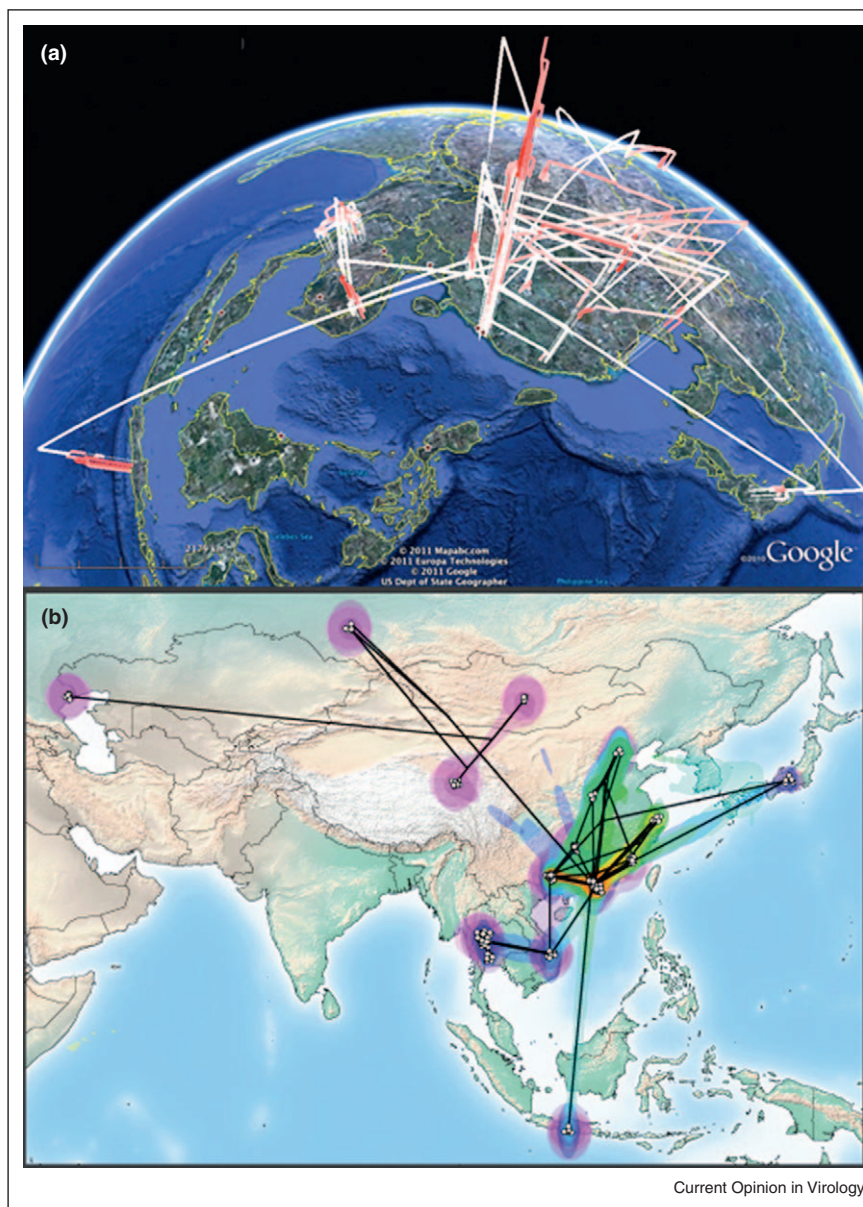
Spatio-temporal reconstruction of avian influenza H5N1 spread in Eurasia. Both panels represent a maximum clade credibility (MCC) tree summarized from a relaxed random walk analysis of previously analyzed H5N1 hemagglutinin sequences [9[•],36]. The MCC is projected in Google Earth **(a)** and Cartographica **(b)** respectively. In panel a, the heights of the branches reflect elapsed time relative to the most recent date of sampling and the gradient red-white indicates older-recent branching events; the map is based on satellite images available in Google Earth (http://earth.google.com). In panel b, transparent polygons indicate the 95% uncertainty intervals around the ancestral locations for the sampled viruses and gradient red-blue indicates recent-older rate of dispersal. The continuous phylogeographic analysis of the H5N1 epidemic corroborates the Chinese province of Guangdong as the hotspot of viral diversity [9[•],36].

employed to identify road distances as a predictor of dog rabies dispersal in North Africa [29[*]] and long-distance swine flows as a major disseminator of human-origin H1 influenza in the US swine population [30]. Stochastic mapping techniques complement this discrete approach by providing estimates of location state transitions throughout the evolutionary history (Markov jumps and rewards, Box 1).

Phylogeographic inference based on diffusion models (such as Brownian motion) in two-dimensional space can yield more realistic reconstructions of spatial history because they allow ancestral viruses to reside at any location in a continuous geographical landscape (Continuous diffusion models, Box 1; Figure 1c,f), but they are restricted by the assumption of constant rates of dispersal through time and space. Variable diffusion rates have recently been accommodated, but such relaxed Brownian processes might still be more applicable to wildlife [10[*]] (Figure 2) or plant viruses [31] than to human pathogens, which are often dispersed within complex, and far ranging, host mobility networks [32[**]]. For human viruses, Brownian-like processes may only be tenable on limited observational scales, such as a single country for locally circulating HIV-1 variants [33]. Estimates for a restricted area also avoid large-scale spatial heterogeneity, but the first steps to incorporate landscape heterogeneity explicitly are being made.

The distinct nature and scope of discrete and continuous phylogeographic inference has elegantly been illustrated by a combined strategy to examine dengue spatial dynamics in Vietnam [34[*]]. On the basis of a large set of complete genomes, this study first demonstrated independently circulating clades using a discrete approach that also considered neighboring countries. Markov jump estimates between different locations pinpointed urban areas as main sources of viral shedding [34[*]]. Brownian diffusion models were subsequently used to demonstrate different dispersion rates in rural and urban areas, suggesting key differences in the epidemiological dynamics in the distinct ecological settings. The ability to quantify dispersal rates is an interesting feature, but the estimates are contingent on the surface area under investigation. A geographically wider sampling for the same epidemic will undoubtedly yield faster diffusion rates than a geographically more restricted sampling, so these estimates find most of their use in comparative analyses on the same scale (e.g. for different viral variants [33]).

Among RNA viruses, arguably most phylogeographic efforts have been invested in characterizing influenza spread, from the emergence of new variants in bird [9[*],35,36] (Figure 2), swine [30], and human [37] populations to the dissemination of drug resistant variants [38] and the local [39] or global seasonal dynamics [40[*],41[**]] of human lineages. Nevertheless considerable debate exists

regarding the contribution of different mobility patterns to human influenza spread, for example, human workflows [42] versus airline travel [43,44] for interregional-influenza spread in the US. At the global scale, source-sink dynamics govern seasonal influenza but the exact source population has also been disputed [40[*],41[**]]. Finally, while air transportation is the most obvious predictor for global influenza spread, and this has been extensively exploited in modeling efforts, its contribution remains to be formally tested. To all these questions, model-based phylogeographic analyses can potentially provide insightful answers if applied to a comprehensive sample of viruses through time. In addition to genetic data, antigenic information has also been exploited to infer spatial dynamics of influenza [41[**]]. We note that continuous diffusion models may provide a natural framework for studying antigenic evolution, which is inherently associated with genetic history. Discrete approaches to study the history of cross-species transmission further demonstrate the generality of phylogenetic diffusion models [45,46[**]], and also risk group dynamics [47] may be within the scope of their application.

Concluding remarks and future areas of research

Recent phylogeographic developments are complementing the phylodynamic framework by providing means to statistically identify viral origins and reservoirs of genetic diversity, while setting the groundwork for quantitative studies on how and to what extent ecological features shape pathogen genetic diversity. The flexible nature of phylogenetic diffusion models allows virologists and epidemiologists to examine genotypic and phenotypic evolution both within and between hosts. Finally, we envisage that future extensions, such as a hybrid approach in which discrete phenotypic traits can act as predictors on the rate of diffusion in continuous space, will further advance our understanding of viral epidemic and evolutionary processes.

Acknowledgements

References and recommended reading
Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Holmes EC: **Evolutionary history and phylogeography of**
•   **human viruses**. *Annu Rev Microbiol* 2008, **62**:307-328.
A comprehensive review of the factors that shape the spatial distributions of human viruses.

2. Pybus OG, Rambaut A: **Evolutionary analysis of the dynamics of**
• **viral infectious disease**. *Nat Rev Genet* 2009, **10**:540-550.
Review of viral phylodynamic processes at different biological scales.

3. Holmes EC, Grenfell BT: **Discovering the phylodynamics of RNA viruses**. *PLoS Comput Biol* 2009, **5**:e1000505.

4. Henning W: *Phylogenetic Systematics*. Urbana: Press UoI; 1966.

5. Bloomquist EW, Lemey P, Suchard MA: **Three roads diverged? Routes to phylogeographic inference**. *Trends Ecol Evol* 2010, **25**:626-632.

6. Lemmon AR, Lemmon EM: **A likelihood framework for**
• **estimating phylogeographic history on a continuous landscape**. *Syst Biol* 2008, **57**:544-561.
Introduction of bivariate Brownian diffusion models in a likelihood framework.

7. Sanmartin I, van der Mark P, Ronquist F: **Inferring dispersal: a Bayesian approach to phylogeny-based island biogeography, with special reference to the Canary Islands**. *J Biogeogr* 2008, **35**:428-449.

8. Drummond AJ, Rambaut A: **BEAST: Bayesian evolutionary analysis by sampling trees**. *BMC Evol Biol* 2007, **7**:214.

9. Lemey P, Rambaut A, Drummond AJ, Suchard MA: **Bayesian**
• **phylogeography finds its roots**. *PLoS Comput Biol* 2009, **5**:e1000520.
Introduction of Bayesian inference methods for discrete viral phylogeographic analyses including a Bayesian stochastic search variable selection procedure.

10. Lemey P, Rambaut A, Welch JJ, Suchard MA: **Phylogeography**
• **takes a relaxed random walk in continuous space and time**. *Mol Biol Evol* 2010, **27**:1877-1885.
Description of a Bayesian implementation of relaxed Brownian diffusion models for continuous phylogeographic analyses.

11. Salemi M, Burkhardt BR, Gray RR, Ghaffari G, Sleasman JW, Goodenow MM: **Phylodynamics of HIV-1 in lymphoid and non-lymphoid tissues reveals a central role for the thymus in emergence of CXCR4-using quasispecies**. *PLoS One* 2007, **2**:e950.

12. Salemi M, Lamers SL, Yu S, de Oliveira T, Fitch WM, McGrath MS: **Phylodynamic analysis of human immunodeficiency virus type 1 in distinct brain compartments provides a model for the neuropathogenesis of AIDS**. *J Virol* 2005, **79**:11343-11352.

13. Navas S, Martin J, Quiroga JA, Castillo I, Carreno V: **Genetic diversity and tissue compartmentalization of the hepatitis C virus genome in blood mononuclear cells, liver, and serum from chronic hepatitis C patients**. *J Virol* 1998, **72**:1640-1646.

14. Alizon S, Luciani F, Regoes RR: **Epidemiological and clinical consequences of within-host evolution**. *Trends Microbiol* 2010, **19**:24-32.

15. Parrish CR, Holmes EC, Morens DM, Park EC, Burke DS, Calisher CH, Laughlin CA, Saif LJ, Daszak P: **Cross-species virus transmission and the emergence of new epidemic diseases**. *Microbiol Mol Biol Rev* 2008, **72**:457-470.

16. Delwart EL, Mullins JI, Gupta P, Learn GH Jr, Holodniy M, Katzenstein D, Walker BD, Singh MK: **Human immunodeficiency virus type 1 populations in blood and semen**. *J Virol* 1998, **72**:617-623.

17. Fulcher JA, Hwangbo Y, Zioni R, Nickle D, Lin X, Heath L, Mullins JI, Corey L, Zhu T: **Compartmentalization of human immunodeficiency virus type 1 between blood monocytes and CD4+ T cells during infection**. *J Virol* 2004, **78**:7883-7893.

18. Kemal KS, Foley B, Burger H, Anastos K, Minkoff H, Kitchen C, Philpott SM, Gao W, Robison E, Holman S *et al.*: **HIV-1 in genital tract and plasma of women: compartmentalization of viral sequences, coreceptor usage, and glycosylation**. *Proc Natl Acad Sci U S A* 2003, **100**:12972-12977.

19. Pillai SK, Good B, Pond SK, Wong JK, Strain MC, Richman DD, Smith DM: **Semen-specific genetic characteristics of human immunodeficiency virus type 1 env**. *J Virol* 2005, **79**:1734-1742.

20. Potter SJ, Lemey P, Achaz G, Chew CB, Vandamme AM, Dwyer DE, Saksena NK: **HIV-1 compartmentalization in diverse leukocyte populations during antiretroviral therapy**. *J Leukoc Biol* 2004, **76**:562-570.

21. Lamers SL, Gray RR, Salemi M, Huysentruyt LC, McGrath MS: **HIV-1 phylogenetic analysis shows HIV-1 transits through the meninges to brain and peripheral tissues**. *Infect Genet Evol* 2011, **11**:31-37.

22. Salemi M, Lamers SL, Huysentruyt LC, Galligan D, Gray RR, Morris A, McGrath MS: **Distinct patterns of HIV-1 evolution within metastatic tissues in patients with non-Hodgkins lymphoma**. *PLoS One* 2009, **4**:e8153.

23. Bunnik EM, Euler Z, Welkers MR, Boeser-Nunnonk BDM, Grijsen ML, Prins JM, Schuitemaker H: **Adaptation of HIV-1 envelope gp120 to humoral immunity at a population level**. *Nat Med* 2010, **16**:995-997.

24. Nelson MI, Holmes EC: **The evolution of epidemic influenza**. *Nat Rev Genet* 2007, **8**:196-205.

25. Smith DJ, Lapedes AS, de Jong JC, Bestebroer TM,
•• Rimmelzwaan GF, Osterhaus AD, Fouchier RA: **Mapping the antigenic and genetic evolution of influenza virus**. *Science* 2004, **305**:371-376.
Elegant demonstration of antigenic cartography based on binding assay data to study the evolution of antigenicity in human influenza.

26. Zarate S, Pond SL, Shapshak P, Frost SD: **Comparative study of methods for detecting sequence compartmentalization in human immunodeficiency virus type 1**. *J Virol* 2007, **81**:6643-6651.

27. Edo-Matas D, Lemey P, Tom JA, Serna-Bolea C, van den Blink AE, van't Wout AB, Schuitemaker H, Suchard MA: **Impact of CCR5delta32 host genetic background and disease progression on HIV-1 intrahost evolutionary processes: efficient hypothesis testing through hierarchical phylogenetic models**. *Mol Biol Evol* 2010, **28**:1605-1616.

28. Suchard MA, Kitchen CM, Sinsheimer JS, Weiss RE: **Hierarchical phylogenetic models for analyzing multipartite sequence data**. *Syst Biol* 2003, **52**:649-664.

29. Talbi C, Lemey P, Suchard MA, Abdelatif E, Elharrak M, Jalal N,
• Faouzi A, Echevarria JE, Vazquez Moron S, Rambaut A *et al.*: **Phylodynamics and human-mediated dispersal of a zoonotic virus**. *PLoS Pathog* 2010, **6**:e1001166.
Shows how different phylogeographic hypotheses for rabies dispersal in North Africa can be rigorously tested using the discrete Bayesian statistical approach.

30. Nelson MI, Lemey P, Tan Y, Vincent A, Tsan-Yuk Lam T, Detmer S, Viboud C, Suchard M, Rambaut A, Holmes EC, Gramer M: **Spatial dynamics of human-origin H1 influenza A virus in North American swine**. *PLoS Pathog* 2011, **7**:e1002077.

31. Monjane AL, Harkins GW, Martin DP, Lemey P, Lefeuvre P, Shepherd DN, Oluwafemi S, Simuyandi M, Zinga I, Komba EK *et al.*: **Reconstructing the history of Maize streak virus strain-A dispersal to reveal diversification hotspots and its origin in southern Africa**. *J Virol* 2011, **85**:9623-9636.

32. Brockmann D, Hufnagel L, Geisel T: **The scaling laws of human**
•• **travel**. *Nature* 2006, **439**:462-465.
Important study demonstrating the use of money circulation data to characterize and quantify patterns of human travel.

33. Faria NR, Suchard MA, Abecasis A, Sousa JD, Ndembi N, Bonfim I, Camacho RJ, Vandamme AM, Lemey P: **Phylodynamics of the HIV-1 CRF02_AG clade in Cameroon**. *Infect Genet Evol* 2011 doi: 10.1016/j.meegid.2011.04.028.

34. Raghwani J, Rambaut A, Holmes EC, Hang VT, Hien TT, Farrar J,
• Wills B, Lennon NJ, Birren BW, Henn MR *et al.*: **Endemic dengue associated with the co-circulation of multiple viral lineages and localized density-dependent transmission**. *PLoS Pathog* 2011, **7**:e1002064.
A recent paper on the application of phylogeographic methods to infer the evolutionary history of dengue virus at different scales.

35. Wallace RG, Fitch WM: **Influenza A H5N1 immigration is filtered out at some international borders**. *PLoS One* 2008, **3**:e1697.

36. Wallace RG, Hodac H, Lathrop RH, Fitch WM: **A statistical phylogeography of influenza A H5N1**. *Proc Natl Acad Sci U S A* 2007, **104**:4473-4478.

37. Lemey P, Suchard M, Rambaut A: **Reconstructing the initial global spread of a human influenza pandemic: a Bayesian spatial-temporal model for the global spread of H1N1pdm**. *PLoS Curr* 2009, **1**:RRN1031.

38. Nelson MI, Simonsen L, Viboud C, Miller MA, Holmes EC: **The origin and global emergence of adamantane resistant A/H3N2 influenza viruses**. *Virology* 2009, **388**:270-278.

39. Nelson MI, Edelman L, Spiro DJ, Boyne AR, Bera J, Halpin R, Sengamalay N, Ghedin E, Miller MA, Simonsen L *et al.*: **Molecular epidemiology of A/H3N2 and A/H1N1 influenza virus during a single epidemic season in the United States**. *PLoS Pathog* 2008, **4**:e1000133.

40. Bedford T, Cobey S, Beerli P, Pascual M: **Global migration**
• **dynamics underlie evolution and persistence of human influenza A (H3N2)**. *PLoS Pathog* 2010, **6**:e1000918.
This study uses a structured coalescent approach to unravel the source-sink dynamics governing the seasonal flu epidemics.

41. Russell CA, Jones TC, Barr IG, Cox NJ, Garten RJ, Gregory V,
•• Gust ID, Hampson AW, Hay AJ, Hurt AC *et al.*: **The global circulation of seasonal influenza A (H3N2) viruses**. *Science* 2008, **320**:340-346.
Important description of the source-sink dynamics in global seasonal flu epidemics using both genetic and antigenic data.

42. Viboud C, Bjornstad ON, Smith DL, Simonsen L, Miller MA, Grenfell BT: **Synchrony, waves, and spatial hierarchies in the spread of influenza**. *Science* 2006, **312**:447-451.

43. Brownstein JS, Wolfe CJ, Mandl KD: **Empirical evidence for the effect of airline travel on inter-regional influenza spread in the United States**. *PLoS Med* 2006, **3**:e401.

44. Viboud C, Miller MA, Grenfell BT, Bjornstad ON, Simonsen L: **Air travel and the spread of influenza: important caveats**. *PLoS Med* 2006, **3**:e503 author reply e502.

45. Hass J, Matuszewski S, Cieslik D, Haase M: **The role of swine as ''mixing vessel'' for interspecies transmission of the influenza A subtype H1N1: a simultaneous Bayesian inference of phylogeny and ancestral hosts**. *Infect Genet Evol* 2010, **11**:437-441.

46. Streicker DG, Turmelle AS, Vonhof MJ, Kuzmin IV, McCracken GF,
•• Rupprecht CE: **Host phylogeny constrains cross-species emergence and establishment of rabies virus in bats**. *Science* 2010, **329**:676-679.
This study elucidates the factors underlying rabies cross-species transmission in a large dataset of North American bats.

47. Holmes EC, Zhang LQ, Robertson P, Cleland A, Harvey E, Simmonds P, Leigh Brown AJ: **The molecular epidemiology of human immunodeficiency virus type 1 in Edinburgh**. *J Infect Dis* 1995, **171**:45-53.

48. Rabaa MA, Ty Hang VT, Wills B, Farrar J, Simmons CP, Holmes EC: **Phylogeography of recently emerged DENV-2 in southern Viet Nam**. *PLoS Negl Trop Dis* 2010, **4**:e766.

49. Edwards CJSM, Lemey P, Welch JJ, Barnes I, Fulton TL, Barnett R, O'Connell TC, Coxon P, Monaghan N, Valdiosera CE, Lorenzen ED, WIllerslev E *et al.*: **Multiple hybridizations between ancient brown and polar bears and an Irish origin for the modern polar bear matriline**. *Curr Biol* 2011, **21**:1251-1258.

50. Suchard MA, Weiss RE, Sinsheimer JS: **Bayesian selection of continuous-time Markov chain evolutionary models**. *Mol Biol Evol* 2001, **18**:1001-1013.

51. Minin VN, Suchard MA: **Fast, accurate and simulation-free stochastic mapping**. *Philos Trans R Soc Lond B Biol Sci* 2008, **363**:3985-3995.

52. Minin VN, Suchard MA: **Counting labeled transitions in continuous-time Markov models of evolution**. *J Math Biol* 2008, **56**:391-412.

53. O'Brien JD, Minin VN, Suchard MA: **Learning to count: robust estimates for labeled distances between molecular sequences**. *Mol Biol Evol* 2009, **26**:801-814.

54. Drummond AJ, Ho SY, Phillips MJ, Rambaut A: **Relaxed phylogenetics and dating with confidence**. *PLoS Biol* 2006, **4**:e88.

55. Bielejec F RA, Suchard MA, Lemey P: **SPREAD: Spatial phylogenetic reconstruction of evolutionary dynamics**. *Bioinformatics* 2011, **27**:2910-2912.

56. Pagel M: **Inferring the historical patterns of biological**
• **evolution**. *Nature* 1999, **401**:877-884.
An excellent review on the problem of ancestral state reconstruction and the statistical approaches to tackle it.

57. Ronquist F, Huelsenbeck JP: **MrBayes 3: Bayesian phylogenetic inference under mixed models**. *Bioinformatics* 2003, **19**:1572-1574.